Econometric Theory, **14**, 1998, 123–138. Printed in the United States of America.

A CONSISTENT NONPARAMETRIC TEST OF PARAMETRIC REGRESSION MODELS UNDER CONDITIONAL QUANTILE RESTRICTIONS

JOHN XU ZHENG
University of Texas

This paper proposes a nonparametric, kernel-based test of parametric quantile regression models. The test statistic has a limiting standard normal distribution if the parametric quantile model is correctly specified and diverges to infinity for any misspecification of the parametric model. Thus the test is consistent against any fixed alternative. The test also has asymptotic power 1 against local alternatives converging to the null at proper rates. A simulation study is provided to evaluate the finite-sample performance of the test.

1. INTRODUCTION

This paper is concerned with testing the correctness of the specification of the following parametric quantile regression model:

$$y_i = g(x_i, \beta_0) + u_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where the conditional θth $(0 < \theta < 1)$ quantile of the error term u_i given x_i is zero, y_i is the dependent variable, x_i is the vector of regressors, and β_0 is the vector of parameters to be estimated. Here the regression function $g(x_i, \beta_0)$ measures the conditional quantile of y_i given x_i, not, as is usual, the conditional mean. Much of the work on quantile estimations of econometric models is due to Koenker and Bassett (1978, 1982) and Powell (1984, 1986). The quantile linear regression method introduced by Koenker and Bassett (1978) has been extended to simultaneous equation models (Amemiya, 1982; Powell, 1983) and nonlinear regression models (Obenhofer, 1982). Powell's least absolute deviations estimation method for the censored regression model (1984) has also been extended to censored regression quantiles (Powell, 1986) and monotonic regression models (Powell, 1991). Newey and Powell (1990) constructed efficient estimators for linear quantile regression models and censored quantile regression models. Chamberlain (1994) and Buchinsky (1994) provided empirical applications of quantile regression techniques in estimating earnings functions. Quantile estimations are

I am grateful to the coeditor Joel Horowitz and two anonymous referees for helpful comments. I also thank Maxwell Stinchcombe for useful comments and proofreading. Any remaining errors are mine only. 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/S0266466698141051 Published online by Cambridge University Press

also useful in testing some interesting features of error distributions, such as homoskedasticity (Koenker and Bassett, 1982; Powell, 1986) and conditional symmetry (Powell, 1986).

However, the aforementioned quantile estimations rely on the assumption that the conditional quantile of the dependent variable given the covariates belongs to a parametric family of known functions, often linear functions. Misspecifying the parametric quantile may lead to incorrect inference. One way to circumvent the possible misspecification is to estimate the conditional quantile nonparametrically. Tsybakov (1982) and Bhattacharya and Gangopadhyay (1990) provided kernel and nearest neighbor estimations of conditional quantiles. White (1992) developed a nonparametric estimator using neural networks. However, nonparametric methods are inefficient compared to the parametric methods if the parametric model is correctly specified. A good specification test may allow the researcher to avoid this loss of efficiency.

To my knowledge, testing functional forms of a parametric quantile regression has not yet been considered in the literature, despite the fact that the related problem of testing regression models under conditional mean restrictions has been extensively investigated. The related work includes that of Bierens (1990), Eubank and Spiegelman (1990), Fan and Li (1992), Härdle and Mammen (1993), Hong and White (1995), Horowitz and Härdle (1994), Lee (1988), Stinchcombe and White (1993), Whang and Andrews (1993), Wooldridge (1992), Yatchew (1992), Zheng (1996a), and others. This paper intends to fill the literature gap by providing a consistent test of the parametric quantile regression model using nonparametric techniques. We shall show that the related work on testing mean restrictions can be extended to the quantile restriction case.

The paper is organized as follows. Section 2 presents the test statistic. Section 3 establishes the asymptotic distribution under the null hypothesis that the parametric model is correct and the powers against both fixed and local alternatives. Section 4 provides a Monte Carlo simulation of the test. Section 5 concludes the paper.

2. THE DERIVATION OF THE 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 $F(y,x)$. In econometrics, Y and X usually denote a dependent variable and a vector of covariates, respectively, as in model (1.1). Denote the conditional distribution function Y given X by $F(y|x)$. The θth $(0<\theta<1)$ conditional quantile $Q_\theta(x)$ of y_i given $x_i=x$ is defined as

$$Q_\theta(x)\equiv \inf\{y: F(y|x)\geq \theta\}. \tag{2.1}$$

If, as is assumed throughout, the conditional distribution function $F(y|x)$ is absolutely continuous in y for almost all x, then $F(Q_\theta(x)|x)=\theta$. Denote $1(\cdot)$ as the

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

indicator function; i.e., $1(A) = 1$ if event A occurs and $1(A) = 0$ otherwise. Define the "check function"

$$\rho_\theta(\epsilon) \equiv [\theta 1(\epsilon \geq 0) - (1 - \theta)1(\epsilon < 0)]\epsilon \tag{2.2}$$

as in Koenker and Bassett (1978). It is well known that

$$Q_\theta(\cdot) = \arg\min_w E[\rho_\theta(y_i - w(x_i))], \tag{2.3}$$

where $w(\cdot)$ are Lebesgue measurable functions.

In parametric quantile regression models, the θth conditional quantile of Y given X is assumed to belong to a parametric family of functions $g(x, \beta)$ where $\beta \in B \subset R^q$. The null hypothesis to be tested is that the conditional quantile is correctly specified:

$$H_0\text{: } \Pr[F(g(x_i, \beta_0)|x_i) = \theta] = 1 \quad \text{for some } \beta_0 \in B, \tag{2.4}$$

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

$$H_1\text{: } \Pr[F(g(x_i, \beta)|x_i) = \theta] < 1 \quad \text{for all } \beta \in B. \tag{2.5}$$

Note that the alternative encompasses all possible departures of the null model. A test is consistent if it has asymptotic power equal to 1 against the infinite-dimensional alternative (2.5).

The null hypothesis is true if and only if $E[1(y_i \leq g(x_i, \beta_0))|x_i] = \theta$ or $1(y_i \leq g(x_i, \beta_0)) = \theta + \epsilon_i$ where $E(\epsilon_i|x_i) = 0$. This reduces the problem of testing a quantile restriction to a problem of testing a particular type of mean restriction. This paper extends the mean restriction tests of Zheng (1995, 1996a, 1996b) to the quantile restriction case. It seems clear that the other specification tests cited earlier can also be extended.

Let ϵ_i denote $1(y_i \leq g(x_i, \beta_0)) - \theta$. Let $p(x)$ be the marginal density function of X. Then $E[\epsilon_i|x_i] = F(g(x_i, \beta_0)|x_i) - \theta$. Observe that

$$\begin{aligned} W &\equiv E\{\epsilon_i E[\epsilon_i|x_i]p(x_i)\} = E\{[E(\epsilon_i|x_i)]^2 p(x_i)\} \\ &= E\{[F(g(x_i, \beta_0)|x_i) - \theta]^2 p(x_i)\} \geq 0, \end{aligned} \tag{2.6}$$

where equality holds if and only if H_0 is true. The test is based on the sample analog of W (weighting W by the density function $p(x)$ serves the same purpose as in Powell, Stock, and Stoker [1989] of avoiding the problem of trimming on the boundary of the density function).

The density weighted conditional expectation $E(\epsilon_i|x_i)p(x_i)$ can be estimated by the kernel regression

$$\hat{E}(\epsilon_i|x_i)\hat{p}(x_i) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right)[1(y_j \leq g(x_j, \beta_0)) - \theta], \tag{2.7}$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. The unknown parameters β_0 can be estimated as in Obenhofer (1982) and Dupacova and Wets (1988). Let $\hat{\beta}$ be an estimator of β_0.

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

Replacing ϵ_i by $e_i \equiv 1(y_i \le g(x_i, \hat{\beta})) - \theta$, we obtain the sample analog of W,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{h^m}K\left(\frac{x_i - x_j}{h}\right)[1(y_i \le g(x_i,\hat{\beta})) - \theta][1(y_j \le g(x_j,\hat{\beta})) - \theta]. \quad \textbf{(2.8)}$$

3. THE LIMITING DISTRIBUTIONS OF THE TEST STATISTIC

We consider the case where some components of x are discrete and some are continuous. Let $x_i = (x_{1i}', x_{2i}')' \in R^{m_1} \times R^{m_2}$ where x_{1i} is discrete, x_{2i} is continuous, and $m_1 + m_2 = m$.

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. $\{y_i, x_i\}_{i=1}^n$ is a random sample from a probability distribution $F(y,x)$ on $R \times R^m$. Let $p_1(x_1)$ be the marginal probability function of x_{1i}. There exists a countable subset X_1 of R^{m_1} such that

(i) $x_1 \in X_1$ implies $p_1(x_1) > 0$;
(ii) $\sum_{x_1 \in X_1} p_1(x_1) = 1$; and
(iii) $\inf\{\|z_1 - z_2\| : z_1, z_2 \in X_1, z_1 \ne z_2\} > 0$.

Assumption 2. Let the marginal density of x_{2i} be $p_2(x_2)$ and the conditional density of x_{2i} given $x_{1i} = x_1$ be $q(x_2|x_1)$. Here $p_2(x_2)$, $q(x_2|x_1)$, and their first-order derivatives with respect to x_2 are uniformly bounded. The conditional density $f(y|x_1,x_2)$ of y give x_1 and x_2 has bounded first-order derivatives with respect to y and x_2. Denote $p(x_1,x_2) \equiv p_1(x_1)q(x_2|x_1)$.

Assumption 3. The parameter space B is a compact and convex subset of R^q.

$$E\left[\sup_{\beta\in B}\left\|\frac{\partial^2 g(x_i,\beta)}{\partial\beta\partial\beta'}\right\|\right] < \infty,$$

$$E\left[\sup_{\beta\in B}\left\|\frac{\partial g(x_i,\beta)}{\partial\beta}\,\frac{\partial g(x_i,\beta)}{\partial\beta'}\right\|\right] < \infty.$$

There exists an estimator $\hat{\beta}$ of $\beta_0 \equiv \arg\min_{\beta\in B} E[\rho_\theta(y_i - g(x_i,\beta))]$ such that under the null $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$, whereas under the alternative, $\hat{\beta} \to_p \beta^* \in B$. Both β_0 and β^* are interior points of B.

Assumption 4. The kernel function K takes the form $K(u_1,u_2) = K_1(u_1)K_2(u_2)$ on $R^{m_1} \times R^{m_2}$. The term $K_1(\cdot)$ has a bounded support on R^{m_1}. Here $K_1(0) = 1$ and $K_2(\cdot)$ is a nonnegative, bounded, symmetric, and Lebesgue measurable function on R^{m_2} such that $\int K_2(u)\,du = 1$ and $\int \|u\| K_2(u)\,du < \infty$.

The condition (iii) of Assumption 1 ensures that the set X_1 does not contain limit points of X_1. The condition is satisfied if x_1 are dummy variables. The estimator in Assumption 3 is satisfied by that of Dupacova and Wets (1988).

Besides testing the null against fixed alternatives (2.5), we also consider testing for local departures from the null. Classical tests usually consider local mis-

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

specifications that converge to the null at the parametric rate $n^{-1/2}$. Here we consider this sequence of local alternatives:

$$H_{1n}: F[(g(x,\beta_0) + d_n l(x))|x] = \theta, \tag{3.1}$$

where $d_n \to 0$ as $n \to \infty$ and the function $l(\cdot)$ and its first-order derivatives are bounded.

The main difficulty in deriving the asymptotic distribution of the statistic defined in (2.8) is that the indicator function $1(y_i \le g(x_i,\beta))$ is not differentiable with respect to β. This prevents taking usual Taylor expansion on $1(y_i \le g(x_i,\hat{\beta}))$ around β_0. To circumvent the problem, we appeal to the work of Sherman (1994) on uniform convergence of *U*-statistic indexed by parameters. As it turns out, the limiting distribution of the statistic in (2.8) only depends on h^{m_2}. To simplify the derivation of its asymptotic distributions, we multiply the statistic in (2.8) by h^{m_1} and drop the terms with $i = j$, which yields a *U*-statistic

$$W_n \equiv \frac{1}{n(n-1)} \sum_{i \ne j} \frac{1}{h^{m_2}} K\left(\frac{x_i - x_j}{h}\right) \times [1(y_i \le g(x_i,\hat{\beta})) - \theta][1(y_j \le g(x_j,\hat{\beta})) - \theta]. \tag{3.2}$$

The following theorem contains the asymptotic properties of this statistic.

THEOREM. *Given Assumptions* 1–4, *if* $h \to 0$ *and* $n^{\alpha}h^{2m_2} \to \infty$ *for some* $0 < \alpha < 1$ *as* $n \to \infty$, *then the following hold*:

(A) *under the null hypothesis* (2.4),

$$nh^{m_2/2}W_n \to_d \mathcal{N}(0,\sigma^2) \tag{3.3}$$

with

$$\sigma^2 \equiv 2\theta^2(1-\theta)^2 \int K_2^2(u)\,du \times E[p(x_{1i},x_{2i})], \tag{3.4}$$

where σ^2 *can be consistently estimated by*

$$\hat{\sigma}^2 \equiv 2\theta^2(1-\theta)^2 \frac{1}{n(n-1)} \sum_{i \ne j} \frac{1}{h^{m_2}} K^2\left(\frac{x_i - x_j}{h}\right); \tag{3.5}$$

(B) *under the alternative hypothesis* (2.5),

$$W_n \to_p E\{[F(g(x_i,\beta^*)|x_i) - \theta]^2 p(x_i)\} > 0; \tag{3.6}$$

(C) *under the local alternatives* (3.1) *with* $d_n = n^{-1/2}h^{-m_2/4}$,

$$W_n \to_d \mathcal{N}(\mu,\sigma^2), \tag{3.7}$$

where

$$\mu = E[f^2(g(x_i,\beta_0)|x_i)l^2(x_i)p(x_i)]. \tag{3.8}$$

Standardizing W_n, we obtain the final test statistic

$$T_n \equiv nh^{m_2/2}W_n/\hat{\sigma}. \tag{3.9}$$

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

Thus, $T_n \to_d \mathcal{N}(0,1)$ under the null, $T_n \to_p \infty$ under *any* fixed alternative, and $T \to_d \mathcal{N}(\mu/\sigma,1)$ under the local alternatives given in (3.1) and (C). By letting h approach zero sufficiently slowly, the test can be made to be consistent against any local alternatives converging to the null arbitrarily close to the parametric rate $n^{-1/2}$. Given their genesis, it is not surprising that the consistency and local power of this test closely parallel the consistency and local power of nonparametric tests of regression functions under conditional mean restrictions.

The behavior of the test may be further understood by comparing it with classical tests. Since the model (1.1) imposes very weak distributional assumptions, the classical tests such as Wald, likelihood ratio (LR), and Lagrange multiplier (LM) that require specifying a complete distribution do not apply here. One feasible test is the χ^2 test. To construct a χ^2 test, partition the product space $R \times R^m$ into subspaces $\{(y,x): y \le g(x,\beta_0)\}$ and $\{(y,x): y > g(x,\beta_0)\}$. The observed frequency falling in $\{(y,x): y \le g(x_i\beta_0)\}$ is $\sum_{i=1}^n 1[y_i \le g(x_i,\beta_0)]$, and its expected frequency under the null is $n\theta$. The observed frequency falling in $\{(y,x): y > g(x,\beta_0)\}$ is $\sum_{i=1}^n 1[y_i > g(x_i,\beta_0)] = n - \sum_{i=1}^n 1[y_i \le g(x_i,\beta_0)]$, and the expected frequency is $n(1-\theta)$. Then a χ^2 test may be based on the statistic

$$\chi_n^2 \equiv \left[\frac{\sum_{i=1}^n 1[y_i \le g(x_i,\hat{\beta})] - n\theta}{\sqrt{n\theta}}\right]^2 + \left[\frac{n - \sum_{i=1}^n 1[y_i \le g(x_i,\hat{\beta})] - n(1-\theta)}{\sqrt{n(1-\theta)}}\right]^2$$

$$= \frac{1}{\theta(1-\theta)} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{1[y_i \le g(x_i,\hat{\beta})] - \theta\}\{1[y_j \le g(x_j,\hat{\beta})] - \theta\}. \quad \textbf{(3.10)}$$

Equation (3.10) bears some resemblance to (2.8). Under the null $\chi_n^2 \to_d \chi^2(1)$, whereas under the alternative

$$\chi_n^2/n \to_p \frac{1}{\theta(1-\theta)} \{E[F(g(x_i,\beta^*)|x_i)] - \theta\}^2. \quad \textbf{(3.11)}$$

Thus the χ^2 test cannot distinguish among those alternatives for which

$$E[F(g(x_i,\beta^*)|x_i)] = \theta,$$

whereas $F[g(x,\beta^*)|x]$ is not identically equal to θ. This shows that the χ^2 test is not consistent against all possible alternatives.

4. A MONTE CARLO SIMULATION

In this section, we conduct a Monte Carlo simulation of the size and power of the test to see how the proposed test behaves in finite samples.

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

We consider five models where a random sample of the dependent variable y is generated as follows:

$$y_i = g(x_{1i}, x_{2i}) + u_i, \qquad i = 1, \ldots, n. \tag{4.1}$$

Here $u_i + 1$ has a standard lognormal distribution so that the median of u_i is zero. The terms u_i, x_{1i}, and x_{2i} are independent; x_{1i} is discrete and has a binomial distribution with the probability of success equal to 0.5 and the number of independent trials equal to 5; and x_{2i} has the standard normal distribution. In models 1–5, the quantile regression functions are set to be $1 + x_1 + x_2$, $1 + x_1 + x_2 + (x_1^2 + x_1x_2 + x_2^2)$, $1 + x_1 + x_2 + (x_1^2 + x_1x_2 + x_2^2)^{3/2}$, $\log(1 + x_1^2 + x_2^2)$, and $\exp(1 + x_1 + x_2)$, respectively.

The null hypothesis H_0 that we want to test is that the conditional median of y given x_1 and x_2 belongs to the parametric family $x'\beta = \beta_0 + \beta_1 x_1 + \beta x_2$, and so $\theta = \frac{1}{2}$. Obviously, H_0 is true for model 1 and false for models 2–5.

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

$$\mathrm{GCV}(h) = \frac{\sum_{i=1}^{n} [1(y_i \le x_i'\hat{\beta}) - \theta - \hat{g}_h(x_i)]^2}{n[1 - \mathrm{tr}(H)/n]^2}, \tag{4.2}$$

where $\hat{g}_h(\cdot)$ is a kernel estimator of the regression function $E\{[1(y_i \le x_i'\beta) - \theta]|x_i\}$

$$\hat{g}_h(x) = \frac{\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)[1(y_j \le x_j'\hat{\beta}) - \theta]}{\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)} \tag{4.3}$$

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

$$h_{i,j} = \frac{K\left(\frac{x_i - x_j}{h}\right)}{\sum_{l=1}^{n} K\left(\frac{x_i - x_l}{h}\right)}. \tag{4.4}$$

In regression cases, minimizing GCV will yield an asymptotically optimal bandwidth that is proportional to $n^{-1/(m+4)}$ (see Härdle, 1990). So we choose the bandwidth h to be $\lambda n^{-1/6}$, where λ is a constant. We use the grid search method to find the optimal λ by letting $0.05 \le \lambda \le 5.0$ and the grid points be 0.05, 0.15,…,4.90, and 5.0 with increment 0.1.

We choose the kernel to be the bivariate Epanechnikov kernel $K(u_1, u_2) = K(u_1)K(u_2)$, where

$$K(u) = 0.75(1 - u^2)1(|u| \le 1).$$

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

TABLE 1. Percentage of rejections in models 1–5

Sample size	1% level	5% level	10% level
Model 1: $y = 1 + x_1 + x_2 + u$			
100	0.7	4.3	9.2
200	1.0	4.7	10.3
300	1.1	5.3	9.5
Model 2: $y = 1 + x_1 + x_2 + (x_1^2 + x_1x_2 + x_2^2) + u$			
100	20.4	38.7	52.7
200	61.4	83.1	90.4
300	92.9	98.4	99.6
Model 3: $y = 1 + x_1 + x_2 + (x_1^2 + x_1x_2 + x_2^2)^{3/2} + u$			
100	84.3	95.2	98.3
200	100	100	100
300	100	100	100
Model 4: $y = \log(1 + x_1^2 + x_2^2) + u$			
100	12.2	26.0	38.6
200	32.7	56.5	69.5
300	62.4	83.4	90.2
Model 5: $y = \exp(1 + x_1 + x_2) + u$			
100	93.6	99.0	99.6
200	100	100	100
300	100	100	100

The simulation is conducted for sample sizes 100, 200, and 300. Each experiment is based on 1,000 replications. Here β_0 are estimated by the least absolute deviation estimation. The critical values for the test are from the standard normal distribution. Let $N_{1-\pi}$ be the $(1 - \pi)$th percentile from the standard normal distribution. Then at $(1 - \pi)$ level, we reject the null if $T_n > N_{1-\pi}$. We let $\pi = 1\%$, 5%, and 10%.

The results of the simulation are shown in Table 1. We can see that the nonparametric test has adequate size. As sample size gets larger, the power increases, in line with the asymptotic theory.

5. CONCLUSION

This paper has provided a consistent test of parametric regression models under conditional quantile restrictions. The test can be extended to testing for the censored regression model $y_i = \max\{0, x_i'\beta_0 + \epsilon_i\}$ or other monotonic regression models considered by Powell (1991) by exploiting the fact that the quantile of a monotonic function is a monotonic function of the quantile. For example, for the censored regression model, a consistent test may be obtained by replacing $1(y_i \leq x_i'\hat{\beta})$ with $1(y_i \leq \max\{0, x_i'\hat{\beta}\})$. The test can also be used to test

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

multiple quantile restrictions, H_0: $q_\theta(x) = g(x, \beta_0(\theta))$ for all $\theta \in \Theta$ by simply substituting $\sum_{\theta\in\Theta}\{1[y_i \le g(x_i, \hat{\beta}(\theta))] - \theta\}\{1[y_i \le g(x_i, \hat{\beta}(\theta))] - \theta\}$ for $\{1[y_i \le g(x_i, \hat{\beta})] - \theta\}\{1[y_i \le g(x_i, \hat{\beta})] - \theta\}$. The test of multiple quantile restrictions may be used to test for homoskedasticity as in Koenker and Bassett (1982) and Powell (1986). Using infinite number of quantile restrictions, we may be able to test a fully specified parametric conditional distribution.

REFERENCES

Amemiya, T. (1982) Two stage least absolute deviations estimators. *Econometrica* 50, 689–711.

Bhattacharya, P.K. & A.K. Gangopadhyay (1990) Kernel and a nearest-neighbor estimation of a conditional quantile. *Annals of Statistics* 18, 1400–1415.

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

Buchinsky, M. (1994) Changes in the U.S. wage structure 1963–1987: Application of quantile regression. *Econometrica* 62, 405–458.

Chamberlain, G. (1994) Quantile regression, censoring, and the structure of wages. In C.A. Sims (ed.), *Advances in Econometrics*, Sixth World Congress, vol. I, pp. 171–209. Cambridge: Cambridge University Press.

Dupacova, J. & R. Wets (1988) Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Annals of Statistics* 16, 1517–1549.

Eubank, R.L. & C.H. Spiegelman (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association* 85, 387–392.

Fan, Y. & Q. Li (1992) The Asymptotic Expansion of Kernel Sum of Squared Residuals and Its Applications in Hypotheses Testing. Manuscript, University of Windsor.

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.

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, 1133–1157.

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

Koenker, R. & G. Bassett (1978) *Regression quantiles. Econometrica* 46, 33–50.

Koenker, R. & G. Bassett (1982) Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43–61.

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

Newey, W.K. & J.L. Powell (1990) Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* 6, 295–317.

Obenhofer, W. (1982) The consistency of nonlinear regression minimizing the L_1 normal. *Annals of Statistics* 10, 316–319.

Powell, J.L. (1983) The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica* 51, 1569–1575.

Powell, J.L. (1984) Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303–325.

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

Powell, J.L. (1991) Estimation of monotonic regression models under quantile restrictions. In W.A. Barnett, J.L. Powell, & G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 357–384. Cambridge: Cambridge University Press.

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

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

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

Stinchcombe, M.B. & H. White (1993) Consistent Specification Testing with Unidentified Nuisance Parameters Using Duality and Banach Space Limit Theory. Manuscript, University of California, San Diego.

Tsybakov, A.B. (1982) Robust estimates of a function. *Problems of Information Transmission* 18, 190–201.

Whang, Y.-J. & Andrews, D.W.K. (1993) Tests of specification for parametric and semiparametric models. *Journal of Econometrics* 57, 277–318.

White, H. (1992) Nonparametric estimation of conditional quantiles using neural networks. In H. White (ed.), *Artificial Neural Networks: Approximation and Learning Theory*, pp. 191–205. Oxford: Blackwell.

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. (1995) 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.

APPENDIX: PROOFS

We first present a result on uniform convergence of *U*-statistic indexed by parameters.

Let $z_1,\ldots,z_n$ be an independent and identically distributed (i.i.d.) random sample with distribution on a set S. Let B be a subset of R^q. For each $\beta \in B$, $H_n(z_1,\ldots,z_k,\beta)$ is a real-valued function on S^k. Following the notations in Sherman (1994), we define the kth-order U-statistic indexed by β as

$$U_n^k H_n(\cdot,\beta) = \frac{k!}{n!} \sum_{i_a \neq i_b, a \neq b} H_n(z_{i_1},\ldots,z_{i_k},\beta). \tag{A.1}$$

If $E[H_n(z_1,\ldots,z_k,\beta)|z_i] = 0$ for $i = 1,\ldots,k$, then $H_n(\cdot,\beta)$ and $U_n^k H_n(\beta)$ are said to be degenerate.

Let $\mathcal{H} \equiv \{H_n(\cdot,\beta):\beta \in B, n = 1,2,\ldots\}$. We say that $\mathcal{H}$ has an envelope $L(\cdot)$ if $\sup_{H_n \in \mathcal{H}} |H_n(\cdot,\beta)| \leq \mathrm{L}(\cdot)$. Under some mild regularity condition called a euclidean (see Sherman, 1994), we can establish the uniform convergence of $U_n^k H_n(\cdot,\beta)$ in a neighborhood of $\beta_0 \in B$.

LEMMA 1. *Let $\mathcal{H} \equiv \{H_n(\cdot,\beta):\beta \in B\}$ be a class of degenerate functions on S^k. If* (i) *$\mathcal{H}$ is Euclidean for a constant envelope and* (ii) *$E[H_n(\cdot,\beta) - H_n(\cdot,\beta_0)]^2 \leq O(n^{-1/2})$ uniformly over $O_p(n^{-1/2})$ neighborhoods of β_0, then for any $0 < \alpha < 1$,*

$$|U_n^k H_n(\cdot,\beta) - U_n^k H_n(\cdot,\beta_0)| \leq O_p(n^{-k/2-\alpha/4})$$

uniformly over $O_p(n^{-1/2})$ neighborhoods of β_0.

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

Proof of Lemma 1. We modify the proof of Corollary 8 of Sherman (1994) by applying it to $G_n(\cdot,\beta) \equiv H_n(\cdot,\beta) - H_n(\cdot,\beta_0)$.

Let $\{M_n\}$ be a sequence of bounded real numbers and $B_n = \{\beta \in B : \sqrt{n}\|\beta - \beta_0\| \le M_n\}$. Applying the main corollary of Sherman (1994) with $p = 1$, we have that for any $0 < \alpha < 1$,

$$E\left[\sup_{B_n}|n^{k/2}U_n^k G_n(\cdot,\beta)|\right] \le \Lambda\left\{E\left[\sup_{B_n}(U_{2n}^k G_n^2(\cdot,\theta))^\alpha\right]\right\}^{1/2}, \tag{A.2}$$

where Λ is a generic constant. By Chebyshev's inequality, we have

$$E\left[\sup_{B_n}(U_{2n}^k G_n^2(\cdot,\beta))^\alpha\right] \le \left\{E\left[\sup_{B_n}(U_{2n}^k G_n^2(\cdot,\beta)\right]\right\}^\alpha. \tag{A.3}$$

Because $\{U_{2n}^k G_n^k(\cdot,\beta) : \beta \in B_n\}$ is a U-process of order k, it can be decomposed into a sum of degenerate U-statistics as in equation (6) of Sherman (1994). Thus, we have

$$E\left[\sup_{B_n}(U_{2n}^k G_n^2(\cdot,\beta))\right] \le \sup_{B_n} E[G_n^2(\cdot,\beta)] + \sum_{l=1}^{k} E\left[\sup_{B_n}|U_{2n}^l g_l(\cdot,\beta)|\right], \tag{A.4}$$

where the class of functions $\{g_l(\cdot,\beta) : \beta \in B_n\}$ are degenerate. Applying Corollary 4(i) of Sherman (1994), we have

$$\sum_{l=1}^{k} E\left[\sup_{B_n}|U_{2n}^l g_l(\cdot,\beta)|\right] = \sum_{l=1}^{k} O(n^{-l/2}) = O(n^{-1/2}). \tag{A.5}$$

Combining the preceding equation with Assumption (ii), we have

$$E\left[\sup_{B_n}(U_{2n}^k G_n^2(\cdot,\beta))\right] = O(n^{-1/2}). \tag{A.6}$$

By (A.2) and (A.6), we have proved the lemma. ■

The following lemma, which is also useful in proving the theorem, follows from Lemma 3.1 of Powell et al. (1989).

LEMMA 2. *Let $U_n^2 H(z_i,z_j)$ be a second-order U-statistic and $r_n(z_i) = E[H_n(z_i,z_j)|z_i]$. If $E[\|H_n(z_i,z_j)\|^2] = o(n)$, then*

$$\sqrt{n}\{U_n^2 H_n(z_i,z_j) - E[H_n(z_i,z_j)]\} = 2n^{-1/2}\sum_{i=1}^{n}\{r_n(z_i) - E[r_n(z_i)]\} + o_p(1)$$

and $U_n^2 H_n(z_i,z_j) = E[H_n(z_i,z_j)] + o_p(1)$.

Proof of Theorem. Note that, by Assumption 4, for sufficiently large n, $K_1[(x_{1i} - x_{1j})/h] = 1(x_{1i} = x_{1j})$, and so we only need to consider

$$W_n = \frac{1}{n(n-1)}\sum_{i\ne j}\frac{1}{h^{m_2}}1(x_{1i} = x_{1j})K_2\left(\frac{x_{2i} - x_{2j}}{h}\right) \times \{1[y_i \le g(x_i,\hat\beta)] - \theta\}\{1[y_j \le g(x_j,\hat\beta)] - \theta\}. \tag{A.7}$$

(A) Denote

$$W_n(\beta) = \frac{1}{n(n-1)}\sum_{i\ne j}\frac{1}{h^{m_2}}1(x_{1i} = x_{1j})K_2\left(\frac{x_{2i} - x_{2j}}{h}\right) \times \{1[y_i \le g(x_i,\beta)] - \theta\}\{1[y_j \le g(x_j,\beta)] - \theta\}. \tag{A.8}$$

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

We shall first show that uniformly over the $O_p(n^{-1/2})$ neighborhood of β_0, $nh^{m_2/2} \times [W_n(\beta) - W_n(\beta_0)] = o_p(1)$. Then we show that $nh^{m_2/2}W_n(\beta_0)$ has a limiting normal distribution. Thus $nh^{m_2/2}W_n = nh^{m_2/2}W_n(\hat{\beta})$ also has a limiting normal distribution.

Step 1. Uniform convergence of $W_n(\beta)$. Denote $g_i(\beta) \equiv g(x_i, \beta)$ and $F_i(\beta) = F[g_i(\beta)|x_i]$. Denote $Q_n(\beta) = [1/n(n-1)]\sum_{i\neq j} H_n(z_i, z_j, \beta)$, where $z_i = (y_i, x_i')'$ and

$$H_n(z_i, z_j, \beta) \equiv 1(x_{1i} = x_{1j})K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)\{1[y_i \leq g_i(\beta)] - \theta\}\{1[y_j \leq g_j(\beta)] - \theta\}. \tag{A.9}$$

Then $W_n(\beta) = Q_n(\beta)/h^{m_2}$. We first show uniform convergence of $Q_n(\beta)$. We decompose $H_n(z_i, z_j, \beta)$ into three parts:

$$\begin{aligned} H_n(z_i, z_j, \beta) &= 1(x_{1i} = x_{1j})K_2\left(\frac{x_{2i} - x_{2j}}{h}\right) \\ &\quad \times \{1[y_i \leq g_i(\beta)] - F_i(\beta)\}\{1[y_j \leq g_j(\beta)] - F_j(\beta)\} \\ &\quad + 2 \times 1(x_{1i} = x_{1j})K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)\{1[y_i \leq g_i(\beta)] - F_i(\beta)\}\{F_j(\beta) - \theta\} \\ &\quad + 1(x_{1i} = x_{1j})K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)[F_i(\beta) - \theta][F_j(\beta) - \theta] \\ &\equiv H_{1n}(z_i, z_j, \beta) + 2H_{2n}(z_i, z_j, \beta) + H_{3n}(z_i, z_j, \beta). \end{aligned} \tag{A.10}$$

Then let $Q_{sn}(\beta) = [1/n(n-1)]\sum_{i\neq j} H_{sn}(z_i, z_j, \beta)$, $(s = 1,2,3)$. Noting that $\theta = F_i(\beta_0)$, we have $Q_{2n}(\beta_0) = Q_{3n}(\beta_0) = 0$. Here H_{1n} has a constant envelope because K_2 is bounded. Denote $F[\min(g_i(\beta), g_i(\beta_0))|x_i] = F_i(\min(\beta, \beta_0))$. Now we verify Condition (ii) of Lemma 1.

$$\begin{aligned} E[H_1(\cdot, \beta) - H_1(\cdot, \beta_0)]^2 &\leq \Lambda E[\{1[y_i \leq g_i(\beta)] - F_i(\beta)\}\{1[y_j \leq g_j(\beta)] - F_j(\beta)\} \\ &\quad - \{1[y_i \leq g_i(\beta)] - F_i(\beta)\} \\ &\quad \times \{1[y_j \leq g_j(\beta)] - F_j(\beta)\}]^2 \\ &= \Lambda E\{F_i(\beta)[1 - F_i(\beta)]F_j(\beta)[1 - F_j(\beta)] + F_i(\beta_0) \\ &\quad \times [1 - F_i(\beta_0)]F_j(\beta_0)[1 - F_j(\beta_0)] \\ &\quad - 2[F_i(\min(\beta, \beta_0)) - F_i(\beta)F_i(\beta_0)] \\ &\quad \times [F_j(\min(\beta, \beta_0)) - F_j(\beta)F_j(\beta_0)]\} \\ &\leq \Lambda\|\beta - \beta_0)\|, \end{aligned} \tag{A.11}$$

where the last inequality can be shown by Assumptions 1 and 2. Thus $\sup_{B_n} E[H_1(\cdot, \beta) - H_1(\cdot, \beta_0)]^2 = O_p(n^{-1/2})$, and so by Lemma 1,

$$\sup_{B_n} |Q_{1n}(\beta) - Q_{1n}(\beta_0)| \leq O_p(n^{-1-\alpha/4}). \tag{A.12}$$

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

Denote $S_i(\beta) \equiv \partial F[g_i(\beta)]/\partial\beta$ and $T_i(\beta) \equiv \partial^2 F[g_i(\beta)]/\partial\beta\partial\beta'$. For $Q_{2n}(\beta)$, by taking a Taylor expansion of $F_i(\beta)$ around β_0, we have

$$\begin{aligned} Q_{2n}(\beta) &= \frac{1}{n(n-1)} \sum_{i\neq j} 1(x_{1i} = x_{1j}) K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)\{1[y_i \leq g_i(\beta)] - F_i(\beta)\}(\beta - \beta_0)' S_j \\ &\quad + \frac{1}{n(n-1)} \sum_{i\neq j} 1(x_{1i} = x_{1j}) K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)\{1[y_i \leq g_i(\beta)] - F_i(\beta)\} \\ &\quad \times (\beta - \beta_0)' T_j(\bar{\beta})(\beta - \beta_0) \\ &= (\beta - \beta_0)' Q_{2n}^A(\beta) + (\beta - \beta_0)' Q_{2n}^B(\beta)(\beta - \beta_0), \end{aligned} \tag{A.13}$$

where $\bar{\beta}$ is between β_0 and β. By (6) of Sherman (1994), $Q_{2n}^A(\beta)$ can be decomposed as a sum of degenerate U-statistics of order 1 and 2. Then by Lemma 1, $\sup_{B_n}\|Q_{2n}^A(\beta)\| \leq O_p(n^{-1/2-\alpha/4}) + O_p(n^{-1-\alpha/4})$. For $Q_{2n}^B(\beta)$,

$$\begin{aligned} E\left[\sup_{B_n}\|Q_{2n}^B(\beta)\|\right] &\leq \Lambda E\left[\frac{1}{n(n-1)} \sum_{i\neq j} K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)\right] \\ &= \int K_2(u) h^{m_2}\, du = O(h^{m_2}). \end{aligned}$$

Thus,

$$\sup_{B_n}|Q_{2n}(\beta)| \leq O_p(n^{-1-\alpha/4}) + O_p(n^{-1}h^{m_2}). \tag{A.14}$$

For $Q_{3n}(\beta)$, taking a Taylor expansion of $F_i(\beta)$ around β_0, we have

$$\begin{aligned} \sup_{B_n}|Q_{3n}(\beta)| &= \sup_{B_n}\left|\frac{1}{n(n-1)} \sum_{i\neq j} K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)(\beta - \beta_0)' T_i(\beta_1) T_j'(\beta_2)(\beta - \beta_0)\right| \\ &\leq \sup_{B_n} \Lambda\|\beta - \beta_0\|^2 \frac{1}{n(n-1)} \sum_{i\neq j} K_2\left(\frac{x_{2i} - x_{2j}}{h}\right) \\ &= \sup_{B_n} \Lambda\|\beta - \beta_0\|^2 O_p(h^{m_2}) = O_p(n^{-1}h^{m_2}), \end{aligned} \tag{A.15}$$

where both β_1 and β_2 are between β_0 and β. Combining (A.12), (A.14), and (A.15), we have proved

$$\sup_{B_n}|Q_n(\beta) - Q_n(\beta_0)| \leq O_p(n^{-1-\alpha/4}) + O_p(n^{-1}h^{m_2}). \tag{A.16}$$

Thus,

$$\sup_{B_n}|nh^{m_2/2}W_n(\beta) - nh^{m_2/2}W_n(\beta_0)| = O_p(n^{-\alpha/4}h^{-m_2/2}) + O_p(h^{m_2/2}) = o_p(1). \tag{A.17}$$

Step 2. Asymptotic distribution of $W_n(\beta_0)$. We apply Theorem 1 of Hall (1984) on asymptotic distribution of a degenerate U-statistic. Let the kernel of W_n be $H_n(z_i, z_j)$. Under the null, note that $E\{[1(y_i \leq g_i(\beta_0)) - \theta]^2|x_i\} = \theta(1-\theta)$. For $i \neq j, j \neq k, k \neq i$, we have

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

$$E[G_n^2(z_i,z_j)] \equiv E\{E[H_n(z_k,z_i)H_n(z_k,z_j)|z_i,z_j]\}^2$$
$$\leq \frac{\Lambda}{h^{4m_2}} E\left\{E\left[K_2\left(\frac{x_{2k}-x_{2i}}{h}\right)K_2\left(\frac{x_{2k}-x_{2j}}{h}\right)\middle| z_i,z_j\right]\right\}^2$$
$$= \frac{\Lambda}{h^{4m_2}} \iint\left[\int K_2(u)K_2(u+v)h^{m_2}\,du\right]^2 p_2(x_{2i})p_2(x_{2i}-hv)h^{m_2}\,dx_{2i}\,dv$$
$$= O(1/h^{m_2}), \tag{A.18}$$

$$E[H_n^2(z_i,z_j)] = \frac{1}{h^{2m_2}} E\left[1(x_{1i}=x_{1j})K_2^2\left(\frac{x_{2i}-x_{2j}}{h}\right)\theta^2(1-\theta)^2\right]$$
$$= \theta^2(1-\theta)^2\,\frac{1}{h^{2m_2}}$$
$$\times \sum_{x_{1i}\in X_1}\iint K_2^2\left(\frac{x_{2i}-x_{2j}}{h}\right)p_1^2(x_{1i})q(x_{2i}|x_{1i})q(x_{2j}|x_{1i})\,dx_{2i}\,dx_{2j}$$
$$= \theta^2(1-\theta)^2\,\frac{1}{h^{m_2}}\int K_2^2(u)\,du$$
$$\times \sum_{x_{1i}\in X_1}\int p_1^2(x_{1i})q^2(x_{2i}|x_{1i})\,dx_{2i} + o\left(\frac{1}{h^{m_2}}\right)$$
$$= \frac{1}{h^{m_2}}\frac{\sigma^2}{2} + o\left(\frac{1}{h^{m_2}}\right) = O\left(\frac{1}{h^{m_2}}\right), \tag{A.19}$$

$$E[H_n^4(z_i,z_j)] \leq \Lambda E\left[\frac{1}{h^{4m_2}}K_2^4\left(\frac{x_{2i}-x_{2j}}{h}\right)\right]$$
$$= \frac{\Lambda}{h^{4m_2}}\iint K_2^4(u)p_2(x_{2i})p_2(x_{2i}-hu)h^{m_2}\,dx_{2i}\,du = O\left(\frac{1}{h^{3m_2}}\right). \tag{A.20}$$

Thus we have

$$\frac{E[G_n^2(z_i,z_j)] + n^{-1}E[H_n^4(z_i,z_j)]}{\{E[H_n^2(z_i,z_j)]\}^2} \leq \frac{O(1/h^{m_2}) + n^{-1}O(1/h^{3m_2})}{O(1/h^{2m_2})}$$
$$= O(h^{m_2}) + O(1/nh^{m_2}) \to 0, \quad \text{as } n\to\infty, \tag{A.21}$$

because $h\to 0$ and $nh^{m_2}\to\infty$ as $n\to\infty$. Thus the condition of Theorem 1 of Hall (1984) is satisfied. By the theorem we have

$$nh^{m_2/2}W_n(\beta_0) \to_d \mathcal{N}(0,\sigma^2). \tag{A.22}$$

From steps 1 and 2, we have proved that $nh^{m_2/2}W_n \to_d \mathcal{N}(0,\sigma^2)$.

Step 3. Consistency of $\hat{\sigma}^2$. The term $\hat{\sigma}^2$ is a standard U-statistic with kernel $H_n(z_i,z_j) = 2\theta^2(1-\theta)^2(1/h^{m_2})K^2[(x_i-x_j)/h]$. We can easily show that $E[\|H_n(z_i,z_j)\|^2] = O(1/h^{m_2}) = o(n)$. Thus the condition of Lemma 2 is satisfied. By equation (A.19), we have, for sufficiently large n,

$$E(\hat{\sigma}^2) = \sigma^2. \tag{A.23}$$

Therefore by Lemma 2, $\hat{\sigma}^2 = \sigma^2 + o_p(1)$.

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

(B) Similarly to step 1 of the theorem, we can show that $W_n = W_n(\beta^*) + o_p(1)$. Similarly to step 3 of (A), the condition of Lemma 2 can be shown to be satisfied. Under the alternative, we have

$$\begin{aligned}
E[W_n(\beta^*)] &= \frac{1}{h^{m_2}} \sum_{x_{1i} \in X_1} \iint K_2\left(\frac{x_{2i} - x_{2j}}{h}\right)[F(g_i(\beta^*)|x_{1i}, x_{2i}) - \theta] \\
&\quad \times [F(g_j(\beta^*)|x_{1j}, x_{2j}) - \theta] \\
&\quad \times p_1^2(x_{1i}) q(x_{2i}|x_{1i}) q(x_{2j}|x_{1i})\, dx_{2i}\, dx_{2j} + o_p(1) \\
&= \frac{1}{h^{m_2}} \sum_{x_{1i} \in X_1} \iint K_2(u)[F(g(x_{1i}, x_{2i}, \beta^*)|x_{1i}, x_{2i}) - \theta] \\
&\quad \times [F(g(x_{1i}, x_{2i} - hu, \beta^*)|x_{1i}, x_{2i} - hu) - \theta] \\
&\quad \times p_1^2(x_{1i}) q(x_{2i}|x_{1i}) q(x_{2i} - hu|x_{1i})\, dx_{2i} h^{m_2}\, du + o_p(1) \\
&= E\{[F(g(x_i, \beta^*)|x_i) - \theta]^2 p(x_i)\} + o(1). \qquad \textbf{(A.24)}
\end{aligned}$$

Therefore by Lemma 2,

$$W_n = E[W_n(\beta^*)] + o_p(1) = E\{[F(g(x_i, \beta^*)|x_i) - \theta]^2 p(x_i)\} + o_p(1). \qquad \textbf{(A.25)}$$

(C) Similarly to step 1 of (A), we can show that $nh^{m_2/2} W_n = nh^{m_2/2} W_n(\beta_0) + o_p(1)$. Taking a Taylor expansion of $F[(g(x_i, \beta_0) + d_n l(x_i))|x_i]$ around $d_n = 0$, we have

$$\begin{aligned}
F[(g(x_i, \beta_0) + d_n l(x_i))|x_i] &= F(g(x_i, \beta_0)|x_i) \\
&\quad + d_n f(g(x_i, \beta_0)|x_i) l(x_i) + o_p(d_n^2). \qquad \textbf{(A.26)}
\end{aligned}$$

Then $W_n(\beta_0)$ can be decomposed as

$$\begin{aligned}
W_n(\beta_0) &= \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^{m_2}} 1(x_{1i} = x_{1j}) K_2\left(\frac{x_{2i} - x_{2j}}{h}\right) \right. \\
&\quad \times [1(y_i \leq g_i(\beta_0)) - \theta + d_n f(g_i(\beta_0)|x_i) l(x_i)] \\
&\quad \left. \times [1(y_j \leq g_j(\beta_0)) - \theta + d_n f(g_j(\beta_0)|x_j) l(x_j)] \right\} \\
&\quad - 2d_n \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^{m_2}} 1(x_{1i} = x_{1j}) K_2\left(\frac{x_{2i} - x_{2j}}{h}\right) \right. \\
&\quad \left. \times [1(y_i \leq g_i(\beta_0)) - \theta + d_n f(g_i(\beta_0)|x_i) l(x_i)] f(g_j(\beta_0)|x_j) l(x_j) \right\} \\
&\quad + d_n^2 \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^{m_2}} 1(x_{1i} = x_{1j}) K_2\left(\frac{x_{2i} - x_{2j}}{h}\right) \right. \\
&\quad \left. \times f(g_i(\beta_0)|x_i) l(x_i) f(g_j(\beta_0)|x_j) l(x_j) \right\} + o_p(d_n^2) \\
&\equiv T_{1n} - 2d_n T_{2n} + d_n^2 T_{3n} + o_p(d_n^2). \qquad \textbf{(A.27)}
\end{aligned}$$

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press

Similarly to the proof of (A), we can show that $nh^{m_2/2}T_{1n} \to_d \mathcal{N}(0, \sigma^2)$. Applying Lemma 2 we can show that $T_{2n} = O_p(n^{-1/2})$ and

$$T_{3n} \to_p E[f^2(g_i(\beta_0)|x_i)l^2(x_i)p(x_i)]. \tag{A.28}$$

Thus, if $d_n = n^{-1/2}h^{-m_2/4}$, then

$$nh^{m_2/2}W_n \to_d \mathcal{N}(\mu, \sigma^2), \tag{A.29}$$

where $\mu = E[f^2(g_i(\beta_0)|x_i)l^2(x_i)p(x_i)]$. ■

https://doi.org/10.1017/S0266466698141051 Published online by Cambridge University Press