Econometric Theory, **14**, 1998, 139–149. Printed in the United States of America.

CONSISTENT SPECIFICATION TESTING FOR CONDITIONAL SYMMETRY

JOHN XU ZHENG
University of Texas

This paper presents a consistent specification test of conditional symmetry using a kernel method. The test statistic is shown to be asymptotically distributed as standard normal under the null hypothesis of conditional symmetry and consistent against any conditional asymmetric distribution. Power against local alternatives is also investigated. A Monte Carlo simulation is provided to evaluate the finite-sample performance of the test.

1. INTRODUCTION

In many econometric models, for the purpose of identification, some distributional assumptions are placed on the relationship between two random vectors. Three often imposed assumptions are conditional moment restrictions, independence, and conditional symmetry (see Newey, 1990). Conditional symmetry implies that the distribution of one random variable, typically the error term, given another random vector, usually the independent variables, has a symmetric form. Though conditional symmetry imposes a strong shape restriction, it allows for heteroskedasticity. There are a few semiparametric estimators proposed under conditional symmetry. Manski (1984) and Newey (1988) estimated regression models under conditional symmetry. Powell (1986a) and Newey (1991) proposed semiparametric estimation for Tobit models under conditional symmetry.

Tests for the conditional symmetry assumption were proposed by Powell (1986b) for censored regression models and by Newey and Powell (1987) for linear regression models via asymmetric least-squares estimation. However these tests are unlikely to be consistent against all conditional asymmetric distributions. The closely related problem of testing for (unconditional) symmetry was investigated by Gupta (1967), Gastwirth (1971), Doksum, Fenstad, and Aaberge (1977), and Randles, Fligner, Policello, and Wolfe (1980). It is not clear whether these tests can be extended to testing for conditional symmetry.

This paper provides a consistent test of conditional symmetry using a kernel estimation method. Here, the null hypothesis imposes only a shape restriction,

I thank the coeditor, Donald W.K. Andrews, for useful comments and editorial assistance and also two anonymous referees for helpful comments. Suggestions from one of the referees led to a significant improvement and extension of the original test. Address correspondence to: John Xu Zheng, Department of Economics, University of Texas at Austin, Austin, TX 78712, USA.

© 1998 Cambridge University Press 0266-4666/98 $9.50

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

not a parametric model as in Zheng (1994, 1996a, 1996b, 1998). The techniques extend those in Zheng (1996a), which tested for functional form in regression models. The paper is organized as follows. Section 2 presents the test statistic. In Section 3, we show that the test statistic is asymptotically standard normal under the null hypothesis of conditional symmetry and consistent against any asymmetric distribution. The test is also shown to be consistent against any local alternatives that approach the null at rates arbitrarily slower than the parametric rate $n^{-1/2}$. Section 4 provides a Monte Carlo study on the finite-sample performance of the test.

2. THE HYPOTHESIS AND TEST STATISTIC

Let $F(y,x)$ be the joint probability distribution function of a random vector (Y,X) defined on some probability space $(\Omega,\mathcal{F},P)$, where (Y,X) takes values on $R \times R^m$. We have a random sample $\{(y_1,x_1),(y_2,x_2),\ldots,(y_n,x_n)\}$ from the probability distribution $F(y,x)$. Denote the conditional distribution function of Y given X and the density function of X by $F(y|x)$ and $p(x)$, respectively.

In applications, the conditional distribution function $F(t|x)$, as a function of t, is assumed to be symmetric about a common center for almost every x. We may normalize the location of the distribution so that the center of symmetry is zero. The null hypothesis to be tested is that the conditional distribution is symmetric,

$$H_0\text{: } \Pr[F(y_i|x_i) = 1 - F(-y_i|x_i)] = 1, \tag{2.1}$$

whereas the alternative to be tested is that the null is false:

$$H_1\text{: } \Pr[F(y_i|x_i) = 1 - F(-y_i|x_i)] < 1. \tag{2.2}$$

Thus the alternative consists of all possible conditionally asymmetric distributions.

Denote $1(\cdot)$ as the indicator function, i.e., $1(A) = 1$ if event A occurs and $1(A) = 0$ otherwise. For any $t \in R$, denote $\epsilon_i(t) \equiv 1(y_i \geq t) - 1(y_i \leq -t)$. Note that H_0 holds if and only if $E[\epsilon_i(t)|x_i] = 1 - F(t|x_i) - F(-t|x_i) = 0$ for all t with probability one. Denote $F_1(t)$ as the marginal distribution function of y_i. Then we have

$$\begin{aligned} W &\equiv \int E\{\epsilon_i(t)E[\epsilon_i(t)|x_i]p(x_i)\}\,dF_1(t) \\ &= \int E\{[1 - F(t|x_i) - F(-t|x_i)]^2 p(x_i)\}\,dF_1(t) \geq 0, \end{aligned} \tag{2.3}$$

where the equality holds if and only if H_0 is true. Thus the sample analog of W can be used to test the null. As in Powell, Stock, and Stoker (1989), weighting W by the density function p avoids the problem of trimming the small values of the density function.

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

In practice, y_i is usually a regression residual $g(w_i, x_i, \theta_0)$ that depends on a dependent variable w_i, some regressors x_i, and some unknown parameters $\theta_0 \in \Theta \subset R^q$. Denote the estimated y_i by $y_i(\hat{\theta}) \equiv g(w_i, x_i, \hat{\theta})$ where $\hat{\theta}$ is an estimator of θ_0. We estimate the marginal distribution function of y_i by the empirical distribution function of $y_i(\hat{\theta})$, $F_{1n}(t|\hat{\theta}) = (1/n)\sum_{i=1}^{n} 1[y_i(\hat{\theta}) \le t]$, and $\epsilon_i(t)$ by $\epsilon_i(t, \hat{\theta}) = 1[y_i(\hat{\theta}) \ge t] - 1[y_i(\hat{\theta}) \le -t]$.

The density weighted conditional expectation $p(x_i)E[\epsilon_i(y)|x_i]$ can be estimated by the delete-one kernel regression method:

$$\hat{p}(x_i)\hat{E}(\epsilon_i(t)|x_i) = \frac{1}{n-1}\sum_{j \neq i} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right)\epsilon_j(t, \hat{\theta}), \tag{2.4}$$

where K is a kernel function and h is a bandwidth.

We then obtain the sample analog of W:

$$W_n \equiv \frac{1}{n(n-1)}\sum_{i \neq j} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right)\int \epsilon_i(t, \hat{\theta})\epsilon_j(t, \hat{\theta})\, dF_{1n}(t|\hat{\theta}), \tag{2.5}$$

where $\sum_{i \neq j}$ sums over all pairs of $(i, j)(i, j = 1, \ldots, n)$ except $i \neq j$.

Denote $F(t|x_i, \theta)$ as the conditional distribution function of $y_i(\theta) \equiv g(w_i, x_i, \theta)$ given x_i and θ and $F_1(t|\theta)$ as the distribution function of $y_i(\theta)$ given θ. For any matrix A, let $\|A\|$ denote its Euclidean norm, i.e., $\|A\| = [\mathrm{tr}(AA')]^{1/2}$. We impose the following regularity assumptions.

Assumption 1. $\{(w_i, x_i)\}_{i=1}^{n}$ is a random sample from a probability distribution $G(w, x)$ on $R \times R^m$. Here $y_i(\theta) \equiv g(w_i, x_i, \theta)$, where $g(\cdot, \cdot, \theta) : R \times R^m \to R$ for every $\theta \in \Theta \subset R^q$. The term $y_i(\theta)$ is twice differentiable with respect to θ, $\sup_{\theta \in \Theta}\|\partial y_i(\theta)/\partial\theta\|^2 \le b(w_i, x_i)$, and $\sup_{\theta \in \Theta}\|\partial^2 y_i(\theta)/\partial\theta\partial\theta'\| \le b(w_i, x_i)$, where $E[b(w_i, x_i)] < \infty$. Here $F(y|x, \theta)$ and $F_1(y|\theta)$ have uniformly bounded first-order and second-order derivatives with respect to y and θ. The density function $p(x)$ of x_i and its first-order derivatives are uniformly bounded. The term $y_i = g(w_i, x_i, \theta_0)$ for an interior point θ_0 of Θ. There exists an estimator $\hat{\theta}$ of θ_0 such that $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$.

Assumption 2. The term $K(u)$ is a nonnegative, bounded, continuous, and symmetric function on R^m such that $\int K(u)\, du = 1$ and $\int \|u\|^2 K(u)\, du < \infty$.

The conditions on $y(\theta)$ in Assumption 1 ensure that its derivatives are appropriately dominated by functions integrable with respect to the distribution $G(w, x)$. For example, in the linear regression model $w_i = x_i'\theta + y_i$, the conditions require that $E(\|x_i\|^2) < \infty$.

For notational simplicity, we let $F(y|x), F(y)$, and y denote $F(y|x, \theta_0), F(y|\theta_0)$, and $y(\theta_0)$, respectively.

3. THE LIMITING NULL DISTRIBUTION AND POWER

The following theorem gives the asymptotic null distribution of the statistic W_n.

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

THEOREM 1. *Given Assumptions* 1 *and* 2, *if* $h \to 0$ *and* $nh^m \to \infty$, *then under the null hypothesis* (2.1),

$$nh^{m/2} W_n \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{3.1}$$

where

$$\sigma^2 = 8 \int K^2(u)\, du \cdot \iint E\{[1 - F(s \vee t|x_i)]^2 p(x_i)\}\, dF_1(t)\, dF_1(s). \tag{3.2}$$

Here $s \vee t \equiv \max\{s, t\}$. *The term* σ^2 *can be consistently estimated by*

$$\hat{\sigma}^2 \equiv \frac{8}{n(n-1)} \sum_{i \neq j} \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) \times \left[\int 1[y_i(\hat{\theta}) \geq t] 1[y_j(\hat{\theta}) \geq t]\, dF_{1n}(t|\hat{\theta})\right]^2. \tag{3.3}$$

The bandwidth condition is the same as one imposed by Prakasa Rao (1983) for obtaining the consistency in quadratic mean of kernel density estimators. The final test statistic T_n is based on W_n:

$$T_n \equiv nh^{m/2} W_n / \hat{\sigma}. \tag{3.4}$$

By Theorem 1, $T_n \to_d \mathcal{N}(0,1)$ under the null.

The following result provides the asymptotic behavior of T_n under the alternative.

THEOREM 2. *Given Assumptions* 1 *and* 2, *if* $h \to 0$ *and* $nh^m \to \infty$, *then under the alternative hypothesis* (2.2),

$$T_n / nh^{m/2} \xrightarrow{p} \int E\{[1 - F(t|x_i) - F(-t|x_i)]^2 p(x_i)\}\, dF_1(t)/\sigma > 0. \tag{3.5}$$

Thus $T_n \to_p \infty$, and the test is consistent.

We next examine the local power of the test. We consider the following sequence of local alternatives:

$$H_{1n}: 1 - F(y|x) - F(-y|x) = d_n l(y, x), \tag{3.6}$$

where $l(y, x)$ is a nonrandom function that has uniformly bounded first-order derivatives with respect to y and x and $\{d_n : n \geq 1\}$ is a sequence of scalar constants for which $\lim_{n\to\infty} d_n = 0$.

THEOREM 3. *Given Assumptions* 1 *and* 2, *if* $h \to 0$, $nh^m \to \infty$, *and* $d_n = n^{-1/2} h^{-m/4}$, *then under the local alternatives* (3.6),

$$nh^{m/2} T_n \xrightarrow{d} \mathcal{N}(\mu, 1), \tag{3.7}$$

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

where

$$\mu = E\left[\int l^2(t, x_i)\, dF_1(t) p(x_i)\right] \Big/ \sigma.$$

Thus, with proper choice of h, the test is consistent against any local alternatives approaching the null at rates arbitrarily slower than the parametric rate $n^{-1/2}$.

4. A MONTE CARLO SIMULATION

In this section, we conduct a Monte Carlo simulation of the size and power of the test to evaluate the finite-sample performance of the test.

The data are generated as follows. Let x and ϵ be two independent random variables. The term x is drawn from the standard normal distribution. A dependent variable w is generated as $w = 1 + x + x\epsilon$. Define $y \equiv x\epsilon = w - 1 - x$. The null hypothesis H_0 that we want to test is that, conditioning on x, y is symmetric about zero. Here y is estimated by the least-squares residual $\hat{y}$ in the regression of w on constant and x.

We draw ϵ from symmetric and asymmetric distributions to yield conditionally symmetric and asymmetric distributions for y. To investigate the size of the test, we first generate ϵ from the standard normal distribution and the t-distribution with 10 degrees of freedom. We then select another six distributions, two symmetric and four asymmetric, from the generalized lambda family (GLF) proposed in Ramberg and Schmeiser (1974). The distributions in this family are defined in terms of the inverses of the cumulative distribution functions: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1 - u)^{\lambda_4}]/\lambda_2$ for $0 < u < 1$. The parameters defining the six distributions are listed in Table 1, along with the values for the

TABLE 1. Distributions used in the Monte Carlo study

Distribution	λ_1	λ_2	λ_3	λ_4	α_3	α_4
Symmetric distributions						
1			Standard normal			
2			t-distribution with 10 degrees of freedom			
3	0.00000	0.16666670	1.0000	1.0000	0.000	1.80
4	0.00000	−0.01412640	−0.0800	−0.0800	0.000	5.99
Asymmetric distributions						
5	−12.60100	−0.00980045	−0.1100	−0.0001	−2.924	19.52
6	9.77260	0.01518780	−0.0010	−0.1300	3.160	23.80
7	−7.84595	−0.02236430	−0.1500	−0.0010	−3.478	30.24
8	6.43871	0.00317949	−0.0001	−0.1700	3.880	40.70

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

associated skewness parameter α_3 and kurtosis parameter α_4. We transform all distributions from this family to have mean equal to 0 and variance equal to 1.

The critical issue in the simulation is how to select the bandwidth h. Here we choose the bandwidth h to minimize the generalized cross-validation (GCV) function

$$\mathrm{GCV}(h) = \sum_{i=1}^{n} [\hat{y} - \hat{g}_h(x_i)]^2/[n(1 - \mathrm{tr}(H)/n)^2], \tag{4.1}$$

where $\hat{g}_h(\cdot)$ is a kernel estimator of the regression function $E[y_i|x_i]$

$$\hat{g}_h(x) = \sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)\hat{y}_j \Big/ \sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right) \tag{4.2}$$

and $\mathrm{tr}(H)$ is the trace of the matrix $H = (h_{i,j})_{n\times n}$ with

$$h_{i,j} = K\left(\frac{x_i - x_j}{h}\right) \Big/ \sum_{l=1}^{n} K\left(\frac{x_i - x_l}{h}\right). \tag{4.3}$$

In the regression case, minimizing GCV will yield an asymptotically optimal bandwidth that is proportional to $n^{-1/5}$ (see Härdle, 1990). So we choose the bandwidth h to be $cn^{-1/5}$ where c is a constant. We use the grid search method to find the optimal c that minimizes GCV. We let $0.05 \le c \le 5.0$ and the grid points be 0.05, 0.15,…,4.90, and 5.0 with increment 0.1.

The kernel function $K(\cdot)$ is chosen to be the Gaussian kernel

$$K(u) = (2\pi)^{-1/2}\exp(-u^2/2). \tag{4.4}$$

TABLE 2. Proportion of rejections for distributions 1–8

Symmetric	n	5%	10%	Asymmetric	n	5%	10%
1	100	5.2	8.6	5	100	14.9	22.9
	200	5.7	10.1		200	32.0	43.9
	300	7.0	9.3		300	47.4	61.6
2	100	6.9	11.4	6	100	16.8	25.7
	200	8.9	13.4		200	36.3	47.9
	300	7.0	10.3		300	54.1	64.5
3	100	3.4	6.9	7	100	18.4	28.6
	200	3.9	6.1		200	40.0	51.3
	300	2.9	5.5		300	56.0	69.2
4	100	6.8	10.1	8	100	21.5	30.3
	200	7.4	13.1		200	43.6	54.1
	300	7.8	12.5		300	61.0	72.2

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

The simulation is conducted with sample sizes 100, 200, and 300. Each experiment is based on 1,000 replications. The critical values for the test are from the standard normal distribution. Let $N_{1-\alpha/2}$ be the $(1 - \alpha/2)$th percentile from the standard normal distribution. Then at $(1 - \alpha)$ level, the null is rejected if $|T_n| > N_{1-\alpha/2}$. We let α be .05 and .10.

The results of the size study are shown in the left panel of Table 2. The actual sizes are reasonably close to their asymptotic sizes. The results of the power study are presented in the right panel of the same table. The power always increases with sample size, which is in line with the asymptotic theory.

REFERENCES

Doksum, K.A., G. Fenstad, & R. Aaberge (1977) Plots and tests for symmetry. *Biometrika* 64, 473–487.

Gastwirth, J.L. (1971) On the sign test for symmetry. *Journal of the American Statistical Association* 66, 821–823.

Gupta, M.K. (1967) An asymptotically nonparametric test of symmetry. *Annals of Mathematical Statistics* 38, 849–866.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.

Härdle, W. (1990) *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

Manski, C. (1984) Adaptive estimation of nonlinear regression models. *Econometric Reviews* 3, 145–194.

Newey, W.K. (1988) Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics* 38, 301–339.

Newey, W.K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.

Newey, W.K. (1991) Efficient estimation of Tobit models under conditional symmetry. In W.A. Barnett, J.L. Powell, & G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 291–336. Cambridge: Cambridge University Press.

Newey, W.K. & J.L. Powell (1987) Asymmetric least squares estimation and testing. *Econometrica* 55, 819–847.

Powell, J.L. (1986a) Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* 54, 1435–1460.

Powell, J.L. (1986b) Censored regression quantiles. *Journal of Econometrics* 32, 143–155.

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.

Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation.* Orlando: Academic Press.

Ramberg, J.S. & B.W. Schmeiser (1974) An approximate method for generating asymmetric random variables. *Communications of the Association for Computing Machinery* 17, 78–82.

Randles, R.H., M.A. Fligner, G.E. Policello, & D.A. Wolfe (1980) An asymptotically distribution-free test for symmetry versus asymmetry. *Journal of the American Statistical Association* 75, 168–172.

Sherman, R.P. (1994) Maximal inequalities for degenerate *U*-processes with applications to optimization estimators. *Annals of Statistics* 22, 439–459.

Zheng, J.X. (1994) A Specification Test of Conditional Parametric Distributions Using Kernel Estimation Methods. Manuscript, University of Texas at Austin.

Zheng, J.X. (1996a) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.

Zheng, J.X. (1996b) A Consistent Nonparametric Test of Heteroskedasticity. Manuscript, University of Texas at Austin.

Zheng, J.X. (1998) A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory*, 14, 123–138.

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

APPENDIX: PROOFS

We will apply U-statistic theories to obtain the asymptotic distributions of the statistic W_n. W_n can be approximated by a U-statistic of order 2 which is of the form

$$U_n \equiv \frac{1}{n(n-1)} \sum_{i \neq j} H_n(z_i, z_j), \tag{A.1}$$

where $\{z_i\}_{i=1}^n$ is an i.i.d. random sample and H_n is a symmetric function, that is, $H_n(z_i, z_j) = H_n(z_j, z_i)$. Define $r_n(z_i) = E[H_n(z_i, z_j)|z_i]$ and $\bar{r}_n = E[H_n(z_i, z_j)]$. The first part of the following lemma follows from Lemma 3.1 of Powell et al. (1989), and the second part follows from the first part.

LEMMA 1. *If* $E[\|H_n(z_i, z_j)\|^2] = o(n)$, *then*

$$n^{1/2}[U_n - \bar{r}_n] = 2n^{-1/2} \sum_{i=1}^{n} [r_n(z_i) - \bar{r}_n] + o_p(1) \quad \textit{and} \quad U_n = \bar{r}_n + o_p(1).$$

The following lemma shows that W_n can be approximated by a second-order U-statistic. Its proof is available from the author upon request. The proof uses results from Sherman (1994).

LEMMA 2. *Under Assumptions* 1 *and* 2, *if* $h \to 0$ *and* $nh^m \to \infty$, *then under the null hypothesis* (2.1), $W_n = V_n + O_p(n^{-1})$, *where*

$$V_n \equiv \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \times \int [1(y_i \geq t) - 1(y_i \leq -t)][1(y_j \geq t) - 1(y_j \leq -t)]\, dF_1(t).$$

Proof of Theorem 1. By Lemma 2, we only need to find the asymptotic distribution for V_n. The term V_n is a U-statistic with kernel

$$H_n(z_i, z_j) = \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \int \epsilon_i(t)\epsilon_j(t)\, dF_1(t), \tag{A.2}$$

where $z_i = (y_i, x_i')'$. Because $E[\epsilon_i(t)|x_i] = 0$, V_n is degenerate.

We apply Theorem 1 of Hall (1984). First we verify its condition. Noting that $|\int \epsilon_i(t)\epsilon_j(t)\, dF_1(t) \leq 4 \int dF_1(t) = 4$ and

$$E[\epsilon_i(s)\epsilon_i(t)|x_i] = 2[1 - F(s \vee t|x_i)], \tag{A.3}$$

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

under the null for any s,t, we have

$$
\begin{aligned}
E[G_n^2(z_1,z_2)] &= E\{E[H_n(z_3,z_1)H_n(z_3,z_2)|z_1,z_2]\}^2 \\
&\le \frac{4^4}{h^{4m}}\,E\left\{E\left[K\left(\frac{x_3-x_1}{h}\right)K\left(\frac{x_3-x_2}{h}\right)\middle| z_1,z_2\right]\right\}^2 \\
&= \frac{4^4}{h^{4m}} \\
&\quad \times \iint\left[\int K(u)K(u+v)h^m p(x_1+hu)\,du\right]^2 p(x_1)p(x_1-hv)h^m\,dx_1\,dv \\
&= O(1/h^m), \qquad \textbf{(A.4)}
\end{aligned}
$$

$$
\begin{aligned}
E[H_n^2(z_1,z_2)] &= \frac{1}{h^{2m}}\,E\left[K^2\left(\frac{x_1-x_2}{h}\right)\iint \epsilon_1(s)\epsilon_1(t)\epsilon_2(s)\epsilon_2(t)\,dF_1(s)\,dF_1(t)\right] \\
&= \frac{1}{h^{2m}}\iint E\left\{K^2\left(\frac{x_1-x_2}{h}\right)\times 2[1-F(s\vee t|x_1)]\right. \\
&\qquad \left.\times\, 2[1-F(s\vee t|x_2)]\right\} dF_1(s)\,dF_1(t) \\
&= \frac{1}{h^{2m}}\iint\left\{\iint K^2(u)\times 2[1-F(s\vee t|x_1)]\times 2[1-F(s\vee t|x_1-hu)]\right. \\
&\qquad \left.\times\, p(x_1)p(x_1-hu)h^m\,du\,dx_1\right\} dF_1(s)\,dF_1(t) \\
&= \frac{4}{h^m}\int K^2(u)\,du \times \iint E\{[1-F(s\vee t|x_i)]^2 p(x_i)\}\,dF_1(s)\,dF_1(t) \\
&\quad + o\left(\frac{1}{h^m}\right) \\
&= \sigma^2/(2h^m) + o(1/h^m) = O(1/h^m), \qquad \textbf{(A.5)}
\end{aligned}
$$

$$
\begin{aligned}
E[H_n^4(z_1,z_2)] &\le \frac{4^4}{h^{4m}}\,E\left[K^4\left(\frac{x_1-x_2}{h}\right)\right] \\
&= \frac{4^4}{h^{4m}}\iint K^4(u)p(x_1)p(x_1-hu)h^m\,dx_1\,du \\
&= O(1/h^{3m}). \qquad \textbf{(A.6)}
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
\frac{E[G_n^2(z_1,z_2)] + n^{-1}E[H_n^4(z_1,z_2)]}{\{E[H_n^2(z_1,z_2)]\}^2} &\le \frac{O(1/h^m) + n^{-1}O(1/h^{3m})}{O(1/h^{2m})} \\
&= O(h^m) + O(1/nh^m) \to 0, \quad \text{as } n\to\infty. \qquad \textbf{(A.7)}
\end{aligned}
$$

The condition of Theorem 1 of Hall (1984) is satisfied, and it follows from (A.5) that $V_n \to_d \mathcal{N}(0,\sigma^2)$.

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

Next we show that $\hat{\sigma}^2$ is a consistent estimator of σ^2. Denote $\hat{\sigma}_1^2$ as the U-statistic with kernel $H_n(z_i, z_j) = (8/h^m)K^2[(x_i - x_j)/h][\int 1(y_i \geq t)1(y_j \geq t)\, dF_1(t)]^2$. Similarly to Lemma 2, we can show that $\hat{\sigma}^2 = \hat{\sigma}_1^2 + o_p(1)$. Similarly to (A.5) we can show that $E[\|H_n(z_i, z_j)\|^2] = O(1/h^m) = o(n)$. Thus the condition of Lemma 1 is satisfied. We have

$$
\begin{aligned}
E[H_n(z_i, z_j)] &= \frac{8}{h^m} E\left\{K^2\left(\frac{x_i - x_j}{h}\right)\right. \\
&\qquad \left. \times \iint E[1(y_i \geq s \vee t)|x_i]E[1(y_j \geq s \vee t)|x_j]\, dF_1(s)\, dF_1(t)\right\} \\
&= \frac{8}{h^m} \iint \left\{ \iint K^2(u) \times [1 - F(s \vee t|x_i)] \times [1 - F(s \vee t|x_i - hu)] \right. \\
&\qquad \left. \times p(x_i)p(x_i - hu)h^m\, du\, dx_i \right\} dF_1(s)\, dF_1(t) \\
&= 8 \int K^2(u)\, du \times \iint E\{[1 - F(s \vee t|x_i)]^2 p(x_i)\}\, dF_1(s)\, dF_1(t) + o(1) \\
&= \sigma^2 + o(1). \qquad \textbf{(A.8)}
\end{aligned}
$$

Therefore by Lemma 1, $\hat{\sigma}^2 = \hat{\sigma}_1^2 + o_p(1) = E[H_n(z_i, z_j)] + o_p(1) = \sigma^2 + o_p(1)$. ■

Proof of Theorem 2. Again similarly to Lemma 2, we can show that $W_n = V_n + o_p(1)$. The condition of Lemma 1 can be shown to be satisfied. Because $E[\epsilon_i(t)|x_i] = 1 - F(t|x_i) - F(-t|x_i)$ under the alternative, we have

$$
\begin{aligned}
E[H_n(z_i, z_j)] &= \frac{1}{h^m} \iiint K\left(\frac{x_i - x_j}{h}\right)[1 - F(t|x_i) - F(-t|x_i)] \\
&\qquad \times [1 - F(t|x_j) - F(-t|x_j)]p(x_i)p(x_j)\, dx_i\, dx_j\, dF_1(t) \\
&= \frac{1}{h^m} \iiint K(u)[1 - F(t|x_i) - F(-t|x_i)] \\
&\qquad \times [1 - F(t|x_i - hu) - F(-t|x_i - hu)] \\
&\qquad \times p(x_i)p(x_i - hu)\, dx_i h^m\, du\, dF_1(t) \\
&= \int E\{[1 - F(t|x_i) - F(-t|x_i)]^2 p(x_i)\}\, dF_1(t) + o(1). \qquad \textbf{(A.9)}
\end{aligned}
$$

Thus by Lemma 1 we have

$$
V_n = \int E\{[1 - F(t|x_i) - F(-t|x_i)]^2 p(x_i)\}\, dF_1(t) + o_p(1). \qquad \textbf{(A.10)}
$$

■

Proof of Theorem 3. Similarly to Lemma 2, we can show that $W_n = V_n + O_p(n^{-1})$. Note that under the local alternatives, $E[\epsilon_i(t)|x_i] = d_n l(t, x_i)$. Here V_n can be decomposed as

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press

$$V_n = \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h} \right) \int [\epsilon_i(t) - d_n l(t,x_i)][\epsilon_j(t) - d_n l(t,x_j)]\, dF_1(t) \right\}$$
$$+ 2d_n \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h} \right) \int [\epsilon_i(t) - d_n l(t,x_i)] l(t,x_j)\, dF_1(t) \right\}$$
$$+ d_n^2 \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h} \right) \int l(t,x_i) l(t,x_j)\, dF_1(t) \right\}$$
$$\equiv V_{1n} + 2d_n V_{2n} + d_n^2 V_{3n}. \tag{A.11}$$

Similarly to the proof of Theorem 1, we can show that $nh^{m/2} V_{1n} \to_d \mathcal{N}(0, \sigma^2)$. Applying Lemma 1 we can show that $V_{2n} = O_p(n^{-1/2})$ and

$$V_{3n} \to_p E\left[\int l^2(t,x_i)\, dF_1(t) p(x_i) \right]. \tag{A.12}$$

Thus, if $d_n = n^{-1/2} h^{-m/4}$, then

$$nh^{m/2} V_n \to_d \mathcal{N}\left(E\left[\int l^2(t,x_i)\, dF_1(t) p(x_i) \right], \sigma^2 \right), \tag{A.13}$$

and $T_n \to \mathcal{N}(\mu, 1)$, where $\mu = E[\int l^2(t,x_i)\, dF_1(t) p(x_i)]/\sigma$. ■

https://doi.org/10.1017/S0266466698141063 Published online by Cambridge University Press