Econometric Theory, **16**, 2000, 667–691. Printed in the United States of America.

A CONSISTENT TEST OF CONDITIONAL PARAMETRIC DISTRIBUTIONS

JOHN XU ZHENG
University of Texas at Austin

This paper proposes a new nonparametric test for conditional parametric distribution functions based on the first-order linear expansion of the Kullback–Leibler information function and the kernel estimation of the underlying distributions. The test statistic is shown to be asymptotically distributed standard normal under the null hypothesis that the parametric distribution is correctly specified, whereas asymptotically rejecting the null with probability one if the parametric distribution is misspecified. The test is also shown to have power against any local alternatives approaching the null at rates slower than the parametric rate $n^{-1/2}$. The finite sample performance of the test is evaluated via a Monte Carlo simulation.

1. INTRODUCTION

The work on testing a parametric distribution dates as early as Pearson's chi-square test, Kolmogorov–Smirnov test, and Cramér–von Mises test (see Darlin, 1957). The Kolmogorov–Smirnov test and the Cramér–von Mises test share similar asymptotic properties in terms of their null distribution, local power, and consistency. The latter two tests have very complicated null distributions, and all three tests are only designed for testing unconditional distribution functions.

Andrews (1988a, 1988b) extended Pearson's chi-square test to testing for parametric conditional distribution assumptions underlying many econometric models. Andrews (1997) also extended the Kolmogorov–Smirnov test to the conditional distribution case. His test is consistent against all possible misspecifications of distributions. The critical values for his test can be obtained by a parametric bootstrap procedure. Zheng (1994) and Stinchcombe and White (1998) provided nonparametric tests of conditional distribution functions, which are also consistent against all possible misspecifications.

This paper proposes a new test for conditional parametric distributions that is based on the classical Kullback–Leibler (1951) information criterion. The information function has been used successfully in testing a parametric model against another parametric model. Testing a parametric model against a non-

I thank co-editor Donald W.K. Andrews and two anonymous referees for helpful comments. Address correspondence to: John Xu Zheng, Department of Economics, University of Texas at Austin, Austin, TX 78712, USA.

© 2000 Cambridge University Press 0266-4666/00 $9.50

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

parametric model requires estimating the information function nonparametrically. However, the general distributional properties of the nonparametric information function are unknown, except the consistency property established by Ahmad and Lin (1976). The difficulty in finding the distribution may be due to the fact that the nonparametric density estimator enters the information function nonlinearly. To overcome the problem, in this paper we linearize the information function by taking its first-order expansion and then obtain its asymptotic distribution via *U*-statistic theory. A test can then be based on the nonparametric estimator of the linearized information function.

The plan of the paper is as follows. Section 2 states the testing problem and provides the motivation for the test statistic. Section 3 shows that the test statistic is asymptotically standard normal under the null. It also establishes the consistency and local power properties of the test. Section 4 provides a Monte Carlo simulation of the finite sample performance of the test. Section 5 summarizes the paper.

2. THE HYPOTHESIS AND TEST STATISTIC

We have observations $\{(y_i, x_i)\}_{i=1}^n$ that are assumed to be independently and identically distributed (i.i.d.) with common density function $p(y, x)$. The term y_i is a vector of dependent variables that takes values in R^l, and x_i is a vector of covariates that takes values in R^m.

In parametric econometric models, the marginal density function $p_1(x)$ of x_i is often unknown and unspecified, but the conditional density function $p(y|x)$ of y_i given x_i is assumed to belong to a parametric family of known real-valued functions $f(y|x, \theta)$ on $R^l \times R^m \times \Theta$, where $\Theta \subset R^k$. The aim of this paper is to develop a test of the parametric distribution that has asymptotic power one against all deviations from the null. Specifically, the null hypothesis to be tested is that the parametric model is correct:

$$H_0: \quad \Pr[\,p(y_i|x_i) = f(y_i|x_i, \theta_0)] = 1 \quad \text{for some } \theta_0 \in \Theta. \tag{2.1}$$

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

$$H_1: \quad \Pr[\,p(y_i|x_i) = f(y_i|x_i, \theta)] < 1 \quad \text{for all } \theta \in \Theta. \tag{2.2}$$

Thus the alternative consists of all the possible deviations from the null. A test is consistent if the asymptotic power of the test is one.

The Kullback–Leibler information function $I(p, f)$ between two conditional distribution functions $p(y|x)$ and $f(y|x, \theta_0)$ or between $p(y, x)$ and $f(y|x, \theta_0)p_1(x)$ is given by

$$I(p, f) \equiv E\left\{\log\left[\frac{p(y_i|x_i)}{f(y_i|x_i, \theta_0)}\right]\right\} = \int p(y, x)\log\left[\frac{p(y|x)}{f(y|x, \theta_0)}\right] dydx. \tag{2.3}$$

It is well known that $I(p, f) \geq 0$ where the equality holds if and only if $p(y|x) = f(y|x, \theta_0)$ almost everywhere (a.e.). This property makes it very appealing to

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

use the nonparametric estimator of the information function as a test statistic. Ahmad and Lin (1976) established the consistency of the nonparametric entropy. But the asymptotic distribution of the nonparametric entropy has not been developed. This makes it difficult to apply the information criterion to hypothesis testing despite its natural appeal. This difficulty was also encountered by Robinson (1991) in testing for independence of two random variables. The problem lies in the fact that the nonparametric density estimator enters the information function nonlinearly. To overcome the problem, we may take the first-order expansion of the information function, that is,

$$E\left\{\log\left[\frac{p(y_i|x_i)}{f(y_i|x_i,\theta_0)}\right]\right\} \approx E\left\{\frac{p(y_i|x_i)}{f(y_i|x_i,\theta_0)} - 1\right\}$$
$$= E\left\{\frac{p(y_i|x_i) - f(y_i|x_i,\theta_0)}{f(y_i|x_i,\theta_0)}\right\}. \tag{2.4}$$

Now the density function $p(y|x)$ enters the information function linearly. So it is easier to work out the asymptotic distribution of the nonparametric information. Note, however, that the asymptotic distribution of $I(p,f)$ may differ from that of its first-order expansion as a result of higher-order terms.

Weighting equation (2.4) by the density function $p_1(x)$ gives us the following function:

$$E\left[\frac{p(y_i,x_i) - f(y_i|x_i,\theta_0)p_1(x_i)}{f(y_i|x_i,\theta_0)}\right]. \tag{2.5}$$

Its sample analogue may be used as a test statistic. As in Powell, Stock, and Stoker (1989), weighting the linearized information by the density function circumvents the problem of trimming the small values of the density function $p_1(x)$ in the denominator of the conditional distribution function $p(y|x)$.

We need to verify if the linearized version (2.5) of $I(p,f)$ can indeed discriminate the null from the alternative. Note that, under either H_0 or H_1, for any $\theta \in \Theta$,

$$E\left[\frac{f(y_i|x_i,\theta)p_1(x_i) - p(y_i,x_i)}{p(y_i|x_i)}\right]$$
$$= E\left\{E\left[\frac{f(y_i|x_i,\theta) - p(y_i|x_i)}{p(y_i|x_i)}\,\middle|\,x_i\right]p_1(x_i)\right\}$$
$$= \int\left[\int f(y|x,\theta)\,dy - \int p(y|x)dy\right]p_1(x)\,dx$$
$$= \int (1-1)p_1(x)\,dx = 0. \tag{2.6}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Thus we have, under either H_0 or H_1, for any $\theta \in \Theta$,

$$
\begin{aligned}
&E\left[\frac{p(y_i,x_i)-f(y_i|x_i,\theta)p_1(x_i)}{f(y_i|x_i,\theta)}\right] \\
&\quad = E\left[\frac{p(y_i,x_i)-f(y_i|x_i,\theta)p_1(x_i)}{f(y_i|x_i,\theta)}\right] + E\left[\frac{f(y_i|x_i,\theta)p_1(x_i)-p(y_i,x_i)}{p(y_i|x_i)}\right] \\
&\quad = E\left\{\left[\sqrt{\frac{p(y_i|x_i)}{f(y_i|x_i,\theta)}}-\sqrt{\frac{f(y_i|x_i,\theta)}{p(y_i|x_i)}}\right]^2 p_1(x_i)\right\} \\
&\quad \geq 0, \qquad \textbf{(2.7)}
\end{aligned}
$$

where the equality holds for some $\theta = \theta_0$ if and only if $\sqrt{p(y_i|x_i)/f(y_i|x_i,\theta_0)} = \sqrt{f(y_i|x_i,\theta_0)/p(y_i|x_i)}$ a.e. or $p(y_i|x_i) = f(y_i|x_i,\theta_0)$ a.e., that is, H_0 holds. Therefore, the linearized version can distinguish the null from the alternative.

The unknown density functions $p(y_i,x_i)$ and $p_1(x_i)$ can be consistently estimated by the kernel density method (Silverman, 1986) as

$$\hat{p}(y_i,x_i) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h^{l+m}}K_2\left(\frac{y_i-y_j}{h}\right)K_1\left(\frac{x_i-x_j}{h}\right), \qquad \textbf{(2.8)}$$

$$\hat{p}_1(x_i) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h^{m}}K_1\left(\frac{x_i-x_j}{h}\right), \qquad \textbf{(2.9)}$$

where K_1 and K_2 are kernel functions and h is a bandwidth. The unknown parameter θ_0 can be estimated by any $\sqrt{n}$-consistent estimator, such as the quasi-maximum likelihood estimator (QMLE) of White (1982). Denote the empirical distributions of (y_i,x_i) and x_i by $P_n(y,x)$ and $F_n(x)$, respectively. Then $\hat{p}(y_i,x_i)$ can be rewritten as

$$\hat{p}(y_i,x_i) = \frac{1}{h^{l+m}}\int K_2\left(\frac{y_i-y}{h}\right)K_1\left(\frac{x_i-x}{h}\right)dP_n(y,x). \qquad \textbf{(2.10)}$$

We use the following smoothed density estimator[1] $\tilde{p}(y,x)$ to estimate $f(y_i|x_i,\theta)p(x_i)$:

$$
\begin{aligned}
\tilde{p}(y_i,x_i) &= \frac{1}{h^{l+m}}\int K_2\left(\frac{y_i-y}{h}\right)K_1\left(\frac{x_i-x}{h}\right)f(y|x,\hat{\theta})\,dydF_n(x) \\
&= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h^{m+l}}\int K_2\left(\frac{y_i-y}{h}\right)K_1\left(\frac{x_i-x_j}{h}\right)f(y|x_j,\hat{\theta})\,dy. \qquad \textbf{(2.11)}
\end{aligned}
$$

Here, $\hat{\theta}$ is an estimator of θ_0 whose properties are specified subsequently.

Substituting $\hat{p}(y_i,x_i)$ and $\tilde{p}(y_i,x_i)$ for $p(y_i,x_i)$ and $f(y_i|x_i,\theta_0)p_1(x_i)$ in (2.5), respectively, we have the sample analogue W_n of $E\big[p(y_i,x_i) - f(y_i|x_i,\theta_0)p_1(x_i)/f(y_i|x_i,\theta_0)\big]$,

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

$$W_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^{l+m}} \times \left[\frac{K_2\left(\frac{y_i - y_j}{h}\right) K_1\left(\frac{x_i - x_j}{h}\right) - \int K_2\left(\frac{y_i - y}{h}\right) K_1\left(\frac{x_i - x_j}{h}\right) f(y|x_j, \hat{\theta}) dy}{f(y_i|x_i, \hat{\theta})} \right], \tag{2.12}$$

where the negligible terms with $i = j$ are dropped to apply U-statistic theory.

Our final test statistic T_n is based on the standardized W_n:

$$T_n \equiv \frac{nh^{(l+m)/2} W_n}{\hat{\sigma}}, \tag{2.13}$$

where $\hat{\sigma}$ is the estimated standard error of W_n and is defined in Theorem 1, which follows. The test based on T_n is one-sided. One rejects the null if T_n is sufficiently large.

Without loss of generality, we assume that y_i has a bounded support, for example, $\Pi_{i=1}^{l}(0,1)$. Otherwise, a suitable transformation can be used to transform the data to have bounded support.

For any matrix A, let $\|A\|$ denote its Euclidean normal, that is, $\|A\| = [\mathrm{tr}(AA')]^{1/2}$. We make the following regularity assumptions.

Assumption A1. $\{(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)\}$ are i.i.d. with a density function $p(y, x)$. Here $p(y, x)$ and its first-order derivatives with respect to y and x are uniformly bounded. The marginal density function $p_1(x)$ of x_i and its first-order derivatives are also uniformly bounded.

Assumption A2. The parameter space Θ is a compact and convex subset of R^k. The parametric conditional density function $f(y|x, \theta)$ and its first-order derivatives with respect to y, x, and θ are uniformly bounded. Under the null, $E[\log f(y_i|x_i, \theta)]$ has a unique maximum at an interior point θ_0 of Θ, and, under the alternative, $E[\log f(y_i|x_i, \theta)]$ has a unique maximum at θ^* of Θ, where the expectations are taken with respect to the true distribution. There exists a measurable function $b(y, x)$ with $\int b(y, x) dy dx < \infty$, $\int b^2(y, x) dy dx < \infty$, such that for all θ in a neighborhood of θ_0 of Θ,

$$0 < f(y|x, \theta),$$

$$\|f(y|x, \theta)\| \leq b(y, x),$$

$$\left\| \frac{\partial f(y|x, \theta)}{\partial \theta} \right\| \leq b(y, x),$$

$$\left\| \frac{\partial^2 \log f(y|x, \theta)}{\partial \theta \partial \theta'} \right\| \leq b(y, x),$$

$$\left\| \frac{\partial \log f(y|x, \theta)}{\partial \theta} \times \frac{\partial \log f(y|x, \theta)}{\partial \theta'} \right\| \leq b(y, x).$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Assumption A3. There exists an estimator $\hat{\theta}$ of θ_0 such that $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$ under the null, whereas under the alternative, $\hat{\theta} \to_p \theta^*$ where $\theta^* \in \Theta$.

Assumption A4. $K_1(u)$ is a nonnegative, bounded, and continuous function such that $\int K_1(u) = 1$. Here $K_2(v)$ is chosen such that $K_2(v) = \Pi_{i=1}^{l}[2e^{-2v_i}/(1 - e^{-2/h})]$ if $0 \leq v_i \leq (1/h)$ $(i = 1,\ldots,l)$ where $v = (v_1,\ldots,v_l)$, $K_2(v_1,\ldots,v_l) = K_2(v_1 + (1/h),\ldots,v_l + (1/h))$ for $-(1/h) \leq v_i \leq 0$ $(i = 1,\ldots,l)$, and $K_2(v) = 0$ otherwise.

Assumption A1 requires that y_i and x_i are continuously distributed. This rules out the cases where x_i are discrete or the density function of x_i is not bounded. The assumptions in A2 are standard for ensuring the consistency and asymptotic normality of the quasi–maximum likelihood estimator of White (1982). The kernel function $K_1(\cdot)$ in A4 is a standard kernel used in the nonparametric literature. We choose $K_2(\cdot)$ so that $\int_0^1 \ldots \int_0^1 K_2[(y_i - y)/h]\,dy/h^l = 1$ for any $y_i \in (0,1)$. This ensures that the U-statistic kernel in (2.12) is a degenerate kernel (see equation (A.6) in the Appendix). Note that the kernel function depends on the sample size n through h. We can also see that $\int_0^1 \ldots \int_0^1 K_2^2[(y_i - y)/h]dy/h^l = (1 + e^{-2/h})/(1 - e^{-2/h}) \to 1$ as $n \to \infty$. Though this kernel is unusual, it does not impose any restrictions on the distribution of the data.

3. THE ASYMPTOTIC NULL DISTRIBUTION AND POWER

First we derive the asymptotic distribution of the statistic W_n under the null hypothesis that $p(y_i|x_i) = f(y_i|x_i,\theta_0)$. The asymptotic null distribution of W_n is provided in the following theorem (all proofs are provided in the Appendix).

THEOREM 1. *Given Assumptions* A1–A4, *if* $h \to 0$ *and* $nh^{l+m} \to \infty$, *then under the null hypothesis* (2.1),

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

where σ^2 *is the asymptotic variance of* $nh^{(l+m)/2}W_n$,

$$\sigma^2 = 2\int K_1^2(u)\,du \int p_1^2(x)\,dx. \tag{3.2}$$

Furthermore, σ^2 *can be consistently estimated by*

$$\hat{\sigma}^2 = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K_1^2\left(\frac{x_i - x_j}{h}\right). \tag{3.3}$$

The condition $nh^{l+m} \to \infty$ places an upper bound on the rate that the bandwidth h converges to 0. The bandwidth condition is the same for obtaining the consistency in quadratic mean of the kernel estimation of a joint density function (see Prakasa Rao, 1983). Note that any $\sqrt{n}$-consistent estimator $\hat{\theta}$ of

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

θ_0, including the QMLE, can be used in the statistic W_n; the asymptotic distribution of W_n does not depend on the limiting distribution of $\hat{\theta}$.

It follows that under the null hypothesis, $T_n \xrightarrow{d} \mathcal{N}(0,1)$. The asymptotic null distribution is similar to that for testing parametric models, mostly regression models.[2]

Next we provide the consistency and power of the test. If the null hypothesis is false, the behavior of W_n is given by the following theorem.

THEOREM 2. *Given Assumptions* A1–A4, *if* $h \to 0$ *and* $nh^{l+m} \to \infty$, *then under the alternative hypothesis* (2.2),

$$W_n \xrightarrow{p} E\left[\frac{p(y_i, x_i) - f(y_i|x_i, \theta^*)p_1(x_i)}{f(y_i|x_i, \theta^*)}\right] > 0. \tag{3.4}$$

Because $\hat{\sigma} \xrightarrow{p} \sigma$ under either the null or the alternative, $T_n \to_p +\infty$. Thus T_n is consistent against any alternative.

To investigate the power of the test against local misspecification, we consider the following sequence of local alternatives:

$$H_{1n}: \quad p(y|x) = f(y|x, \theta_0) + d_n \times l(y, x), \tag{3.5}$$

where $d_n \to 0$ as $n \to \infty$ and the function $l(\cdot, \cdot)$ is continuously differentiable and uniformly bounded. Because both $p(y|x)$ and $f(y|x, \theta_0)$ are densities, $\int l(y, x)dy = 0$ for any x. The power of our test against the local misspecification (3.5) is given in the following theorem.

THEOREM 3. *Given Assumptions* A1–A4, *if* $h \to 0$ *and* $nh^{l+m} \to \infty$, *then under the local misspecification* (3.5) *with* $d_n = n^{-1/2}h^{-(l+m)/4}$,

$$T_n \xrightarrow{d} \mathcal{N}(\delta, 1), \tag{3.5}$$

where

$$\delta = E\left[\frac{l^2(y_i, x_i)}{f^2(y_i|x_i, \theta_0)} p_1(x_i)\right] \Big/ \sigma > 0. \tag{3.7}$$

Here σ *is given in Theorem* 1.

Thus the test is consistent against local alternatives converging to the null arbitrarily close to, but slower than, the parametric rate $n^{-1/2}$. The local power result is similar to that for the smoothing-based nonparametric tests of regression models and Zheng (1994). The tests of Andrews (1997) and Stinchcombe and White (1998) have better local power properties than the test proposed in this paper—those two tests can detect $n^{-1/2}$ local misspecification.

4. A MONTE CARLO SIMULATION

In this section, a Monte Carlo simulation is conducted to evaluate the finite sample performance of the test.

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

In all of the following models, the covariate x_i is generated from the standard normal distribution, and the error term ϵ_i is independent of x_i.

In models 1a–1e, we consider testing for possible misspecifications in classical linear regression models. In model 1a, the dependent variable y is generated as $y_i = 1 + x_i + \epsilon_i$ $(i = 1,\ldots,n)$, where ϵ_i has the standard normal distribution. In model 1b, $y_i = 1 + x_i + \epsilon_i$, where ϵ_i has the standard logistic distribution. In model 1c, $y_i = 1 + x_i + \epsilon_i$, where ϵ_i has the t-distribution with five degrees of freedom. In model 1d, $y_i = 1 + x_i + x_i^2 + \epsilon_i$, where ϵ_i has the standard normal distribution. In model 1e, $y_i = 1 + x_i + x_i\epsilon_i$, where ϵ_i has the standard normal distribution. The null hypothesis is that the model is linear and homoskedastic with normal error, that is,

$$H_0: \quad p(y|x) = \phi[(y - \beta_1 - \beta_2 x)/\sigma]/\sigma,$$

where ϕ is the standard normal density function and $\theta \equiv (\beta_1, \beta_2, \sigma^2)' \in R \times R \times R_{++}$. Under the null, model 1a is true and models 1b–1e are false. Here $\theta_0 = (1,1,1)'$ and is estimated by the maximum likelihood method.

In models 1a–1e, we use the standard logistic distribution function, denoted by $G_1(y)$, to transform the data $\{y_i\}_{i=1}^n$ with unbounded support $(-\infty,\infty)$ to $\{z_i \equiv G_1(y_i)\}_{i=1}^n$, which has bounded support (0,1). Then the null hypothesis is tested via testing the correct specification of the conditional density function of z_i.

In models 2a–2e, we consider specification testing of truncated regression models, where the latent dependent variables y_i^* are generated the same as in models 1a–1e, respectively. We observe (y_i^*, x_i) if and only $y_i^* > 0$. The null hypothesis is that the observed $y_i = y_i^*$ has the conditional density function $f(y|x,\theta) = 1(y > 0)[\phi((y - \beta_1 - \beta_2 x)/\sigma)/\sigma]/\Phi((\beta_1 + \beta_2 x)/\sigma)$, where Φ is the standard normal distribution function and $\theta = (\beta_1, \beta_2, \sigma^2)'$. Thus, under the null, model 2a is true and models 2b–2e are false. Here $\theta_0 = (1,1,1)'$ and is estimated by the maximum likelihood method.

We use a similar transformation as G_1, $G_2(y) \equiv 2G_1(y) - 1$, to transform the data $\{y_i\}_{i=1}^n$ with unbounded support $(0,\infty)$ to $\{z_i \equiv G_2(y_i)\}_{i=1}^n$, which has bounded support (0,1). Then the null hypothesis is tested via testing the conditional density function of z_i.

To see whether the size and power are sensitive to the choice of bandwidth h, we let $h = c \times s_x \times n^{-1/5}$ for $c = 0.5$, 1.0, and 2.0, where s_x is the sample standard deviation of $\{x_i\}_{i=1}^n$. The kernel function K_1 is chosen to be the Epanechnikov kernel

$$K_1(u) = 0.75(1 - u^2)1(|u| \le 1), \tag{4.1}$$

and the kernel function K_2 is chosen as in Assumption A4. The simulation is conducted with sample sizes 50, 100, 200, and 300. Each experiment is based on 1,000 replications. The integral in the test statistic is calculated using the Gauss–Legendre procedure INTQUAD1 in GAUSS. We let the significance level α be .01, .05, and .10.

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

From the simulation results on models 1a and 2a in Tables 1 and 2, we can see that, at the .05 and 0.1 levels, the sizes of the test are below the asymptotic sizes and, at the .01 level, they are closer to each other. For any given choice of bandwidth, the power increases as the sample size increases, in line with the consistency property of the test. In both the linear regression models and the truncated regression models, the test has high power detecting the regression functional form misspecification or heteroskedasticity. Because the logistic distribution is closer to the normal distribution than the *t*-distribution with five degrees of freedom, in both types of models, the test has higher power against the *t*-distribution than the logistic distribution. Though the percentages of rejections are below the nominal sizes for model 2b, nevertheless, they are higher than that for model 2a, suggesting that with proper control of the size, the test can detect logistic distribution in the truncated regression models. The simula-

TABLE 1. Percentage of rejections in classical linear regression models

	$c = 0.5$			$c = 1.0$			$c = 2.0$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
			Model 1a: $y = 1 + x + \epsilon$,		$\epsilon \sim \mathcal{N}(0,1)$				
50	1.4	1.9	2.9	1.3	1.9	2.7	1.3	1.8	2.3
100	1.2	2.0	3.1	1.5	2.7	3.9	1.6	3.4	4.0
200	1.3	2.2	4.6	1.7	2.5	4.0	2.8	3.9	4.7
300	1.9	3.5	5.5	1.8	2.7	4.0	2.6	4.4	5.3
			Model 1b: $y = 1 + x + \epsilon$,		$\epsilon \sim$ Logistic				
50	4.4	6.1	7.9	3.9	5.3	6.7	4.0	5.1	6.7
100	7.2	10.2	12.1	8.2	10.5	11.6	9.3	12.1	14.0
200	12.4	15.1	17.2	12.1	14.2	16.7	18.7	22.8	25.6
300	13.0	16.5	19.5	16.3	19.1	21.4	22.8	26.9	30.0
			Model 1c: $y = 1 + x + \epsilon$,		$\epsilon \sim t(5)$				
50	6.5	11.1	14.1	8.3	13.5	17.5	9.4	15.5	20.6
100	14.1	20.1	24.8	18.1	26.6	32.1	22.0	34.5	42.0
200	31.6	40.2	46.1	43.7	55.0	59.5	55.5	66.6	74.0
300	45.6	51.9	57.6	54.2	64.2	70.5	72.9	82.0	85.6
			Model 1d: $y = 1 + x + x^2 + \epsilon$,		$\epsilon \sim \mathcal{N}(0,1)$				
50	30.6	50.2	61.8	56.3	75.6	84.0	76.9	88.1	92.0
100	79.8	89.7	92.8	93.3	96.9	97.5	97.7	98.8	98.9
200	99.6	100	100	100	100	100	100	100	100
300	100	100	100	100	100	100	100	100	100
			Model 1e: $y = 1 + x + x\epsilon$,		$\epsilon \sim \mathcal{N}(0,1)$				
50	50.1	69.0	78.6	28.4	44.3	56.8	20.1	27.5	33.8
100	93.8	95.5	95.9	81.0	88.8	91.2	50.3	62.2	69.6
200	95.1	95.3	95.3	96.3	96.6	96.7	89.8	93.1	94.0
300	97.0	97.0	97.2	97.1	97.3	97.3	97.5	97.6	97.7

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

TABLE 2. Percentage of rejections in truncated regression models

	$c = 0.5$			$c = 1.0$			$c = 2.0$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
				Model 2a: $y^* = 1 + x + \epsilon$,		$\epsilon \sim \mathcal{N}(0,1)$			
50	0.8	1.4	3.1	0.7	1.0	2.1	0.2	0.5	0.9
100	0.9	1.9	3.7	0.9	2.1	3.2	0.8	1.0	1.5
200	1.0	2.2	3.3	0.6	1.3	2.3	0.7	1.6	2.3
300	1.5	2.7	4.9	0.8	2.2	3.6	0.9	1.6	2.8
				Model 2b: $y^* = 1 + x + \epsilon$,		$\epsilon \sim$ Logistic			
50	0.7	1.9	3.4	0.5	0.9	2.0	0.9	1.6	2.7
100	1.3	2.7	4.6	1.1	1.8	2.9	0.8	1.1	2.0
200	2.1	3.4	5.6	1.2	2.1	3.7	2.2	3.1	4.5
300	2.6	4.6	6.7	2.8	5.1	7.1	2.9	4.5	6.2
				Model 2c: $y^* = 1 + x + \epsilon$,		$\epsilon \sim t(5)$			
50	1.2	2.5	4.6	1.3	2.8	4.2	1.0	3.3	6.6
100	0.8	3.2	6.5	2.6	5.9	10.2	3.4	9.6	15.7
200	2.2	7.2	12.6	4.5	11.3	18.2	10.5	21.8	30.6
300	6.4	13.3	22.6	9.9	20.4	28.8	19.3	32.4	44.2
				Model 2d: $y^* = 1 + x + x^2 + \epsilon$,		$\epsilon \sim \mathcal{N}(0,1)$			
50	7.2	19.0	29.9	15.2	38.0	54.5	46.1	72.0	83.9
100	26.8	53.0	65.9	58.0	83.6	91.2	94.1	98.5	98.8
200	81.1	94.6	97.4	99.2	99.6	99.9	100	100	100
300	97.3	99.4	99.5	100	100	100	100	100	100
				Model 2e: $y^* = 1 + x + x\epsilon$,		$\epsilon \sim \mathcal{N}(0,1)$			
50	56.4	74.3	84.3	30.7	47.7	60.4	10.2	22.0	33.2
100	96.7	98.1	98.4	86.8	94.2	96.3	42.7	62.8	72.4
200	99.4	99.5	99.6	99.0	99.0	99.1	92.1	95.5	96.1
300	99.8	99.8	99.8	99.7	99.7	99.7	98.6	99.0	99.3

tions also show that the power and size are bandwidth dependent in small samples. How to select the bandwidth via a data-driven method in this testing problem is an important issue and is left for future research.

5. CONCLUSION

This paper provides a new consistent specification test of conditional parametric distributions against nonparametric alternatives. The test is based on the first-order expansion of the well-known Kullback–Leibler information criterion function. The test is shown to have the standard normal distribution under the null and to be consistent against any fixed alternatives. The test can be applied to testing for conditional distributions underlying econometric models. Though

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

the paper only deals with i.i.d. data, it is possible to extend the test to time series cases.

NOTES

1. A similar smoothing device was also used by Härdle and Mammen (1993) and Fan (1994).
2. The related work includes Bierens (1990), Fan and Li (1996), Gozalo (1993), Härdle and Mammen (1993), Hong and White (1995), Horowitz and Härdle (1994), Lavergne and Vuong (1995), Lee (1988), Li (1994), Stinchcombe and White (1998), Whang and Andrews (1993), Wooldridge (1992), Yatchew (1992), Zheng (1996, 1998a, 1998b), and others.

REFERENCES

Ahmad, I.A. & P.E. Lin (1976) A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Transactions on Information* 22, 372–375.

Andrews, D.W.K. (1988a) Chi-square diagnostic tests for econometric models: Theory. *Econometrica* 56, 1419–1453.

Andrews, D.W.K. (1988b) Chi-square diagnostic tests for econometric models: Introduction and application. *Journal of Econometrics* 37, 135–156.

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097–1128.

Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.

Darling, D.A. (1957) The Kolmogorov–Smirnov, Cramér–von Mises tests. *Annals of Mathematical Statistics* 28, 823–838.

Fan, Y. (1994) Testing the goodness-of-fit of a parametric density function by kernel method. *Econometric Theory* 10, 316–356.

Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables, parametric and semiparametric functional forms. *Econometrica* 64, 865–890.

Gozalo, P.L. (1993) A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* 9, 451–477.

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

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.

Horowitz, J.L. & W. Härdle (1994) Testing a parametric model against a nonparametric alternative. *Econometric Theory* 10, 821–848.

Kullback, S. & R.A. Leibler (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.

Lavergne, P. & Q. Vuong (1995) Nonparametric Significance Testing. Manuscript, INRA.

Lee, B.-J. (1988) *A Nonparametric Model Specification Test Using a Kernel Regression Method*. Ph.D. Dissertation, University of Wisconsin-Madison.

Li, Q. (1994). A Consistent Test for Linearity in Partially Linear Regression Models. Manuscript, University of Guelph.

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*. New York: Academic Press.

Robinson, P.M. (1991) Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58, 437–453.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.
Whang, Y.-J. & D.W.K. Andrews (1993) Tests of specification for parametric and semiparametric models. *Journal of Econometrics* 57, 277–318.
White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.
Wooldridge, J.M. (1992) A test for functional form against nonparametric alternatives. *Econometric Theory* 8, 452–475.
Yatchew, A.J. (1992) Nonparametric regression tests based on least squares. *Econometric Theory* 8, 435–451.
Zheng, J.X. (1994) A Specification Test of Conditional Parametric Distributions using Kernel Estimation Methods. Manuscript, University of Texas at Austin.
Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.
Zheng, J.X. (1998a) A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory* 14, 123–138.
Zheng, J.X. (1998b) Consistent specification testing for conditional symmetry. *Econometric Theory* 14, 139–149.

APPENDIX: PROOFS

We first review some of the theories on nondegenerate and degenerate *U*-statistic. The general "second-order" *U*-statistic is of the form

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

where $\{z_i\}_{i=1}^n$ is an i.i.d. random sample for a random vector Z and q_n is any dimensional-symmetric function, that is, $q_n(z_i, z_j) = q_n(z_j, z_i)$. The following lemma follows from Lemma 3.1 of Powell et al. (1989) and Lemma 3.1 of Zheng (1996).

LEMMA 1. *If* $E[\|q_n(z_i, z_j)\|^2] = o(n)$, *then* $U_n = E[r_n(z_i)] + o_p(1)$.

The preceding result is useful if $E[q_n(z_i, z_j)|z_j] \neq 0$ or the *U*-statistic is non-degenerate. A *U*-statistic is degenerate if $E[q_n(z_i, z_j)|z_i] = 0$ for $i \neq j$. For a one-dimensional degenerate *U*-statistic, denote

$$G_n(z_1, z_2) \equiv E[q_n(z_3, z_1) q_n(z_3, z_2)|z_1, z_2]. \tag{A.2}$$

The next lemma follows from Hall (1984).

LEMMA 2. *Assume* $E[q_n(z_1, z_2)|z_1] = 0$ *and* $E[q_n^2(z_1, z_2)] < \infty$ *for each n. If*

$$\frac{E[G_n^2(z_1, z_2)] + n^{-1} E[q_n^4(z_1, z_2)]}{\{E[q_n^2(z_1, z_2)]\}^2} \to 0, \quad \text{as } n \to \infty, \tag{A.3}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

then

$$U_n \equiv \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq 1}^{n} q_n(z_i, z_j) \right\} \Bigg/ \left\{ \frac{2E[q_n^2(z_1, z_2)]}{n^2} \right\} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof of Theorem 1. By taking a Taylor expansion of W_n around θ_0 up to the second order, we decompose W_n into three parts:

$$\begin{aligned}
W_n = & \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h^{l+m}} K_1\left(\frac{x_i - x_j}{h}\right) \frac{K_2\left(\frac{y_i - y_j}{h}\right) - \int K_2\left(\frac{y_i - y}{h}\right) f(y|x_j, \theta_0) dy}{f(y_i|x_i, \theta_0)} \right\} \\
& - \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h^{l+m}} \frac{K_1\left(\frac{x_i - x_j}{h}\right)}{f(y_i|x_i, \theta_0)} \right. \\
& \quad \times \left[K_2\left(\frac{y_i - y_j}{h}\right) \frac{\partial \log f(y_i|x_i, \theta_0)}{\partial \theta'} \right. \\
& \qquad - \int K_2\left(\frac{y_i - y}{h}\right) f(y|x_j, \theta_0) dy \, \frac{\partial \log f(y_i|x_i, \theta_0)}{\partial \theta'} \\
& \qquad \left. \left. + \int K_2\left(\frac{y_i - y}{h}\right) \frac{\partial f(y|x_j, \theta_0)}{\partial \theta'} dy \right] \right\} (\hat{\theta} - \theta_0) \\
& - (\hat{\theta} - \theta_0)' \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h^{l+m}} \frac{K_1\left(\frac{x_i - x_j}{h}\right)}{f(y_i|x_i, \bar{\theta})} \right. \\
& \quad \times \left[K_2\left(\frac{y_i - y_j}{h}\right) \left(\frac{\partial^2 \log f(y_i|x_i, \bar{\theta})}{\partial \theta \partial \theta'} - \frac{\partial \log f(y_i|x_i, \bar{\theta})}{\partial \theta} \frac{\partial \log f(y_i|x_i, \bar{\theta})}{\partial \theta'} \right) \right. \\
& \qquad - \int K_2\left(\frac{y_i - y}{h}\right) f(y|x_j, \bar{\theta}) dy \\
& \qquad \times \left(\frac{\partial^2 \log f(y_i|x_i, \bar{\theta})}{\partial \theta \partial \theta'} - \frac{\partial \log f(y_i|x_i, \bar{\theta})}{\partial \theta} \frac{\partial \log f(y_i|x_i, \bar{\theta})}{\partial \theta'} \right) \\
& \qquad - 2 \int K_2\left(\frac{y_i - y}{h}\right) \frac{\partial f(y|x_j, \bar{\theta})}{\partial \theta} dy \, \frac{\partial \log f(y_i|x_i, \bar{\theta})}{\partial \theta'} \\
& \qquad \left. \left. + \int K_2\left(\frac{y_i - y}{h}\right) \frac{\partial^2 f(y|x_j, \bar{\theta})}{\partial \theta \partial \theta'} dy \right] \right\} (\hat{\theta} - \theta_0) \\
\equiv & \; W_{1n} - W_{2n}(\hat{\theta} - \theta_0) - (\hat{\theta} - \theta_0)' W_{3n} (\hat{\theta} - \theta_0), \qquad \textbf{(A.4)}
\end{aligned}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

where $\bar{\theta}$ lies between θ_0 and $\hat{\theta}$. Under the null, we will show that $nh^{(l+m)/2}W_{1n}$ is normally distributed, $nh^{(l+m)/2}W_{2n}(\hat{\theta}-\theta_0)=o_p(1)$, and $nh^{(l+m)/2}(\hat{\theta}-\theta_0)'W_{3n}(\hat{\theta}-\theta_0)=o_p(1)$.

First we establish the asymptotic distribution of W_{1n} under the null hypothesis. The term W_{1n} is a U-statistic with kernel

$$q_n(z_i,z_j)=\frac{1}{2h^{l+m}}K_1\left(\frac{x_i-x_j}{h}\right)\frac{K_2\left(\frac{y_i-y_j}{h}\right)-\int K_2\left(\frac{y_i-y}{h}\right)f(y|x_j,\theta_0)\,dy}{f(y_i|x_i,\theta_0)}$$

$$+\frac{1}{2h^{l+m}}K_1\left(\frac{x_j-x_i}{h}\right)\frac{K_2\left(\frac{y_j-y_i}{h}\right)-\int K_2\left(\frac{y_j-y}{h}\right)f(y|x_i,\theta_0)\,dy}{f(y_j|x_j,\theta_0)} \tag{A.5}$$

$$\equiv J_n(z_i,z_j)+J_n(z_j,z_i). \tag{A.5}$$

Under the null hypothesis, because

$$\int_0^1 K_2\left(\frac{y_i-y_j}{h}\right)dy_j=h^l\int_{y_i/h-1/h}^{y_i/h}K_2(u)\,du=h^l\int_0^{1/h}K_2(u)\,du=h^l \tag{A.6}$$

for any $0<y_i<1$, it is easy to show that $E[q_n(z_1,z_2)|z_1]=0$. Thus W_{1n} is a degenerate U-statistic.

We next verify the condition of Lemma 2. We have

$$\begin{aligned}E[G_n^2(z_i,z_j)]&=E\{E[q_n(z_3,z_1)q_n(z_3,z_2)|z_1,z_2]\}\\&=E\{E[J_n(z_3,z_1)J_n(z_3,z_2)|z_1,z_2]+E[J_n(z_3,z_1)J_n(z_2,z_3)|z_1,z_2]\\&\quad+E[J_n(z_1,z_3)J_n(z_3,z_2)|z_1,z_2]+E[J_n(z_1,z_3)J_n(z_2,z_3)|z_1,z_2]\}^2\\&\le 4E\{E[J_n(z_3,z_1)J_n(z_3,z_2)|z_1,z_2]\}^2\\&\quad+4E\{E[J_n(z_3,z_1)J_n(z_2,z_3)|z_1,z_2]\}^2\\&\quad+4E\{E[J_n(z_1,z_3)J_n(z_3,z_2)|z_1,z_2]\}^2\\&\quad+4E\{E[J_n(z_1,z_3)J_n(z_2,z_3)|z_1,z_2]\}^2.\end{aligned} \tag{A.7}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

For the first term in the preceding equation, we have

$$
\begin{aligned}
&E\{E[J_n(z_3,z_1)J_n(z_3,z_2)|z_1,z_2]\}^2 \\
&\quad = \frac{1}{2^4 h^{4l+4m}} \iiiint \\
&\qquad \times \left\{ \iint \frac{K_1\left(\frac{x_3-x_1}{h}\right)K_1\left(\frac{x_3-x_2}{h}\right)}{f^2(y_3|x_3,\theta_0)} \right. \\
&\qquad\quad \times \left[K_2\left(\frac{y_3-y_1}{h}\right)K_2\left(\frac{y_3-y_2}{h}\right) \right. \\
&\qquad\qquad - K_2\left(\frac{y_3-y_1}{h}\right)\int K_2\left(\frac{y_3-y_2}{h}\right) f(y_2|x_2,\theta_0)\,dy_2 \\
&\qquad\qquad - K_2\left(\frac{y_3-y_2}{h}\right)\int K_2\left(\frac{y_3-y_1}{h}\right) f(y_1|x_1,\theta_0)\,dy_1 \\
&\qquad\qquad \left. + \int K_2\left(\frac{y_3-y_1}{h}\right) f(y_1|x_1,\theta_0)\,dy_1 \times \int K_2\left(\frac{y_3-y_2}{h}\right) f(y_2|x_2,\theta_0)\,dy_2 \right] \\
&\qquad\quad \left. \times f(y_3|x_3,\theta_0)p_1(x_3)dy_3\,dx_3 \right\}^2 \\
&\qquad \times f(y_2|x_2,\theta_0)p_1(x_2)f(y_1|x_1,\theta_0)p_1(x_1)dy_2\,dx_2\,dy_1\,dx_1 \\
&\quad = \frac{1}{2^4 h^{4l+4m}} \iiiint \\
&\qquad \times \left\{ \iint \frac{K_1(u)K_1(u+v)}{f^2(y_1+hw|x_1+hu,\theta_0)} \right. \\
&\qquad\quad \times \left[K_2(w)K_2(w+z) - K_2(w)\int K_2(y) f(y_3-hy|x_1+hu,\theta_0)h^l dy \right. \\
&\qquad\qquad - K_2(w+z)\int K_2(y) f(y_3-hy|x_1,\theta_0)h^l dy \\
&\qquad\qquad + \int K_2(y) f(y_3-hy|x_1,\theta_0)h^l dy \\
&\qquad\qquad \left. \times \int K_2(y) f(y_3-hy|x_1+hu,\theta_0)h^l dy \right] \\
&\qquad\quad \left. \times f(y_1+hw|x_1+hu,\theta_0)p_1(x_1+hu)h^l dw h^m du \right\}^2 \\
&\qquad \times f(y_1-hz|x_1-hv,\theta_0)p_1(x_1-hv)f(y_1|x_1,\theta_0)p_1(x_1)h^l dz h^m dv dy_1\,dx_1 \\
&\quad = O(h^{-(l+m)}), \qquad \textbf{(A.8)}
\end{aligned}
$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

where we have changed variables according to $x_3 = x_1 + hu$, $y_3 = y_1 + hw$, $x_2 = x_1 - hv$, and $y_2 = y_1 - hz$.

Similarly we can show that

$$E\{E[J_n(z_3,z_1)J_n(z_2,z_3)|z_1,z_2]\}^2 = O(h^{-(l+m)}),$$

$$E\{E[J_n(z_1,z_3)J_n(z_3,z_2)|z_1,z_2]\}^2 = O(h^{-(l+m)}),$$

$$E\{E[J_n(z_1,z_3)J_n(z_1,z_3)|z_1,z_2]\}^2 = O(h^{-(l+m)}).$$

Thus we have shown that $E[G_n^2(z_1,z_2)] = O(h^{-(l+m)})$.

Next we show that $E[q_n^2(z_1,z_2)] = O(h^{-(l+m)})$.

$$\begin{aligned}
&E[J_n^2(z_1,z_2)] \\
&\quad = \frac{1}{4h^{2l+2m}} \iiiint \\
&\qquad \times \left\{ \frac{K_1^2\left(\frac{x_1-x_2}{h}\right)K_2^2\left(\frac{y_1-y_2}{h}\right)}{f^2(y_1|x_1,\theta_0)} \right. \\
&\qquad\quad - 2\frac{K_1^2\left(\frac{x_1-x_2}{h}\right)K_2\left(\frac{y_1-y_2}{h}\right)\int K_2\left(\frac{y_1-y_2}{h}\right)f(y_2|x_2,\theta_0)\,dy_2}{f^2(y_1|x_1,\theta_0)} \\
&\qquad\quad \left. + \frac{K_1^2\left(\frac{x_1-x_2}{h}\right)\left[\int K_2\left(\frac{y_1-y_2}{h}\right)f(y_2|x_2,\theta_0)\,dy_2\right]^2}{f^2(y_1|x_1,\theta_0)} \right\} \\
&\qquad \times p_1(x_1)p_1(x_2)f(y_1|x_1,\theta_0)f(y_2|x_2,\theta_0)dx_1\,dx_2\,dy_1\,dy_2 \\
&\quad = \frac{1}{4h^{2l+2m}} \iiiint_{y_1/h-1/h}^{y_1/h} \\
&\qquad \times \left\{ \frac{K_1^2(u)K_2^2(v)}{f^2(y_1|x_1,\theta_0)} - 2\frac{K_1^2(u)K_2(v)\int K_2(v)f(y_1-hv|x_1-hu,\theta_0)h^l dv}{f^2(y_1|x_1,\theta_0)} \right. \\
&\qquad\quad \left. + \frac{K_1^2(u)\left[\int K_2(v)f(y_1-hv|x_1-hu,\theta_0)h^l dv\right]^2}{f^2(y_1|x_1,\theta_0)} \right\} \\
&\qquad\quad \times p_1(x_1)p_1(x_1-hu)f(y_1|x_1,\theta_0)f(y_1-hv|x_1-hu,\theta_0)\,dx_1\,h^m du dy_1\,h^l dv
\end{aligned}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

$$= \frac{1}{4h^{l+m}} \iiiint_{y_1/h-1/h}^{y_1/h} \times \left\{ \frac{K_1^2(u)K_2^2(v)}{f^2(y_1|x_1,\theta_0)} - 2\frac{K_1^2(u)K_2(v)\int K_2(v)f(y_1|x_1,\theta_0)h^l dv}{f^2(y_1|x_1,\theta_0)} + \frac{K_1^2(u)\left[\int K_2(v)f(y_1|x_1,\theta_0)h^l dv\right]^2}{f^2(y_1|x_1,\theta_0)} \right\} \times p_1^2(x_1)f^2(y_1|x_1,\theta_0)\,dx_1\,du dy_1\,dv + o\left(\frac{1}{h^{m+l}}\right)$$

$$= \frac{1}{4h^{m+l}} \int K_1^2(u)du \int p_1^2(x)dx \int_0^1 \times \left\{ \int_{(y_1-1)/h}^{y_1/h} K_2^2(v)\,dv - h^l\left[\int_{(y_1-1)/h}^{y_1/h} K_2(v)dv\right]^2 \right\} dy_1 + o\left(\frac{1}{h^{m+l}}\right)$$

$$= \frac{1}{4h^{m+l}} \int K_1^2(u)\,du \int p_1^2(x)\,dx \left\{ \left(\frac{1+e^{-2/h}}{1-e^{-2/h}}\right)^l - h^l \right\} + o\left(\frac{1}{h^{m+l}}\right)$$

$$= \frac{1}{4h^{m+l}} \int K_1^2(u)\,du \int p_1^2(x)\,dx + o\left(\frac{1}{h^{m+l}}\right) = O\left(\frac{1}{h^{m+l}}\right), \tag{A.9}$$

where we have made similar variable changes as in (A.8).

Similarly, we can show that

$$E[J_n(z_1,z_2)J_n(z_2,z_1)] = \frac{1}{4h^{m+l}} \int K_1^2(u)\,du \int p_1^2(x)\,dx + o\left(\frac{1}{h^{m+l}}\right) = O\left(\frac{1}{h^{m+l}}\right).$$

Thus we have

$$E[q_n^2(z_1,z_2)] = E[2J_n^2(z_1,z_2) + 2J_n(z_1,z_2)J_n(z_2,z_1)]$$

$$= \frac{1}{h^{m+l}} \int K_1^2(u)\,du \int p_1^2(x)\,dx + o\left(\frac{1}{h^{m+l}}\right) = O\left(\frac{1}{h^{m+l}}\right). \tag{A.10}$$

Next we show that $E[q_n^4(z_1,z_2)] \le O(1/h^{3l+3m})$. Because

$$\{E[q_n^4(z_1,z_2)]\}^{1/4} \le \{E[J_n^4(z_1,z_2)]\}^{1/4} + \{E[J_n^4(z_2,z_1)]\}^{1/4}$$

$$= 2\{E[J_n^4(z_1,z_2)]\}^{1/4}, \tag{A.11}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

and

$$
\begin{aligned}
&E[J_n^4(z_1,z_2)]\\
&\quad\le 4E\left[\frac{1}{2^4h^{4l+4m}}\frac{K_1^4\left(\frac{x_1-x_2}{h}\right)K_2^4\left(\frac{y_1-y_2}{h}\right)}{f^4(y_1|x_1,\theta_0)}\right]\\
&\qquad+4E\left[\frac{1}{2^4h^{4l+4m}}\frac{K_1^4\left(\frac{x_1-x_2}{h}\right)\left[\int K_2\left(\frac{y_1-y_2}{h}\right)f(y_2|x_2,\theta_0)\,dy_2\right]^4}{f^4(y_1|x_1,\theta_0)}\right]\\
&\quad\le 4E\left[\frac{1}{2^4h^{4l+4m}}\frac{K_1^4\left(\frac{x_1-x_2}{h}\right)K_2^4\left(\frac{y_1-y_2}{h}\right)}{f^4(y_1|x_1,\theta_0)}\right]\\
&\qquad+4E\left[\frac{1}{2^4h^{4l+4m}}\frac{K_1^4\left(\frac{x_1-x_2}{h}\right)\left[\int K_2^4\left(\frac{y_1-y_2}{h}\right)f(y_2|x_2,\theta_0)\,dy_2\right]}{f^4(y_1|x_1,\theta_0)}\right]\\
&\quad=\frac{4}{2^4h^{4l+4m}}\int\frac{K_1^4\left(\frac{x_1-x_2}{h}\right)K_2^4\left(\frac{y_1-y_2}{h}\right)}{f^4(y_1|x_1,\theta_0)}\\
&\qquad\times f(y_1|x_1,\theta_0)p_1(x_1)f(y_2|x_2,\theta_0)p_1(x_2)\,dy_1\,dx_1\,dy_2\,dx_2\\
&\qquad+\frac{4}{2^4h^{4l+4m}}\int\frac{K_1^4\left(\frac{x_1-x_2}{h}\right)\int K_2^4\left(\frac{y_1-y_2}{h}\right)f(y_2|x_2,\theta_0)\,dy_2}{f^4(y_1|x_1,\theta_0)}\\
&\qquad\times f(y_1|x_1,\theta_0)p_1(x_1)\,dy_1\,dx_1\,dx_2\\
&\quad=\frac{4}{2^4h^{4l+4m}}\int\frac{K_1^4(u)K_2^4(v)}{f^4(y_1|x_1,\theta_0)}\\
&\qquad\times f(y_1|x_1,\theta_0)p_1(x_1)f(y_1-hv|x_1-hu,\theta_0)p_1(x_2)\,dy_1\,dx_1\,h^l\,dv\,h^m\,du\\
&\qquad+\frac{4}{2^4h^{4l+4m}}\int\frac{K_1^4(u)\int K_2^4(v)f(y_1-hv|x_1-hu,\theta_0)h^l\,dv}{f^4(y_1|x_1,\theta_0)}\\
&\qquad\times f(y_1|x_1,\theta_0)p_1(x_1)\,dy_1\,dx_1\,h^m\,du\\
&\quad=O(h^{-(3l+3m)}),
\end{aligned}
\tag{A.12}
$$

we have

$$
E[q_n^4(z_1,z_2)]\le O(h^{-(3l+3m)}). \tag{A.13}
$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Thus

$$\frac{E[G_n^2(z_1,z_2)]+n^{-1}E[q_n^4(z_1,z_2)]}{\{E[q_n^2(z_1,z_2)]\}^2} \leq \frac{O(h^{-(l+m)})+n^{-1}O(h^{-(3l+3m)})}{O(h^{-(2l+2m)})}$$

$$= O(h^{l+m})+O(1/nh^{(l+m)}) \to 0, \tag{A.14}$$

as $n \to \infty$. Thus the condition of Lemma 2 is satisfied, and it follows from (A.10),

$$nh^{(l+m)/2}W_{1n} \xrightarrow{d} \mathcal{N}\left(0, 2\int K_1^2(u)\,du \int p_1^2(x)\,dx\right). \tag{A.15}$$

Next we show that $W_{2n} = O_p(n^{-1/2})$. We decompose W_{2n} as the sum of Q_{1n} and Q_{2n} where

$$Q_{1n} \equiv \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{h^{l+m}}\frac{K_1\left(\frac{x_i-x_j}{h}\right)}{f(y_i|x_i,\theta_0)}\frac{\partial \log f(y_i|x_i,\theta_0)}{\partial \theta'}$$

$$\times\left[K_2\left(\frac{y_i-y_j}{h}\right)-\int K_2\left(\frac{y_i-y}{h}\right)f(y|x_j,\theta_0)\,dy\right]$$

$$\equiv \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} q_{1n}(z_i,z_j),$$

$$Q_{2n} \equiv \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{h^{l+m}}\frac{K_1\left(\frac{x_i-x_j}{h}\right)}{f(y_i|x_i,\theta_0)}\int K_2\left(\frac{y_i-y}{h}\right)\frac{\partial f(y|x_j,\theta_0)}{\partial \theta'}\,dy$$

$$\equiv \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} q_{2n}(z_i,z_j). \tag{A.16}$$

Here Q_{1n} is a U-statistic with kernel

$$q_n(z_i,z_j) = \tfrac{1}{2}q_{1n}(z_i,z_j)+\tfrac{1}{2}q_{1n}(z_j,z_i). \tag{A.17}$$

Similarly to (A.10), we can show that $E[\|q_n(z_i,z_j)\|^2] = O(h^{-(l+m)})$. Thus $E[\|q_n(z_i,z_j)\|^2] = O(n/nh^{(l+m)}) = o(n)$, as $nh^{l+m} \to \infty$. The condition of Lemma 1 is thus satisfied. For the preceding $q_n(z_i,z_j)$, we have

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

$$\bar{r}_n = E[q_n(z_i, z_j)] = E[q_{1n}(z_i, z_j)] = E\{E[q_{1n}(z_i, z_j)|z_i, x_j]\}$$

$$= \frac{1}{h^{l+m}} E\left\{ \frac{K_1\left(\frac{x_i - x_j}{h}\right)}{f(y_i|x_i, \theta_0)} \frac{\partial \log f(y_i|x_i, \theta_0)}{\partial \theta'} \right.$$

$$\left. \times \int \left[K_2\left(\frac{y_i - u}{h}\right) - \int K_2\left(\frac{y_i - y}{h}\right) f(y|x_j, \theta_0)\, dy \right] f(u|x_j, \theta_0)\, du \right\}$$

$$= 0, \tag{A.18}$$

$$r_n(z_i) = \tfrac{1}{2}E[q_{1n}(z_i, z_j)|z_i] + \tfrac{1}{2}E[q_{1n}(z_j, z_i)|z_i] = \tfrac{1}{2}E[q_{1n}(z_j, z_i)|z_i]$$

$$= \frac{1}{2h^{l+m}} \int \frac{K_1\left(\frac{x - x_i}{h}\right)}{f(u|x, \theta_0)} \frac{\partial \log f(u|x, \theta_0)}{\partial \theta'}$$

$$\times \left[K_2\left(\frac{u - y_i}{h}\right) - \int K_2\left(\frac{u - y}{h}\right) f(y|x_i, \theta_0)\, dy \right] f(u|x, \theta_0) p_1(x)\, du dx$$

$$= \frac{1}{2h^{l+m}} \int K_1(u) \frac{\partial \log f(y_i + hv|x_i + hu, \theta_0)}{\partial \theta'}$$

$$\times \left[K_2(v) - \int K_2(w) f(y_i - hw|x_i, \theta_0) h^l dw \right] p_1(x_i + hu) h^l dv h^m du$$

$$= \frac{1}{2} \int K_1(u) \frac{\partial \log f(y_i|x_i, \theta_0)}{\partial \theta'} K_2(v) p_1(x_i)\, dv du + o_p(h)$$

$$\equiv \frac{1}{2} \frac{\partial \log f(y_i|x_i, \theta_0)}{\partial \theta'} p_1(x_i) + t_n(z_i). \tag{A.19}$$

Thus

$$\sqrt{n}\hat{U}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(y_i|x_i, \theta_0)}{\partial \theta'} p_1(x_i) + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} t_n(z_i). \tag{A.20}$$

Note that $|t_n(z_i)| \leq Mh$ for some constant M. From (A.18) $E[r_n(z_i)] = \bar{r}_n = 0$. Under the null and the assumptions, it is well known that the conditional score has zero expectation, that is,

$$E\left[\frac{\partial \log f(y_i|x_i, \theta_0)}{\partial \theta'} \middle| x_i \right] = 0. \tag{A.21}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Thus $E\left[\dfrac{\partial \log f(y_i|x_i,\theta_0)}{\partial \theta'} p_1(x_i)\right] = 0$. We have $E[t_n(z_i)] = 0$ and

$$E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_n(z_i)\right]^2 = \frac{1}{n}\sum_{i=1}^{n} E[t_n^2(z_i)] \le M^2 h^2. \tag{A.22}$$

So $n^{-1/2}\sum_{i=1}^{n} t_n(z_i) = o_p(1)$, and the asymptotic distribution of $\sqrt{n}\hat{U}_n$ is the same as that of $(1/\sqrt{n})\sum_{i=1}^{n} \dfrac{\partial \log f(y_i|x_i,\theta_0)}{\partial \theta'} p_1(x_i)$. By the Lindeberg–Levy central limit theorem, we have

$$\sqrt{n}Q_{1n} \equiv \sqrt{n}U_n \to \mathcal{N}\left(0, E\left[\frac{\partial \log f(y_i|x_i,\theta_0)}{\partial \theta'} p_1(x_i)\right]^2\right\}. \tag{A.23}$$

Thus $Q_{1n} = O_p(n^{-1/2})$.

The term Q_{2n} is a U-statistic with

$$q_n(z_i, z_j) = \tfrac{1}{2}q_{2n}(z_i, z_j) + \tfrac{1}{2}q_{2n}(z_j, z_i). \tag{A.24}$$

Similarly, we can show that the condition of Lemma 1 is satisfied. For the preceding $q_n(z_i, z_j)$, noting that from (A.19)

$$\int \frac{\partial f(y_j|x_j,\theta_0)}{\partial \theta'} dy_j = E\left[\frac{\partial \log f(y_j|x_j,\theta_0)}{\partial \theta'}\middle| x_i\right] = 0, \tag{A.25}$$

we have

$$\begin{aligned}
\bar{r}_n &= E[q_n(z_i, z_j)] = E[q_{2n}(z_i, z_j)] = E\{E[q_{2n}(z_i, z_j)|x_j, x_i]\} \\
&= \frac{1}{h^{l+m}} E\left\{\frac{K_1\left(\dfrac{x_i - x_j}{h}\right)}{f(u|x_i,\theta_0)} \int K_2\left(\frac{u-v}{h}\right)\frac{\partial f(v|x_j,\theta_0)}{\partial \theta'} dv f(u|x_i,\theta_0)\, du\right\} \\
&= \frac{1}{h^{l+m}} E\left\{K_1\left(\frac{x_i - x_j}{h}\right)\int \frac{\partial f(v|x_j,\theta_0)}{\partial \theta'} \int K_2\left(\frac{u-v}{h}\right) du dv\right\} \\
&= \frac{1}{h^{l+m}} E\left\{K_1\left(\frac{x_i - x_j}{h}\right)\int \frac{\partial f(v|x_j,\theta_0)}{\partial \theta'} h^l dv\right\} = 0.
\end{aligned} \tag{A.26}$$

Similarly to the rest of the proof that $Q_{1n} = O_p(n^{-1/2})$, we can show that $Q_{2n} = O_p(n^{-1/2})$. Thus we have shown that $W_{2n} = Q_{1n} + Q_{2n} = O_p(n^{-1/2})$. Because $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$,

$$nh^{(l+m)/2} W_{2n}(\hat{\theta} - \theta_0) = O_p(h^{(l+m)/2}) = o_p(1). \tag{A.27}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Next we show that $W_{3n} = O_p(1)$. By Assumption A2, we have

$$E(\|W_{3n}\|) \le E\left\{\frac{1}{h^{l+m}}\frac{K_1\left(\frac{x_i - x_j}{h}\right)}{f(y_i|x_i,\theta_0)} \times \left[2K_2\left(\frac{y_i - y_j}{h}\right)b(y_i,x_i) + 2\int K_2\left(\frac{y_i - u}{h}\right)du \times b^2(y_i,x_i) + 2\int K_2\left(\frac{y_i - u}{h}\right)du \times b^2(y_i,x_i) + 2\int K_2\left(\frac{y_i - u}{h}\right)du \times b(y_i,x_i)\right]\right\}$$

$$= \int \frac{1}{h^{l+m}}\frac{K_1\left(\frac{x - v}{h}\right)}{f(y|x,\theta_0)}\left[2K_2\left(\frac{y - u}{h}\right)b(y,x) + 4h^l b^2(y,x) + 2h^l b(y,x)\right] \times f(y|x,\theta_0)f(u|v,\theta_0)p_1(x)p_1(v)dydudxdv$$

$$= \int \frac{1}{h^{l+m}}K_1(u)[2K_2(v)b(y,x) + 4h^l b^2(y,x) + 2h^l b(y,x)] \times f(y - hv|x - hu,\theta_0)p_1(x)p_1(x - hv)\,dyh^l dvdxh^m du$$

$$= 2E[b(y_i,x_i)p_1(x_i)] + o(1) = O(1). \quad \textbf{(A.28)}$$

Thus $W_{3n} = O_p(1)$ and

$$(\hat{\theta} - \theta_0)W_{3n}(\hat{\theta} - \theta_0)' = O_p(h^{(l+m)/2}) = o_p(1). \quad \textbf{(A.29)}$$

By (A.4), (A.15), (A.27), and (A.29), we have proved the first part of Theorem 1, that is,

$$nh^{(l+m)/2}W_n = nh^{(l+m)/2}W_{1n} + o_p(1) \xrightarrow{d} \mathcal{N}(0,\sigma^2), \quad \textbf{(A.30)}$$

where σ^2 is the asymptotic variance of $nh^{(l+m)/2}W_n$,

$$\sigma^2 = 2\int K_1^2(u)\,du \int p_1^2(x)\,dx. \quad \textbf{(A.31)}$$

Next we show that $\hat{\sigma}^2$ is a consistent estimator of σ^2. Here σ^2 is a standard U-statistic with kernel

$$q_n(z_i,z_j) = \frac{2}{h^m}K_1^2\left(\frac{x_i - x_j}{h}\right). \quad \textbf{(A.32)}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Similarly to the previous proofs, we can easily show that the condition of Lemma 1 is satisfied. For the preceding $q_n(z_i, z_j)$, because

$$\bar{r}_n = E\left[\frac{2}{h^m} K_1^2\left(\frac{x_i - x_j}{h}\right)\right]$$

$$= \frac{2}{h^m} \iint K_1^2\left(\frac{x - v}{h}\right) p_1(x) p_1(v)\, dxdv$$

$$= 2\int K_1^2(u)\, du \int p_1^2(x)\, dx + o(1) = \sigma^2 + o(1), \tag{A.33}$$

by Lemma 1, $\hat{\sigma}^2 = \bar{r}_n + o_p(1) = \sigma^2 + o_p(1)$. ■

Proof of Theorem 2. Similarly to Theorem 1, we can show that

$$W_n = W_{1n}^* + o_p(1), \tag{A.34}$$

where W_{1n}^* is defined similarly to W_{1n} in (A.4) except θ_0 is replaced by θ^*. The kernel for W_{1n}^* is defined similarly to that in (A.5). Under the alternative, we have

$$\bar{r}_n = 2E[J_n(z_i, z_j)]$$

$$= \frac{1}{h^{l+m}} \int \frac{K_1\left(\frac{x - v}{h}\right) K_2\left(\frac{y - u}{h}\right)}{f(y|x, \theta^*)} p(y, x) p(u, v)\, dydxdudv$$

$$- \frac{1}{h^{l+m}} \int \frac{K_1\left(\frac{x - v}{h}\right)}{f(y|x, \theta^*)} \left[\int K_2\left(\frac{y - u}{h}\right) f(u|v, \theta^*)\, du\right] p(y, x) p_1(x)\, dydxdv$$

$$= \frac{1}{h^{l+m}} \int \frac{K_1(u) K_2(v)}{f(y|x, \theta^*)} p(y, x) p(y - hv, x - hu)\, dydxh^l dvh^m du$$

$$- \frac{1}{h^{l+m}} \int \frac{K_1(u)}{f(y|x, \theta^*)} \left[\int K_2(v) f(y - hv|x - hu, \theta^*) h^l dv\right] p(y, x) p_1(x)\, dydxh^m du$$

$$= \int \frac{p(y, x)}{f(y|x, \theta^*)} p(y, x)\, dydx - \int \frac{f(y|x, \theta^*) p_1(x)}{f(y|x, \theta^*)} p(y, x)\, dydx$$

$$= E\left[\frac{p(y_i, x_i) - f(y_i|x_i, \theta^*) p_1(x_i)}{f(y_i|x_i, \theta^*)}\right] + o(1). \tag{A.35}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Therefore by Lemma 1,

$$W_n = W_{1n} + o_p(1) = \bar{r}_n + o_p(1) = E\left[\frac{p(y_i, x_i) - f(y_i|x_i, \theta^*)p_1(x_i)}{f(y_i|x_i, \theta^*)}\right] + o(1). \quad \textbf{(A.36)}$$

■

Proof of Theorem 3. Similarly to previous proofs, under the local alternatives (3.5), we have

$$W_n = W_{1n} + o_p(1). \quad \textbf{(A.37)}$$

Expanding the following term as a function of d_n around $d_n = 0$, we have

$$\begin{aligned} Q_{3n} &\equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^{l+m}} K_1\left(\frac{x_i - x_j}{h}\right) \\ &\quad \times \frac{K_2\left(\frac{y_i - y_j}{h}\right) - \int K_2\left(\frac{y_i - y}{h}\right)[\, f(y|x_j, \theta_0) + d_n l(y, x_j)]\, dy}{f(y_i|x_i, \theta_0) + d_n l(y_i, x_i)} \\ &= W_{1n} - d_n \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^{l+m}} \frac{l(y_i, x_i) K_1\left(\frac{x_i - x_j}{h}\right)}{f^2(y_i|x_i, \theta_0)} \right. \\ &\qquad \left. \times \left[K_2\left(\frac{y_i - y_j}{h}\right) - \int K_2\left(\frac{y_i - y}{h}\right) f(y|x_j, \theta_0)\, dy \right] \right\} \\ &\quad - d_n \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^{l+m}} \frac{K_1\left(\frac{x_i - x_j}{h}\right)}{f(y_i|x_i, \theta_0)} \int K_2\left(\frac{y_i - y}{h}\right) l(y, x_j)\, dy \right\} + o(d_n) \\ &\equiv W_{1n} - d_n Q_{4n} - d_n Q_{5n} + o(d_n). \end{aligned} \quad \textbf{(A.38)}$$

Thus

$$W_n = Q_{3n} + d_n Q_{4n} + d_n Q_{5n} + o(d_n). \quad \textbf{(A.39)}$$

Noting that under the local alternatives (3.5), the true conditional distribution is $f(y|x, \theta_0) + d_n \times l(y, x)$, similarly to the proof of Theorem 1, we can show that

$$nh^{(l+m)/2} Q_{3n} \xrightarrow{d} \mathcal{N}(0, \sigma^2). \quad \textbf{(A.40)}$$

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press

Similarly to the proof that $Q_{1n} = O_p(n^{-1/2})$, we can show that $Q_{4n} = O_p(n^{-1/2})$. For the kernel of Q_{5n},

$$\begin{aligned}
\bar{r}_n &= \frac{1}{h^{l+m}} E\left[\frac{K_1\left(\frac{x_i - x_j}{h}\right)}{f(y_i|x_i,\theta_0)} \int K_2\left(\frac{y_i - y}{h}\right) l(y, x_j)\, dy \right] \\
&= \frac{1}{h^{l+m}} \int \frac{K_1\left(\frac{x - v}{h}\right)}{f(y|x,\theta_0)} \int K_2\left(\frac{y_i - u}{h}\right) l(u,v)\, du \\
&\quad \times [\, f(y|x,\theta_0) + d_n l(y,x)] p_1(x) p_1(v)\, dydxdv \\
&= \frac{1}{h^{l+m}} \int \frac{K_1(u)}{f(y|x,\theta_0)} \int K_2(w) l(y - hw, x - hu) h^l dw \\
&\quad \times [\, f(y|x,\theta_0) + d_n l(y,x)] p_1(x) p_1(x - hu)\, dydx h^m du \\
&= \int l(y,x) p_1^2(x) dydx + d_n \int \frac{l^2(y,x) p_1^2(x)}{f(y|x,\theta_0)}\, dydx + o(d_n) \\
&= d_n E\left[\frac{l^2(y_i, x_i)}{f^2(y_i|x_i,\theta_0)} p_1(x_i) \right] + o(d_n), \qquad \textbf{(A.41)}
\end{aligned}$$

where $\int l(y, x_i) dy = 0$. By Lemma 1, we have

$$Q_{5n}/d_n \xrightarrow{p} E\left[\frac{l^2(x_i)}{f^2(y_i|x_i)} p_1(x_i) \right] > 0. \qquad \textbf{(A.42)}$$

If $d_n = n^{-1/2} h^{-(l+m)/4}$, then

$$nh^{(l+m)/2} d_n Q_{4n} = h^{(l+m/4)} n^{-1/2} Q_{4n} \xrightarrow{p} 0, \qquad \textbf{(A.43)}$$

$$nh^{(l+m)/2} d_n Q_{5n} = Q_{5n}/d_n \xrightarrow{p} E\left[\frac{l^2(y_i, x_i)}{f^2(y_i|x_i,\theta_0)} p_1(x_i) \right] > 0. \qquad \textbf{(A.44)}$$

Hence,

$$nh^{(l+m)/2} W_n \xrightarrow{d} \mathcal{N}\left(E\left[\frac{l^2(y_i, x_i)}{f^2(y_i|x_i,\theta_0)} p_1(x_i) \right], \sigma^2 \right), \qquad \textbf{(A.45)}$$

$$T_n \xrightarrow{d} \mathcal{N}(\mu, 1). \qquad \textbf{(A.46)}$$

■

https://doi.org/10.1017/S026646660016503X Published online by Cambridge University Press