A CONSISTENT TEST FOR CONDITIONAL SYMMETRY AND ITS ASYMPTOTICAL NORMALITY

CHEN Min-qiong

(*School of Economics and Trade, Xinhua College of Sun Yat-Sen University, Guangzhou 510520, China*)

(*School of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China*)

Abstract: In this paper, we investigate the problem of testing the conditional symmetry of a random vector given another random vector. We propose a new test based on the concept of conditional energy distance. The test statistic has the form of a U-statistic with random kernel. By using the theory of U-statistic, we prove that the test statistic is asymptotically normal under the null hypothesis of conditional symmetry and consistent against any conditional asymmetric distribution.

Keywords: conditional symmetry test; conditional energy distance; U-statistic with random kernel; consistent; asymptotical normality

2010 MR Subject Classification: 62G10

Document code: A   **Article ID:** 0255-7797(2019)02-0159-12

1 Introduction

In many regression models, specially the econometric models, for the purpose of identification, some distributional assumptions are often imposed on the error term. The assumptions are conditional moment restrictions, independence between observations, and conditional symmetry around zero given the independent variables. There were a few semiparametric estimators proposed under conditional symmetry. Manski [1] and Newey [2] estimated regression models under conditional symmetry. Powell [3] and Newey [4] proposed semiparametric estimations for Tobit models under conditional symmetry.

Despite the wide use of the property of conditional symmetry, tests for conditional symmetry were not addressed very much in the literature. The first tests were proposed by Powell [5] for censored regression models and by Newey and Powell [6] for linear regression models via asymmetric least squares estimation. However these tests are unlikely to be consistent against all conditional asymmetric distributions. Zheng [7] proposed a consistent test of conditional symmetry using a kernel method, but the test statistic contains integral term

and is hard to implement. Bai and Ng [8] proposed an alternative test for conditional symmetry for time series models. The test relied on the correct specification of both conditional mean and conditional variance. Hyndman and Yao [9] developed a bootstrap test for the symmetry of conditional density functions based on their improved methods for conditional density estimation, but they didn't discuss the asymptotic properties of the test statistic, so it is not clear whether the test be consistent or not. Su [10] gave a simple consistent nonparametric test of conditional symmetry based on the principle of conditional characteristic functions, and he [11] also gave an unconditional method by transforming the conditional symmetry test problem to a unconditional test one. Both of the test statistics he presented in paper need a given characteristic function of the probability measure for the value space of the conditional variable.

In this paper, we propose a simple test for conditional symmetry based on the concept of conditional energy distance. The test is shown to be asymptotically normal under the null hypothesis of conditional symmetry and consistent against any conditional asymmetric distribution. Our test statistic only contains the Euclidean distances and kernel function, so it is easy to compute.

2 The Test Statistic for Conditional Symmetry

Székely [12] introduced a new concept named energy distance to measure the difference between two independent probability distributions. If X and Y are independent random vectors in $\mathbb{R}^p$ with cumulative distribution functions (cdf) F and G respectively, then the energy distance between the distributions F and G is defined as

$$\varepsilon(F,G) = 2E|X-Y| - E|X-X'| - E|Y-Y'|, \tag{2.1}$$

where X' is an i.i.d. copy of X and Y' is an i.i.d. copy of Y, E is the expected value, and $|\,.\,|$ denotes the Euclidean norm. One can also write $\varepsilon(F,G)$ as $\varepsilon(X,Y)$, and call it the energy distance of X and Y. Székely [12] proved that for real-valued random variables this distance is exactly twice Harald Cramér's distance, that is

$$2\int_{-\infty}^{\infty}(F(t)-G(t))^2 dt = 2E|X-Y| - E|X-X'| - E|Y-Y'|.$$

In higher dimensions, however, the two distances are different because the energy distance is rotation invariant while Cramér's distance is not. The equality becomes

$$2E|X-Y| - E|X-X'| - E|Y-Y'| = \frac{1}{c_p}\int_{\mathbb{R}^p}\frac{|\phi_X(t)-\phi_Y(t)|^2}{|t|^{p+1}}dt, \tag{2.2}$$

where $\phi_X(t)$ is X's characteristic function and $\phi_Y(t)$ is Y's characteristic function, $c_p = \frac{\pi^{(p+1)/2}}{\Gamma(\frac{p+1}{2})}$. Thus $\varepsilon(F,G) \geq 0$ with equality to zero if and only if $F = G$. This property makes it possible to use $\varepsilon(F,G)$ for testing goodness-of-fit, homogeneity, etc. in a consistent way. We shall draw the consistent test statistic for conditional symmetry from the thought of energy distance.

Let X be a p dimensional random vector in Euclidean space $\mathbb{R}^p$, Z be a r dimensional random vector in Euclidean space $\mathbb{R}^r$. Denote $f(x|z)$ as the conditional density function of X given Z. Consider the hypothesis

$$H_0 : f(x|z) = f(-x|z) \text{ for all } x \in \mathbb{R}^p, z \in S(Z), \tag{2.3}$$

where $S(Z)$ denotes the support of the density function of Z. Note that the null hypothesis (2.3) can be expressed equivalently as

$$H_0 : X|Z \overset{D}{=} -X|Z. \tag{2.4}$$

Analogous to the concept of energy distance for two independent vectors, we can also define the conditional energy distance between X and $-X$ given Z as follows.

Definition 2.1 The conditional energy distance $\epsilon(X, -X|Z)$ between X and $-X$ with finite first moment given Z is defined as the square root of

$$\varepsilon^2(X, -X|Z) = \|\phi_{X|Z}(t) - \phi_{X|Z}(-t)\|^2 := \frac{1}{c_p} \int_{\mathbb{R}^p} \frac{|\phi_{X|Z}(t) - \phi_{X|Z}(-t)|^2}{|t|^{p+1}} dt, \tag{2.5}$$

where $\phi_{X|Z}(t)$ is the conditional characteristic function of X given Z. Therefore H_0 holds if and only if $\epsilon(X, -X|Z) = 0$.

Let $W_i = (X_i, Z_i), i = 1, 2, \cdots, n$ be a sample from the distribution of (X, Z) and denote $\mathbf{W} = (\mathbf{X}, \mathbf{Z}) = \{W_1, W_2, \cdots, W_n\}$. Then for the specific value of $\varepsilon^2(X, -X|Z)$ when given $Z = z$, $\epsilon(X, -X|Z)$ can be rewritten as the form of expectation by the following lemma.

Lemma 2.1 $\varepsilon^2(X, -X|Z = z)$ can be rewritten as the form of

$$\varepsilon^2(X, -X|Z = z) = 2E[|X_1 + X_2| - |X_1 - X_2||Z_1 = Z_2 = z]. \tag{2.6}$$

Therefore, $X|Z = z \overset{D}{=} -X|Z = z$ for any z if and only if

$$E[|X_1 + X_2| - |X_1 - X_2||Z_1 = Z_2 = z] = 0.$$

Proof Given the event $Z = z$, we consider

$$|\phi_{X|Z=z}(t) - \phi_{X|Z=z}(-t)|^2$$
$$= \phi_{X|Z=z}(t)\overline{\phi_{X|Z=z}(t)} + \phi_{X|Z=z}(-t)\overline{\phi_{X|Z=z}(-t)}$$
$$\quad - \phi_{X|Z=z}(t)\overline{\phi_{X|Z=z}(-t)} - \phi_{X|Z=z}(-t)\overline{\phi_{X|Z=z}(t)}$$
$$= E[\exp(i < t, X_1 - X_2 >)|Z_1 = z, Z_2 = z]$$
$$\quad + E[\exp(i < t, -X_1 + X_2 >)|Z_1 = z, Z_2 = z]$$
$$\quad - E[\exp(i < t, X_1 + X_2 >)|Z_1 = z, Z_2 = z]$$
$$\quad - E[\exp(i < t, -X_1 - X_2 >)|Z_2 = z, Z_2 = z]$$
$$= 1 - E[\exp(i < t, X_1 + X_2 >)|Z_1 = z, Z_2 = z] + 1$$
$$\quad - E[\exp(i < t, -X_1 - X_2 >)|Z_1 = z, Z_2 = z]$$
$$\quad - (1 - E[\exp(i < t, X_1 - X_2 >)|Z_1 = z, Z_2 = z])$$
$$\quad - (1 - E[\exp(i < t, -X_1 + X_2 >)|Z_1 = z, Z_2 = z]).$$

According to the equation [12],

$$\int_{\mathbb{R}^p} \frac{1 - \exp(i\langle t, X\rangle)}{|t|^{p+1}} dt = c_p |X|_p, \tag{2.7}$$

we have

$$\varepsilon^2(X, -X|Z=z) = \|\phi_{X|Z=z}(t) - \phi_{X|Z=z}(-t)\|^2$$
$$= \frac{1}{c_p} \int_{\mathbb{R}^p} \frac{|\phi_{X|Z=z}(t) - \phi_{X|Z=z}(-t)|^2}{|t|^{p+1}} dt = 2E[|X_1 + X_2| - |X_1 - X_2||Z_1 = z, Z_2 = z].$$

Let

$$\mathcal{S}_a = E\left[\frac{1}{2}\varepsilon^2(X, -X|Z)f^2(Z)\right], \tag{2.8}$$

where $f(Z)$ is the density function of Z. Consequently, $X|Z \overset{D}{=} -X|Z$ if and only if $\mathcal{S}_a = 0$. Naturally, we can choose test statistic for H_0 as

$$\mathcal{U}_n := \frac{1}{C_n^3} \sum_{i<j<k} (|X_i + X_j| - |X_i - X_j|)\omega_i(Z_k)\omega_j(Z_k)$$
$$= \frac{1}{C_n^3} \sum_{i<j<k} (|X_i + X_j| - |X_i - X_j|)K_H(Z_i - Z_k)K_H(Z_j - Z_k)$$
$$= \frac{1}{C_n^3|H|^2} \sum_{i<j<k} (|X_i + X_j| - |X_i - X_j|)K_{ik}K_{jk},$$

where $K_{ik} = K(H^{-1}(Z_i - Z_k))$.

The test statistic $\mathcal{U}_n$ has the advantage that it has zero mean under H_0 and hence it does not have a finite sample bias term. We show the consistent of $\mathcal{U}_n$ and its asymptotical normality under H_0.

Here, we choose the Gaussian kernel

$$K_H(\mathbf{u}) = |H|^{-1}K(H^{-1} \cdot \mathbf{u}) = (2\pi)^{-\frac{r}{2}}|H|^{-1}\exp\left(-\frac{1}{2}\mathbf{u}'H^{-2}\mathbf{u}\right)$$

in $\mathbb{R}^r$, where H is a diagonal matrix $\text{diag}\{h, h, \cdots, h\}$ determined by bandwidth h. With the Gaussian kernel, $\sum_i \omega_i(Z)/n$ is known to be consistent under the following regularity conditions.

(C1)

$$\int_{\mathbb{R}^r} \mathbf{u}K(\mathbf{u})d\mathbf{u} = 0, \int_{\mathbb{R}^r} K(\mathbf{u})d\mathbf{u} = 1, \int_{\mathbb{R}^r} |K(\mathbf{u})|d\mathbf{u} < \infty,$$
$$\int_{\mathbb{R}^r} K^2(\mathbf{u})d\mathbf{u} < \infty, \int_{\mathbb{R}^r} \mathbf{u}^2 K(\mathbf{u})d\mathbf{u} < \infty.$$

(C2) $h^r \longrightarrow 0$ and $nh^r \longrightarrow \infty$ as $n \longrightarrow \infty$. This requires h to be chosen appropriately according to n.

(C3) The density function of Z and the conditional density function $f(\cdot|z)$ are twice differentiable and all of the derivatives are bounded.

3 Asymptotical Normality of $\mathcal{U}_n$ under Null Hypothesis

Using the theory of U-statistic discussed by Fan and Li [13] and Lee [14], we have the following asymptotical normality.

Theorem 3.1 (Weak convergence) Assume that conditions (C1)–(C3) hold and the second moment of X exists, if the conditional density of X given Z is symmetric and if $h \longrightarrow 0$ and $nh^r \longrightarrow \infty$ as $n \longrightarrow \infty$, we have $nh^{r/2}\mathcal{U}_n \xrightarrow[n\to\infty]{d} N(0,\sigma^2)$, where σ^2 is given in (3.5).

Proof Let $P_n(W_1,W_2,W_3) = (|X_1+X_2|-|X_1-X_2|)K_{13}K_{23}$. Note that $P_n(W_1,W_2,W_3)$ is not symmetric with respect to W_1,W_2,W_3, so we symmetrize P_n as

$$\mathcal{P}_n(W_1,W_2,W_3) = \frac{1}{3}[P_n(W_1,W_2,W_3) + P_n(W_3,W_2,W_1) + P_n(W_1,W_3,W_2)],$$

then $\mathcal{U}_n$ can be expressed as a U-statistic of degree 3 with random kernel,

$$\mathcal{U}_n = \frac{1}{C_n^3 h^{2r}} \sum_{i<j<k} \mathcal{P}_n(W_i,W_j,W_k).$$

Denote that

$$\begin{aligned}
&P_{n1}(W_i) = E[P_n(W_1,W_2,W_3)|W_i], i=1,2,3, \\
&P_{nc}(W_1,\cdots,W_c) = E[P_n(W_1,W_2,W_3)|W_1,\cdots,W_c], c=2,3, \\
&\mathcal{P}_{n1}(W_1) = E[\mathcal{P}_n(W_1,W_2,W_3)|W_1], \\
&\mathcal{P}_{nc}(W_1,\cdots,W_c) = E[\mathcal{P}_n(W_1,W_2,W_3)|W_1,\cdots,W_c], c=2,3,
\end{aligned} \tag{3.1}$$

and

$$\sigma_{n1}^2 = \mathrm{Var}(\mathcal{P}_{n1}(W_1)), \quad \sigma_{n2}^2 = \mathrm{Var}(\mathcal{P}_{n2}(W_1,W_2)), \quad \sigma_{n3}^2 = \mathrm{Var}(\mathcal{P}_n(W_1,W_2,W_3)). \tag{3.2}$$

We use Lemma B.4 in Fan and Li [13] to obtain the asymptotical distribution of $\mathcal{U}_n$ under H_0 in the following steps.

Step 1 Under H_0, $E\mathcal{P}_n(W_1,W_2,W_3) = 0$. Note that

$$E[\mathcal{P}_n(W_1,W_2,W_3)] = E[(|X_1+X_2|-|X_1-X_2|)K_{13}K_{23}]$$

$$= \int (|x_1+x_2|-|x_1-x_2|)K(H^{-1}(z_1-z_3))K(H^{-1}(z_2-z_3))$$

$$\cdot f(x_1,z_1)f(x_2,z_2)f(x_3,z_3)dx_1dx_2dx_3dz_1dz_2dz_3$$

$$= h^{2r} \int (|x_1+x_2|-|x_1-x_2|)K(z_{13})K(z_{23})f(z_3+Hz_{13})f(x_1|z_3+Hz_{13})$$

$$\cdot f(z_3+Hz_{23})f(x_2|z_3+Hz_{23})f(z_3)dx_1dx_2dz_{13}dz_{23}dz_3$$

$$= h^{2r} \int K(z_{13})K(z_{23})f(z_3)f(z_3+Hz_{13})f(z_3+Hz_{23})$$

$$\cdot \left(\int (|x_1+x_2|-|x_1-x_2|)f(x_1|z_3+Hz_{13})f(x_2|z_3+Hz_{23})dx_1dx_2 \right)dz_{13}dz_{23}dz_3 = 0.$$

Step 2   Under H_0, $E[\mathcal{P}_n(W_1, W_2, W_3)|W_1] = 0$. Because

$$
E[P_n(W_1, W_2, W_3)|W_1]
$$
$$
= \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(H^{-1}(z_2 - z_3))f(x_2, z_2)f(z_3)dx_2dz_2dz_3
$$
$$
= h^{2r} \int (|x_1 + x_2| - |x_1 - x_2|)K(z_{13})K(z_{23})f(z_1 - Hz_{13})
$$
$$
\cdot f(z_1 + H(z_{13} - z_{23}))f(x_2|z_1 + H(z_{23} - z_{13}))dx_2dz_{13}dz_{23}
$$
$$
= h^{2r} \int f(z_1 + H(z_{23} - z_{13}))f(z_1 - Hz_{13})K(z_{13})K(z_{23})
$$
$$
\cdot \left\{ \int (|x_1 + x_2| - |x_1 - x_2|)f(x_2|z_1 + H(z_{23} - z_{13}))dx_2 \right\}dz_{13}dz_{23} = 0,
$$

which also implies that $E[P_n(W_1, W_3, W_2)|W_1] = 0$. Moreover, note that

$$
E[P_n(W_1, W_2, W_3)|W_3]
$$
$$
= \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(H^{-1}(z_2 - z_3))f(x_1, z_1)f(x_2, z_2)dx_1dx_2dz_1dz_2
$$
$$
= h^{2r} \int f(z_3 + Hz_{13})f(z_3 + Hz_{23})K(z_{13})K(z_{23})
$$
$$
\cdot \left\{ \int (|x_1 + x_2| - |x_1 - x_2|)f(x_1|z_3 + Hz_{13})f(x_2|z_3 + Hz_{23})dx_1dx_2 \right\}dz_{13}dz_{23}
$$
$$
= 0
$$

implies $E[P_n(W_3, W_2, W_1)|W_1] = 0$. By the definition of $\mathcal{P}_n(W_1, W_2, W_3)$, we have

$$
E[\mathcal{P}_n(W_1, W_2, W_3)|W_1]
$$
$$
= \frac{1}{3}\{E[P_n(W_1, W_2, W_3)|W_1] + E[P_n(W_3, W_2, W_1)|W_1] + E[P_n(W_1, W_3, W_2)|W_1]\}
$$
$$
= 0.
$$

Step 3 $\sigma_{n3}^2/\sigma_{n2}^2 = o(n)$. Obviously, under H_0,

$$
\sigma_{n2}^2 = E\mathcal{P}_{n2}^2(W_1, W_2), \ \ \sigma_{n3}^2 = E\mathcal{P}_n^2(W_1, W_2, W_3).
$$

For $E\mathcal{P}_{n2}^2(W_1, W_2)$, we have

$$
E\mathcal{P}_{n2}^2(W_1, W_2)
$$
$$
= \frac{1}{9}\left\{ EP_{n2}^2(W_1, W_2) + 4EP_{n2}^2(W_1, W_3) + 2EP_{n2}(W_1, W_2)P_{n2}(W_1, W_3) \right\}, \tag{3.3}
$$

where

$$
\begin{aligned}
&EP_{n2}^2(W_1, W_2) \\
&= \int \left(\int (|x_1 + x_2| - |x_1 - x_2|) K(H^{-1}(z_1 - z_3)) K(H^{-1}(z_2 - z_3)) f(z_3) dz_3 \right)^2 \\
&\quad \cdot f(x_1, z_1) f(x_2, z_2) dx_1 dx_2 dz_1 dz_2 \\
&= h^{2r} \int \left(\int (|x_1 + x_2| - |x_1 - x_2|) K(z_{31}) K(H^{-1}(z_1 - z_2) + z_{31}) f(z_1 + Hz_{31}) dz_{31} \right)^2 \\
&\quad \cdot f(x_1, z_1) f(x_2, z_2) dx_1 dx_2 dz_1 dz_2 \\
&= h^{3r} \int \left(\int (|x_1 + x_2| - |x_1 - x_2|) K(z_{31}) K(z_{12} + z_{31}) f(z_1 + Hz_{31}) dz_{31} \right)^2 \\
&\quad \cdot f(x_1, z_2 + Hz_{12}) f(x_2, z_2) dx_1 dx_2 dz_{12} dz_2 \\
&= O_p(h^{3r}).
\end{aligned}
$$

By considering $P_{n2}^2(W_1, W_3)$ and $EP_{n2}(W_1, W_2)P_{n2}(W_1, W_3)$ in a similar way, we get

$$
P_{n2}^2(W_1, W_3) = O_p(h^{3r}) \quad \text{and} \quad EP_{n2}(W_1, W_2)P_{n2}(W_1, W_3) = O_p(h^{3r}),
$$

which implies that $\sigma_{n2}^2 = EP_{n2}^2(W_1, W_2) = O_p(h^{3r})$.

For $EP_n^2(W_1, W_2, W_3)$, we have

$$
\begin{aligned}
&E\mathcal{P}_n^2(W_1, W_2, W_3) \\
&= \frac{1}{9} \big(3EP_n^2(W_1, W_2, W_3) + 2EP_n(W_1, W_2, W_3) P_n(W_3, W_2, W_1) \\
&\quad + 2EP_n(W_1, W_2, W_3) P_n(W_1, W_3, W_2) + 2EP_n(W_3, W_2, W_1) P_n(W_1, W_3, W_2) \big)
\end{aligned} \tag{3.4}
$$

with

$$
\begin{aligned}
&EP_n^2(W_1, W_2, W_3) \\
&= \int (|x_1 + x_2| - |x_1 - x_2|)^2 K^2(H^{-1}(z_1 - z_3)) \\
&\quad K^2(H^{-1}(z_2 - z_3)) f(z_3) f(x_1, z_1) f(x_2, z_2) dx_1 dx_2 dz_1 dz_2 dz_3 \\
&= h^{2r} \int (|x_1 + x_2| - |x_1 - x_2|)^2 K^2(z_{13}) K^2(z_{23}) \\
&\quad f(z_3) f(x_1, z_3 + Hz_{13}) f(x_2, z_3 + Hz_{23}) dx_1 dx_2 dz_{13} dz_{23} dz_3 \\
&= O_p(h^{2r}).
\end{aligned}
$$

Similarly, we can prove the rest three terms in (3.4) are all $O_p(h^{2r})$, which implies that $\sigma_{n3}^2 = E\mathcal{P}^2(W_1, W_2, W_3) = O_p(h^{2r})$. Thus $\sigma_{n3}^2/\sigma_{n2}^2 = O_P(\frac{1}{h^r}) = o(n)$ holds.

Step 4 We need to prove that, when $n \longrightarrow \infty$,

$$
\frac{E\mathcal{G}_n^2(W_1, W_2)) + n^{-1} E\mathcal{P}_{n2}^4(W_1, W_2)}{(E\mathcal{P}_{n2}^2(W_1, W_2))^2} \longrightarrow 0,
$$

where

$$G_n(W_1, W_2) = E(P_{n2}(W_1, W_3)P_{n2}(W_2, W_3)|W_1, W_2),$$
$$\mathcal{G}_n(W_1, W_2) = E(\mathcal{P}_{n2}(W_1, W_3)\mathcal{P}_{n2}(W_2, W_3)|W_1, W_2).$$

As we discussed in Step 3, $E\mathcal{P}_{n2}^2(W_1, W_2) = O_p(h^{3r})$. Thus $(E\mathcal{P}_{n2}^2(W_1, W_2))^2 = O_p(h^{6r})$. Analogously to $E\mathcal{P}_{n2}^2(W_1, W_2)$, we can prove that $E\mathcal{P}_{n2}^4(W_1, W_2) = O_p(h^{5r})$ by noting that

$$EP_{n2}^4(W_1, W_2)$$
$$= \int \left(\int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(H^{-1}(z_2 - z_3))f(z_3)dz_3 \right)^4$$
$$\cdot f(x_1, z_1)f(x_2, z_2)dx_1dx_2dz_1dz_2$$
$$= h^{4r} \int \left(\int (|x_1 + x_2| - |x_1 - x_2|)K(z_{31})K(H^{-1}(z_1 - z_2) + z_{31})f(z_1 + Hz_{31})dz_{31} \right)^4$$
$$\cdot f(x_1, z_1)f(x_2, z_2)dx_1dx_2dz_1dz_2$$
$$= h^{5r} \int \left(\int (|x_1 + x_2| - |x_1 - x_2|)K(z_{31})K(z_{12} + z_{31})f(z_1 + Hz_{31})dz_{31} \right)^4$$
$$\cdot f(x_1, z_2 + Hz_{12})f(x_2, z_2)dx_1dx_2dz_{12}dz_2$$
$$= O_p(h^{5r}).$$

Moreover,

$$P_{n2}(W_1, W_3)$$
$$= \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(H^{-1}(z_2 - z_3))f(x_2, z_2)dx_2dz_2$$
$$= h^r \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(z_{23})f(x_2, z_3 + Hz_{23})dx_2dz_{23},$$
$$P_{n2}(W_2, W_3)$$
$$= \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(H^{-1}(z_2 - z_3))f(x_1, z_1)dx_1dz_1$$
$$= h^r \int (|x_1 + x_2| - |x_1 - x_2|)K(z_{13})K(H^{-1}(z_2 - z_3))f(x_1, z_3 + Hz_{13})dx_1dz_{13}.$$

Therefore

$$G_n(W_1, W_2)$$
$$= E(P_{n2}(W_1, W_3)P_{n2}(W_2, W_3)|W_1, W_2)$$
$$= h^{2r} \int \left(\int (|x_1 + x_2| - |x_1 - x_2|)K(z_{13})K(H^{-1}(z_2 - z_3))f(x_1, z_3 + Hz_{13})dx_1dz_{13} \right.$$
$$\left. \cdot \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(z_{23})f(x_2, z_3 + Hz_{23})dx_2dz_{23} \right)$$
$$\cdot f(z_3)dz_3$$
$$= h^{3r} \int \left(\int (|x_1 + x_2| - |x_1 - x_2|)K(z_{13})K(z_4 + H^{-1}(z_1 - z_2))f(x_1, z_1 + H(z_{13} + z_4))dx_1dz_{13} \right.$$

$$
\cdot \int (|x_1+x_2|-|x_1-x_2|)K(z_4)K(z_{23})f(x_2,z_1+H(z_{23}+z_4))dx_2dz_{23}\Big)
$$
$$
\cdot f(z_1+Hz_4)dz_4.
$$

We can verify that $EG_n^2(W_1,W_2)=O_p(h^{7r})$ with more transformation $z_1=z_2+Hz_{21}$ in the integral. Furthermore $EG_n^2(W_1,W_2)=O_p(h^{7r})$.

Therefore, under the conditions $nh^r \longrightarrow \infty$ and $h^r \longrightarrow 0$, we obtain that

$$
\frac{EG_n^2(W_1,W_2))+n^{-1}E\mathcal{P}_{n2}^4(W_1,W_2)}{(E\mathcal{P}_{n2}^2(W_1,W_2))^2}=\frac{O_p(h^{7r})+\frac{1}{n}O_p(h^{5r})}{O_p(h^{6r})}=O_p(h^r)+O_p(\frac{1}{nh^r})\longrightarrow 0.
$$

According to Lemma B.4 in Fan and Li [13], it follows that

$$
\frac{nh^{2r}}{\sqrt{18}\sigma_{n2}}\mathcal{U}_n \xrightarrow[n\longrightarrow\infty]{\mathcal{D}} N(0,1),
$$

where

$$
\sigma_{n2}^2=\frac{1}{9}\{EP_{n2}^2(W_1,W_2)+4EP_{n2}^2(W_1,W_3)+2EP_{n2}(W_1,W_2)P_{n2}(W_1,W_3)\}
$$

with

$$
\begin{aligned}
&EP_{n2}^2(W_1,W_2)\\
=&h^{3r}\int\Big(\int(|x_1+x_2|-|x_1-x_2|)K(z_{31})K(z_{12}+z_{31})f(z_2)dz_{31}\Big)^2\\
&\cdot f(x_1,z_2)f(x_2,z_2)dx_1dx_2dz_{12}dz_2+O_p(h^{3r}),\\
&EP_{n2}^2(W_1,W_3)\\
=&h^{3r}\int\Big(\int(|x_1+x_2|-|x_1-x_2|)K(z_{13})f(x_2,z_3)dx_2\Big)^2\\
&\cdot f(x_1,z_3)f(z_3)dx_1dz_{13}dz_3+O_p(h^{3r})
\end{aligned}
$$

and

$$
\begin{aligned}
&EP_{n2}(W_1,W_2)P_{n2}(W_1,W_3)\\
=&h^{3r}\int K^2(z_{13})(|x_1+x_2|-|x_1-x_2|)\Big(\int(|x_1+x_2|-|x_1-x_2|)f(x_2,z_3)dx_2\Big)\\
&\cdot f(z_2)f(x_1,z_3)f(x_2,z_2)f(z_3)dx_1dx_2dz_{13}dz_2dz_3+O_p(h^{3r}).
\end{aligned}
$$

Therefore, we finally obtain that $nh^{r/2}\mathcal{U}_n \xrightarrow[n\to\infty]{D} N(0,\sigma^2)$ with

$$
\begin{aligned}
\sigma^2=&2\int\Big(\int(|x_1+x_2|-|x_1-x_2|)K(z_4)K(z_3+z_4)f(z_2)dz_4\Big)^2\\
&f(x_1,z_2)f(x_2,z_2)dx_1dx_2dz_2dz_3\\
&+8\int\Big(\int(|x_1+x_2|-|x_1-x_2|)f(x_2,z_3)dx_2\Big)^2K^2(z_4)f(x_1,z_3)f(z_3)dx_1dz_3dz_4 \quad (3.5)\\
&+4\int K^2(z_4)(|x_1+x_2|-|x_1-x_2|)\Big(\int(|x_1+x_2|-|x_1-x_2|)f(x_2,z_3)dx_2\Big)\\
&\cdot f(z_2)f(x_1,z_3)f(x_2,z_2)f(z_3)dx_1dx_2dz_2dz_3dz_4.
\end{aligned}
$$

4 Consistency of $\mathcal{U}_n$

The following result provides the consistency of $\mathcal{U}_n$.

Theorem 4.1 (Consistency) Assume that conditions (C1)–(C3) hold and the second moment of X exists, then as $n \longrightarrow \infty$, we have $\mathcal{U}_n \xrightarrow[n\to\infty]{P} \mathcal{S}_a$.

Proof We will complete this proof by two steps.

Step 1 $\mathcal{U}_n = E[\mathcal{U}_n] + o_p(1)$.

We follow the notation in (3.1) and (3.2). According to Lee [14], we have

$$
\begin{aligned}
\mathrm{Var}(\mathcal{U}_n) =& \frac{1}{h^{4r}} \Big[\frac{3C_{n-3}^2}{C_n^3} \sigma_{n1}^2 + \frac{3(n-3)}{C_n^3} \sigma_{n2}^2 + \frac{1}{C_n^3} \sigma_{n3}^2 \Big] \\
=& \frac{1}{h^{4r}} \Big[O(\frac{1}{n}) \sigma_{n1}^2 + O(\frac{1}{n^2}) \sigma_{n2}^2 + O(\frac{1}{n^3}) \sigma_{n3}^2 \Big].
\end{aligned}
$$

First, we consider σ_{n1}^2 as follows

$$
\sigma_{n1}^2 = \mathrm{Var}(\mathcal{P}_{n1}(W_1)) \le E\mathcal{P}_{n1}^2(W_1) = \frac{1}{3} E\{P_{n1}^2(W_1) + E^2[P_n(W_1, W_2, W_3)|W_3]\},
$$

where

$$
\begin{aligned}
& E P_{n1}^2(W_1) \\
=& \int \Big[\int (|x_1 + x_2| - |x_1 - x_2|) K(H^{-1}(z_1 - z_3)) K(H^{-1}(z_2 - z_3)) f(x_2, z_2) f(z_3) dx_2 dz_2 dz_3 \Big]^2 \\
& \cdot f(x_1, z_1) dx_1 dz_1 \\
=& h^{4r} \int \Big[\int (|x_1 + x_2| - |x_1 - x_2|) K(z_{31}) K(z_{32}) f(x_2, z_1 + H(z_{31} - z_{32})) f(z_1 + Hz_{31}) dx_2 dz_{31} dz_{32} \Big]^2 \\
& \cdot f(x_1, z_1) dx_1 dz_1 \\
=& O_p(h^{4r}), \\
& E^2[P_n(W_1, W_2, W_3)|W_3] \\
=& h^{4r} \int \Big[\int (|x_1 + x_2| - |x_1 - x_2|) K(z_{13}) K(z_{23}) f(x_1, z_3 + Hz_{12}) f(x_2, z_3 + Hz_{23}) dx_1 dx_2 dz_{13} dz_{23} \Big]^2 \\
& \cdot f(z_3) dz_3 \\
=& O_p(h^{4r}),
\end{aligned}
$$

which means $\sigma_{n1}^2 \le E\mathcal{P}_{n1}^2(W_1) = O_p(h^{4r})$.

Analogously to σ_{n1}^2, we can obtain that

$$
\sigma_{n2}^2 \le E\mathcal{P}_{n2}^2(W_1, W_2) = O_p(h^{3r}), \quad \sigma_{n3}^2 \le E\mathcal{P}_n^2(W_1, W_2, W_3) = O_p(h^{2r}).
$$

Therefore, we get

$$
\begin{aligned}
\mathrm{Var}(\mathcal{U}_n) =& \frac{1}{h^{4r}} \Big[O(\frac{1}{n}) \sigma_{n1}^2 + O(\frac{1}{n^2}) \sigma_{n2}^2 + O(\frac{1}{n^3}) \sigma_{n3}^2 \Big] \\
\le& O(\frac{1}{n}) \Big(1 + O_p(\frac{1}{nh^r}) + O_p(\frac{1}{(nh^r)^2}) \Big) \\
=& o_p(1).
\end{aligned}
$$

So $\mathcal{U}_n = E[\mathcal{U}_n] + o_p(1)$ by the Chebyshev's inequality.

Step 2 $E\mathcal{U}_n = E[\frac{1}{2}\varepsilon^2(X, -X|Z)f^2(Z)] + O_p(h^2)$.

Due to the definition of $\mathcal{P}_n(W_1, W_2, W_3)$, it's easy to verify that

$$E\mathcal{U}_n = \frac{1}{h^{2r}}E\mathcal{P}_n(W_1, W_2, W_3) = \frac{1}{h^{2r}}EP_n(W_1, W_2, W_3)$$
$$= \frac{1}{h^{2r}}E[(|X_1 + X_2| - |X_1 - X_2|)K_{13}K_{23}].$$

Consider $E[(|X_1 + X_2| - |X_1 - X_2|)K_{13}K_{23}]$ as follows

$$E[(|X_1 + X_2| - |X_1 - X_2|)K_{13}K_{23}]$$
$$= \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(H^{-1}(z_2 - z_3))$$
$$\cdot f(x_1, z_1)f(x_2, z_2)f(x_3, z_3)dx_1 dx_2 dx_3 dz_1 dz_2 dz_3$$
$$= \int (|x_1 + x_2| - |x_1 - x_2|)K(H^{-1}(z_1 - z_3))K(H^{-1}(z_2 - z_3))$$
$$\cdot f(z_1)f(x_1|z_1)f(z_2)f(x_2|z_2)f(z_3)dx_1 dx_2 dz_1 dz_2 dz_3$$
$$= h^{2r}\int (|x_1 + x_2| - |x_1 - x_2|)K(z_{13})K(z_{23})f(z_3 + Hz_{13})f(x_1|z_3 + Hz_{13})$$
$$\cdot f(z_3 + Hz_{23})f(x_2|z_3 + Hz_{23})f(z_3)dx_1 dx_2 dz_{13} dz_{23} dz_3$$
$$= h^{2r}\int (|x_1 + x_2| - |x_1 - x_2|)K(z_{13})K(z_{23})f(z_3)f(x_1|z_3)f(z_3)f(x_2|z_3)f(z_3)dx_1 dx_2 dz_{13} dz_{23} dz_3$$
$$+ O_p(h^{2r+2})$$
$$= h^{2r}\int (|x_1 + x_2| - |x_1 - x_2|)f(z_3)f(x_1|z_3)f(z_3)f(x_2|z_3)f(z_3)dx_1 dx_2 dz_3 + O_p(h^{2r+2})$$
$$= h^{2r}\int E[(|X_1 + X_2| - |X_1 - X_2|)|z]f^3(z)dz + O_p(h^{2r+2})$$
$$= h^{2r}E[E[(|X_1 + X_2| - |X_1 - X_2|)|Z]f^2(Z)] + O_p(h^{2r+2}).$$

Thus, we get

$$E\mathcal{U}_n = E\left[\frac{1}{2}\varepsilon^2(X, -X|Z)f^2(Z)\right] + O_p(h^2).$$

Combining the results in Step 1 and Step 2, we can finally obtain that

$$\mathcal{U}_n \xrightarrow[n\to\infty]{P} \mathcal{S}_a = E\left[\frac{1}{2}\varepsilon^2(X, -X|Z)f^2(Z)\right].$$

References

[1] Manski C F. Adaptive estimation of non-linear regression models[J]. Econ. Rev., 1984, 3(2): 145–194.

[2] Newey W K. Adaptive estimation of regression models via moment restrictions[J]. J. Econ., 1988, 38(3): 301–339.

[3] Powell J L. Symmetrically trimmed least squares estimation for Tobit models[J]. Econ.: J. Econ. Soc., 1986: 1435–1460.

[4] Newey W K. Efficient estimation of Tobit models under conditional symmetry[J]. Nonpara. Semipara. Meth. Econ. Stat., 1991: 291–336.

[5] Powell J L. Censored regression quantiles[J]. J. Econ., 1986, 32(1): 143–155.

[6] Newey W K, Powell J L. Asymmetric least squares estimation and testing[J]. Econ., 1987, 55(4): 819–847.

[7] Zheng J X. Consistent specification testing for conditional symmetry[J]. Econ. The., 1998, 14(01): 139–149.

[8] Bai J, Ng S. A consistent test for conditional symmetry in time series models[J]. J. Econ., 2001, 103(1): 225–258.

[9] Yao Q, Hyndman R J. Nonparametric estimation and symmetry tests for conditional density functions[J]. J. Nonpara. Stat., 2002, 14(3): 259–278.

[10] Su L, Jin S. A bootstrap test for conditional symmetry[J]. Ann. Econ. Fin., 2005, 6(2): 251–261.

[11] Su L. A simple test for multivariate conditional symmetry[J]. Econ. Lett., 2006, 93(3): 374–378.

[12] Székely G J. *E*-statistics: the energy of statistical samples[R]. Bowling Green State University, Dpt. Math. Stat. Tech. Rep., 2003, 3-5: 2000–2003.

[13] Fan Y, Li Q. Consistent model specification tests: omitted variables and semiparametric functional forms[J]. Econ., 1996, 64(4): 865–890.

[14] Lee J. *U*-statistics: theory and practice[M]. Boca Raton, FL: CRC Press, 1990.

条件分布对称性的一个一致性检验及渐近正态性

陈敏琼

(中山大学新华学院经济与贸易学院, 广东 广州 510520)

(中山大学数学学院, 广东 广州 510275)

摘要: 本文研究了条件分布对称性检验的问题. 借助能量距离的概念与思路, 提出了条件能量距离的概念. 基于条件能量距离构造出一个新的条件分布对称性的检验统计量, 该统计量具有带随机核的U-统计量的形式. 利用带随机核U-统计量理论证明得到该检验统计量的一致性及在原假设下的渐近正态性的结果.

关键词: 条件分布对称性; 能量距离; 随机核U-统计量; 一致性; 渐近正态性

MR(2010)主题分类号: 62G10   中图分类号: O212.7