ELSEVIER

Journal of Econometrics 75 (1996) 263–289

JOURNAL OF
Econometrics

A consistent test of functional form via nonparametric estimation techniques

John Xu Zheng

Department of Economics, University of Texas at Austin, Austin, TX 78712, USA

(Received December 1992; final version received April 1995)

Abstract

This paper presents a consistent test of functional form of nonlinear regression models. The test combines the methodology of the conditional moment test and nonparametric estimation techniques. Using degenerate and nondegenerate U-statistic theories, the test statistic is shown to be asymptotically distributed standard normal under the null hypothesis that the parametric model is correct, while diverging to infinity at a rate arbitrarily close to n, the sample size, if the parametric model is misspecified. Therefore, the test is consistent against all deviations from the parametric model. The test is robust to heteroskedasticity. A version of the test can be constructed which will have asymptotic power equal to 1 against any local alternatives approaching the null at rates slower than the parametric rate $1/\sqrt{n}$. A simulation study reveals that the test has good finite-sample properties.

Key words: Specification test; Consistent test; Conditional moment test; Kernel estimation; Degenerate and nondegenerate U-statistics; Local alternatives
JEL classification: C12; C14; C21

1. Introduction

Specification testing of function form is one of the most important problems in econometrics. Following the work of Hausman (1978), the research on this area has been growing. The work includes Ruud (1984), Newey (1985a, 1985b),

This paper is part of my Ph.D. dissertation at Princeton University. I am indebted to Whitney Newey and especially Jim Powell for invaluable advice. I am also grateful to an associate editor, referees, David Card, and Maxwell Stinchcombe for many useful comments. The work was supported by the National Science Foundation through grant SES 91-96185.

SSDI 0304-4076(95)01760-B

Tauchen (1985), White (1982, 1987), Bierens (1990), and many others. Most of the specification tests can be put in the framework of Newey (1985a) and Tauchen's (1985) conditional moment test. However, most of the tests have the drawback of being not consistent against general alternatives or an infinite-dimensional alternative since they are designed to test a parametric null against some parametric alternatives or finite-dimensional alternative. For example, Hausman's (1978) test and White's (1982) information matrix test are among many of the tests that are not consistent against all deviations from the parametric model. Among the tests mentioned above, the only consistent model specification test is Bierens' test which is based on Newey and Tauchen's conditional moment test. Bierens uses a family of exponential functions to generate an infinite number of moment conditions required for the consistency of the conditional moment test. But calculation of Bierens' test statistic requires computing a maximum over an infinite set and this can impose a major computational burden in practice. To overcome the problem, he proposes drawing a sequence of elements from the infinite set at random, then calculating the maximum. Thus the implementation of his test relies on arbitrary selection of moment conditions. Based on different choices of these conditions, different researchers may reach different conclusions on whether a model should be accepted.

Recent developments in nonparametric methods offer powerful tools to tackle the inconsistency problem of earlier specification tests. To obtain a consistent test, we may estimate the infinite-dimensional alternative or true model by nonparametric methods and compare the nonparametric model with the parametric model. The work along this line includes Lee (1988), Yatchew (1992), Eubank and Spiegelman (1990), Wooldridge (1992), and Härdle and Mammen (1993). Both Lee's and Yatchew's tests are based on comparing the parametric sum of squared residuals with the nonparametric sum of squared residuals. However, Lee's procedure is not robust to heteroskedasticity. Yatchew's approach relies on sample splitting and it also assumes homoskedasticity of the error term. Eubank and Spiegelman obtain their test by fitting a spline smoothing to the residuals from a linear regression. However, their test is limited by the normality assumption on the error term. Wooldridge's test is based on Davidson–MacKinnon's (1981) residual-based test and sieve estimation of the infinite-dimensional alternative. His test requires that the alternative models be nonnested. Härdle and Mammen's test is based on the integrated squared difference between the parametric fit and the nonparametric fit. The power of the test against fixed alternatives is not investigated.

In this paper, we propose a test that combines the idea of the conditional moment test and the methodology of nonparametric estimations. We use the kernel method to construct a moment condition which can be used to distinguish the null from the alternative. The test is more powerful than Bierens' consistent conditional moment test and most of the nonparametric tests since the rate at which our test diverges to infinity under the alternative can be constructed to be arbitrarily close to n, the sample size, which is faster than $\sqrt{n}$ achieved by these tests. Another advantage of the test over Bierens' test is that the computation of the test is simpler and only one parameter, the bandwidth, needs to be chosen. The literature on bandwidth selection can shed some light on selection of the parameter in a finite sample. The test has another advantage over the tests based on measuring distance between a parametric model and a nonparametric model in that it does not have the drawbacks of those tests mentioned earlier and it imposes very few regularity conditions beyond those commonly imposed on nonlinear least squares and kernel functions. For example, neither higher-order kernel functions nor trimming on the boundary of a density function, used in many applications of nonparametric regressions, are needed. The bandwidth condition imposed in this paper is no stronger than the condition sufficient for consistency in quadratic mean of kernel density estimators.

The plan of the paper is follows. Section 2 states the testing problem and presents the test statistic. In Section 3, using degenerate and nondegenerate U-statistic theories, we show that the test statistic is asymptotically standard normal under the null hypothesis that the parametric model is correct and tends to infinity in probability under the alternatives. Therefore, the test is consistent. In Section 4, we analyze the power of the test against local misspecifications. We show that the test has power 1 against local misspecifications approaching the null at rates slower than the parametric rate $1/\sqrt{n}$. It can no longer distinguish the null from the alternatives converging to the null at rates faster than or equal to the parametric rate $1/\sqrt{n}$. Section 5 contains some Monte Carlo results. Section 6 summarizes the paper.

2. The hypothesis and test statistic

We have observations $\{(x_i, y_i)\}_{i=1}^n$ where x_i is a $m \times 1$ vector and y_i is a scalar. If $\mathrm{E}[|y_i|] < \infty$, then there exists a Borel measurable function g such that $\mathrm{E}(y_i \mid x_i = x) = g(x)$ where $x \in R^m$.

In a parametric regression model, $g(x)$ is assumed to belong to a parametric family of known real functions $f(x, \theta)$ on $R^m \times \Theta$ where $\Theta \subset R^l$. To justify the use of a parametric model, a specification test is needed. Thus, the null hypothesis to be tested is that the parametric model is correct:

$$\mathrm{H}_0\colon \quad \Pr[\mathrm{E}(y_i \mid x_i) = f(x_i, \theta_0)] = 1 \quad \text{for some} \quad \theta_0 \in \Theta, \tag{2.1}$$

while, without a specific alternative model, the alternative to be tested will be that the null is false:

$$\mathrm{H}_1\colon \quad \Pr[\mathrm{E}(y_i \mid x_i) = f(x_i, \theta)] < 1 \quad \text{for all} \quad \theta \in \Theta, \tag{2.2}$$

where θ_0 is defined as $\theta_0 = \arg\min_{\theta \in \Theta} \mathrm{E}[y_i - f(x_i, \theta)]^2$. Thus the alternative encompasses all the possible departures from the null model. A test that has asymptotic power equal to 1 is said to be consistent.

The idea of our test is as follows. Denote $\varepsilon_i \equiv y_i - f(x_i, \theta_0)$ and let $p(\cdot)$ be the density function of x_i. Then under H_0, since $\mathrm{E}(\varepsilon_i \mid x_i) = 0$, we have

$$\mathrm{E}[\varepsilon_i \, \mathrm{E}(\varepsilon_i \mid x_i) p(x_i)] = 0, \tag{2.3}$$

while under H_1, since $\mathrm{E}(\varepsilon_i \mid x_i) = g(x_i) - f(x_i, \theta_0)$, we have

$$\begin{aligned} \mathrm{E}[\varepsilon_i E(\varepsilon_i \mid x_i) p(x_i)] &= \mathrm{E}\{[\mathrm{E}(\varepsilon_i \mid x_i)]^2 \, p(x_i)\} \\ &= \mathrm{E}\{[g(x_i) - f(x_i, \theta_0)]^2 \, p(x_i)\} \\ &> 0. \end{aligned} \tag{2.4}$$

Therefore, we may use the sample analogue of $\mathrm{E}[\varepsilon_i \, \mathrm{E}\,(\varepsilon_i \mid x_i) \, p(x_i)]$ to form a test.

The test may be understood in light of the Newey and Tauchen's conditional moment test with a single weighting function

$$w(x_i) = \mathrm{E}(\varepsilon_i \mid x_i) p(x_i). \tag{2.5}$$

As in Powell, Stock, and Stoker (1989), the inclusion of the density function avoids the problem of trimming the small values of the density function commonly used in applications of kernel regressions.

The unknown functions g and p can be estimated by various nonparametric methods. Here, we use analytically simpler kernel regression and density methods to estimate g (cf. Härdle, 1990) and p (cf. Silverman, 1986). A kernel estimator of the regression function $\mathrm{E}(\varepsilon_i \mid x_i)$ can be written in the form

$$\hat{\mathrm{E}}(\varepsilon_i \mid x_i) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_j \Big/ \hat{p}(x_i), \tag{2.6}$$

where $\hat{p}$ is a kernel estimator of the density function of p,

$$\hat{p}(x_i) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right), \tag{2.7}$$

where K is a kernel function, h, depending on sample size n, is a bandwidth parameter. θ_0 can be estimated by any $\sqrt{n}$-consistent method, for example the nonlinear least squares method.

Under some mild regularity conditions (cf. Jennrich, 1969; White, 1981, 1982), the nonlinear least squares estimator $\hat{\theta}$ is a consistent and asymptotically normally distributed estimator of θ_0 even in the presence of model misspecification. Replacing ε_i by $e_i \equiv y_i - f(x_i, \hat{\theta})$, we have the sample analogue[1] of $\mathrm{E}[\varepsilon_i \mathrm{E}(\varepsilon_i \mid x_i) p(x_i)]$,

$$V_n \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) e_i e_j. \tag{2.8}$$

For any $n_1 \times n_2$ matrix $A = (a_{ij})$, let $\|A\|$ denote its Euclidean norm, i.e., $\|A\| = [\mathrm{tr}(AA')]^{1/2}$. The following regularity assumptions are sufficient for obtaining the asymptotic distributions of V_n under both the null and the alternative.

Assumption 1. $\{(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)\}$ *is a random sample from a probability distribution* $F(y, x)$ *on* $R \times R^m$. *The density function* $p(x)$ *of* x_i *and its first-order derivatives are uniformly bounded.* $\mathrm{E}(y_i^4 \mid x_i)$ *is continuously differentiable and bounded by a measurable function* $b(x)$ *such that* $\mathrm{E}[b^2(x_i)] < \infty$.

Assumption 2. The parameter space Θ *is a compact and convex subset of* R^l. $f(x, \theta)$ *is a Borel measurable function on* R^m *for each* θ *and a twice continuously differentiable real function on* Θ *for each* $x \in R^m$. *Moreover,*

$$\mathrm{E}\left[\sup_{\theta \in \Theta} |f^2(x_i, \theta)|\right] < \infty,$$

$$\mathrm{E}\left[\sup_{\theta \in \Theta} \left\| \frac{\partial f(x_i, \theta)}{\partial \theta} \cdot \frac{\partial f(x_i, \theta)}{\partial \theta'} \right\|\right] < \infty,$$

$$\mathrm{E}\left[\sup_{\theta \in \Theta} \left\| (y_i - f(x_i, \theta))^2 \cdot \frac{\partial f(x_i, \theta)}{\partial \theta} \cdot \frac{\partial f(x_i, \theta)}{\partial \theta'} \right\|\right] < \infty,$$

$$\mathrm{E}\left[\sup_{\theta \in \Theta} \left\| (y_i - f(x_i, \theta))^2 \cdot \frac{\partial^2 f(x_i, \theta)}{\partial \theta \partial \theta'} \right\|\right] < \infty.$$

Assumption 3. $\mathrm{E}[(y_i - f(x_i, \theta))^2]$ *takes a unique minimum at* $\theta_0 \in \Theta$. *Under* H_0, θ_0 *is an interior point of* Θ.

Assumption 4. The matrix

$$\mathrm{E}\left[\frac{\partial f(x_i, \theta_0)}{\partial \theta} \cdot \frac{\partial f(x_i, \theta_0)}{\partial \theta'}\right]$$

is nonsingular.

---

[1] A referee suggested the statistic V_n as an alternative to an earlier, more complicated (but asymptotically equivalent) statistic. The same statistic V_n is also independently proposed in Zheng (1993).

Assumption 5. *$K(u)$ is a nonnegative, bounded, continuous, and symmetric function such that $\int K(u)\,\mathrm{d}u = 1$.*

Assumptions 1–4 are essentially the same as assumptions used by Bierens (1990, App. A) and standard for ensuring the consistency and asymptotic normality of nonlinear least squares. The kernel function in Assumption 5 is the most commonly used one in nonparametric literature.

3. The limiting distributions of the test statistic

Our statistic can be approximated by a standard one-sample 'second-order' U-statistic. The general 'second-order' U-statistic is of the form

$$U_n \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} H_n(z_i, z_j) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} H_n(z_i, z_j), \tag{3.1}$$

where $\{z_i\}_{i=1}^n$ is an i.i.d. random sample and H_n is any function symmetric in its arguments, i.e., $H_n(z_i, z_j) = H_n(z_j, z_i)$. For the statistic V_n, the function H_n is

$$H_n = \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) e_i e_j.$$

Define

$$r_n(z_i) \equiv \mathrm{E}[H_n(z_i, z_j) \mid z_i], \tag{3.2}$$

$$\bar{r}_n \equiv \mathrm{E}[r_n(z_i)] = \mathrm{E}[H_n(z_i, z_j)], \tag{3.3}$$

$$\hat{U}_n \equiv \bar{r}_n + \frac{2}{n} \sum_{i=1}^{n} [r_n(z_i) - \bar{r}_n], \tag{3.4}$$

where we assume that $\bar{r}_n$ exists. $\hat{U}_n$ is called the 'projection' of the statistic U_n (cf. Hoeffding, 1948). Since $\hat{U}_n$ is an average of independent random variables, its asymptotic distribution can be easily obtained by applying central limit theorems and laws of large numbers. If $\mathrm{E}[\|H_n(z_i, z_j)\|^2] = \mathrm{o}(n)$, then by Lemma 3.1 of Powell, Stock, and Stoker (1989), we have $\sqrt{n}(U_n - \hat{U}_n) = \mathrm{o}_p(1)$. Since the projection $\hat{U}_n$ is a sample average, standard calculations show that $\hat{U}_n - \bar{r}_n$ converges to zero in mean squares.

Summarizing the above results, we have:

Lemma 3.1. *If $\mathrm{E}[\|H_n(z_i, z_j)\|^2] = \mathrm{o}(n)$, then*

$$\sqrt{n}(U_n - \hat{U}_n) = \mathrm{o}_p(1) \quad \textit{and} \quad U_n = \bar{r}_n + \mathrm{o}_p(1).$$

The above lemma is useful if $\mathrm{E}[H_n(z_i, z_j) \mid z_i] \neq 0$ or the U-statistic is non-degenerate. A U-statistic is said to be degenerate if $\mathrm{E}[H_n(z_i, z_j)|z_i] = 0$, almost surely, for $i \neq j$. For a one-dimensional degenerate U-statistic, denote

$$G_n(z_1, z_2) = \mathrm{E}[H_n(z_3, z_1)\, H_n(z_3, z_2) \mid z_1, z_2]. \tag{3.5}$$

Applying Theorem 1 of Hall (1984), we obtain the asymptotic distribution of a degenerate U-statistic.

Lemma 3.2. *Assume* $\mathrm{E}[H_n(z_1, z_2) \mid z_1] = 0$ *almost surely and* $\mathrm{E}[H_n^2(z_1, z_2)] < \infty$ *for each n. If*

$$\frac{\mathrm{E}[G_n^2(z_1, z_2)] + n^{-1}\mathrm{E}[H_n^4(z_1, z_2)]}{\{\mathrm{E}[H_n^2(z_1, z_2)]\}^2} \to 0 \quad as \quad n \to \infty, \tag{3.6}$$

then

$$n \cdot U_n / \{2\mathrm{E}[H_n^2(z_i, z_j)]\}^{1/2}$$

has a limiting standard normal distribution.

Applying Lemmas 3.1 and 3.2, we obtain the asymptotic distribution of V_n under the null hypothesis (all proofs are given in the Appendix):

Lemma 3.3. *Given Assumptions 1–5, if* $h \to 0$ *and* $nh^m \to \infty$, *then under the null hypothesis (2.1),*

$$nh^{m/2} V_n \xrightarrow{d} \mathrm{N}(0, \Sigma), \tag{3.7}$$

where Σ *is the asymptotic variance of* $nh^{m/2} V_n$,

$$\Sigma = 2 \int K^2(u)\mathrm{d}u \cdot \int [\sigma^2(x)]^2 p^2(x)\mathrm{d}x. \tag{3.8}$$

Moreover, Σ *can be consistently estimated by* $\hat{\Sigma}$,

$$\hat{\Sigma} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) e_i^2 e_j^2. \tag{3.9}$$

The condition nh^m places an upper bound on the rate at which the bandwidth h converges to 0. The bandwidth condition turns out to be the same as one used by Prakasa Rao (1983, Thm. 3.1.2, p. 181) for obtaining the consistency in quadratic mean of kernel density estimators.

Finally, define a standardized version of the test statistic T_n as

$$T_n \equiv \sqrt{\frac{n-1}{n}} \cdot \frac{nh^{m/2}V_n}{\sqrt{\hat{\Sigma}}}$$

$$= \frac{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} K\left(\frac{x_i - x_j}{h}\right) e_i e_j}{\left\{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} 2K^2\left(\frac{x_i - x_j}{h}\right) e_i^2 e_j^2\right\}^{1/2}}. \tag{3.10}$$

The asymptotic distribution of T_n under the null then follows from Lemma 3.3.

Theorem 1. Given Assumptions 1–5, if $h \to 0$ and $nh^m \to \infty$, then under the null hypothesis (2.1),

$$T_n \xrightarrow{d} \mathrm{N}(0, 1). \tag{3.11}$$

Theorem 1 could be used to calculate the asymptotic critical value for our test. To know the power and consistency of the test, we next obtain the asymptotic distribution of the test statistic T_n under a fixed alternative hypothesis.

Applying Lemma 3.1, we obtain the asymptotic distribution of V_n under the alternative.

Lemma 3.4. Given Assumptions 1–5, if $h \to 0$ and $nh^m \to \infty$, then under the alternative hypothesis (2.2),

$$V_n \xrightarrow{p} \mathrm{E}\{[g(x_i) - f(x_i, \theta_0)]^2 p(x_i)\} > 0 \tag{3.12}$$

and

$$\hat{\Sigma} \xrightarrow{p} 2\int K^2(u)\,\mathrm{d}u \cdot \int \{\sigma^2(x) + [g(x) - f(x, \theta_0)]^2\}^2 p^2(x)\,\mathrm{d}x > 0. \tag{3.13}$$

The asymptotic distribution of the test statistic T_n under the alternative then follows.

Theorem 2. Given Assumptions 1–5, if $h \to 0$ and $nh^m \to \infty$, then under the alternative hypothesis (2.2),

$$T_n / nh^{m/2} \xrightarrow{p} \frac{\mathrm{E}\{[g(x_i) - f(x_i, \theta_0)]^2 p(x_i)\}}{\{2\int K^2(u)\mathrm{d}u \cdot \int \{\sigma^2(x) + [g(x) - f(x, \theta_0)]^2\}^2 p^2(x)\mathrm{d}x\}^{1/2}} > 0. \tag{3.14}$$

Thus $T_n \to \infty$ in probability and the asymptotic power of the test is 1. Since $nh^{m/2}/\sqrt{n} = \sqrt{nh^m} \to \infty$, the convergence rate $nh^{m/2}$ of T_n going to infinity is faster than those obtained by Bierens (1990) and Wooldridge (1992) which are $\sqrt{n}$.[2] Our convergence rate can be made arbitrarily close to n by letting h approach to zero slowly. The convergence rate of Eubank and Spiegelman's (1990) test can also be made arbitrarily close to n. The same normalized factor $nh^{m/2}$ used by Härdle and Mammen's (1993) suggests that their test should have the same convergence rate as one in this paper. Thus Eubank and Spiegelman's test, Härdle and Mammen's test, and our test should be more powerful in large samples than those of Lee, Yatchew, and Wooldridge. Comparing with Härdle and Mammen's test, our test is easier to compute. To apply their test one needs to calculate the integration and estimate its asymptotic mean. In our case, the asymptotic mean is zero.

Though the motivation of the test proposed here is very different from other tests, it turns out that there are some interesting connections among those tests. Lee and Yatchew's procedures are based on the sum of squared residuals from the parametric model $SSR_P = \sum_{i=1}^{n}[y_i - f(x_i, \hat{\theta})]^2/n$ and the sum of squared residuals from the nonparametric model $SSR_N = \sum_{i=1}^{n}[y_i - \hat{g}(x_i)]^2/n$, where $\hat{g}(x_i)$ is a nonparametric estimator of g. Wooldridge's procedure is based on the statistic $W_n = \sum_{i=1}^{n}[\hat{g}(x_i) - f(x_i, \hat{\theta})][y_i - f(x_i, \hat{\theta})]/n$. Since

$$\begin{aligned} SSR_P - SSR_N &= \frac{1}{n}\sum_{i=1}^{n}[y_i - f(x_i, \hat{\theta})]^2 - \frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{g}(x_i)]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}[\hat{g}(x_i) - f(x_i, \hat{\theta})][y_i - f(x_i, \hat{\theta})] \qquad (3.15) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}[\hat{g}(x_i) - f(x_i, \hat{\theta})][y_i - \hat{g}(x_i)], \end{aligned}$$

Lee and Yatchew's procedure differs from Wooldridge's test by its inclusion of the second term in the above equation, which converges to zero under both the null and the alternative.[3] If we look at the density weighted version of those tests, we can also see some relations between our test and Wooldridge's test, and thus Lee and Yatchew's tests. Denote

$$\hat{g}(x_i) = \frac{1}{n-1}\sum_{\substack{j=1 \\ j \neq i}}^{n}\left(\frac{1}{h}\right)^m K\left(\frac{x_i - x_j}{h}\right) y_j \Big/ \hat{p}(x_i), \tag{3.16}$$

[2] The same $\sqrt{n}$ convergence rate is expected from tests of Lee (1988) and Yatchew (1992), though no proofs are given by them.

[3] I thank a referee for pointing out the connection.

and the smoothed version of $f(x, \hat{\theta})$ by $\hat{f}(x_i, \hat{\theta})$,

$$\hat{f}(x_i, \hat{\theta}) \equiv \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\frac{1}{h}\right)^m K\left(\frac{x_i - x_j}{h}\right) f(x_j, \hat{\theta}) \Big/ \hat{p}(x_i), \tag{3.17}$$

then our test statistic V_n can be rewritten as

$$\begin{aligned} V_n &= \frac{1}{n} \sum_{i=1}^{n} [\hat{g}(x_i) - \hat{f}(x_i, \hat{\theta})][y_i - f(x_i, \hat{\theta})]\,\hat{p}(x_i) \\ &= \left\{\frac{1}{n} \sum_{i=1}^{n} [\hat{g}(x_i) - f(x_i, \hat{\theta})][y_i - f(x_i, \hat{\theta})]\hat{p}(x_i)\right\} \\ &\quad + \left\{\frac{1}{n} \sum_{i=1}^{n} [f(x_i, \hat{\theta}) - \hat{f}(x_i, \hat{\theta})][y_i - f(x_i, \hat{\theta})]\hat{p}(x_i)\right\}. \end{aligned} \tag{3.18}$$

The first term in the above equation is the density-weighted version of Wooldridge's test statistic, while the second term converges in probability to zero under both the null and the alternative. There is also a connection between our test and Härdle and Mammen's test. Their test is based on the statistic

$$\int [\hat{g}(x) - \hat{f}(x, \hat{\theta})]^2\, \pi(x)\,\mathrm{d}x,$$

where π is a weighting function.

Our test statistic and Wooldridge's test statistic are both degenerate under the null. The degeneracy may cause the null distribution of the test statistics to be ill-behaved under the usual $\sqrt{n}$ normalization. Despite the close connections among various tests, our test turns out to be more powerful than Wooldridge, Lee, and Yatchew's tests. This is because our test exploits the degeneracy property of the test statistics under the null for our advantage while others avoid this problem through different bias control methods. Wooldridge avoids this by properly controlling the number of series terms used in a sieve estimation of the alternative model. Yatchew avoids the degeneracy problem by splitting the sample into two parts, one for calculating SSR_P and another for SSR_N.

4. Tests of local alternatives

In Section 3, we have shown that our test is consistent against all fixed alternatives. It would be interesting to know how the test behaves under the local alternatives. To investigate the power of a test, classical tests usually consider local misspecifications converging to the null at the parametric rate $1/\sqrt{n}$, the familiar Pitman (1949) drift.

We consider a sequence of local alternatives

$$\mathrm{H}_{1n}: \quad \mathrm{E}(y_i \mid x_i) = f(x_i, \theta_0) + \delta_n \cdot l(x_i), \tag{4.1}$$

where the known function $l(\cdot)$ is continuously differentiable and bounded by the measurable function $b(\cdot)$ in Assumption 1, and $\delta_n \to 0$ as $n \to \infty$.

The following theorem gives the power of our test against local misspecifications (4.1):

Theorem 3. Given Assumptions 1–5, if $h \to 0$ and $nh^m \to \infty$, then under the local alternatives (4.1), if $\delta_n = n^{-1/2}h^{-m/4}$,

$$T_n \xrightarrow{d} \mathrm{N}(\mu, 1), \tag{4.2}$$

where

$$\mu = \mathrm{E}[l^2(x_i)p(x_i)]/\sqrt{\Sigma},$$

where Σ is given in Lemma 3.3.

Tests of local misspecifications are also considered by Eubank and Spiegelman (1990) and Härdle and Mammen (1993). Eubank and Spiegelman show that their test has some power detecting local alternatives approaching to the null at rate $n^{-1/2}D^{1/4}$, where D is the dimension of the departure of the alternative from the null. Härdle and Mammen obtain similar local properties as ours.

It should be clear that the local convergence rate δ_n which those tests have power to detect is the square root of the rate at which the test statistics converge to infinity under the alternative. Since Wooldridge's test statistic converges to infinity at rate $\sqrt{n}$, it has only power to detect local alternatives approaching to the null at rates slower or equal to $n^{-1/4}$. Thus our test is also more powerful than Wooldridge's test in detecting local misspecifications.[4] Note that none of these tests, including ours, can detect $1/\sqrt{n}$-local alternatives.

5. A Monte Carlo study

To investigate how the test proposed in the paper behaves in finite samples, we conduct a Monte Carlo simulation of the size and power of the test.

The data is generated as follows. Let z_{1i} and z_{2i} be independent number drawing from the standard normal distribution. Two regressors, x_1 and x_2, are defined as

$$x_{1i} = z_{1i}, \qquad x_{2i} = (z_{1i} + z_{2i})/\sqrt{2}. \tag{5.1}$$

The error term ε_i is also drawn independently from the standard normal distribution. x_{1i}, x_{2i}, and ε_i all have the same variance 1.

[4] I thank a referee for pointing this out.

The null hypothesis we want to test is that the linear model is correct:

$$H_0:\quad E(y_i \mid x_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} \quad \text{for some} \quad \alpha = (\alpha_0, \alpha_1, \alpha_2)' \in R^3. \tag{5.2}$$

In the four models considered, the dependent variable y is generated as follows. To investigate the size of the test, we consider model 1, where the dependent variable y is generated to be

$$y_i = 1 + x_{1i} + x_{2i} + \varepsilon_i. \tag{5.3}$$

To see if the test has power to detect high-order terms, we consider model 2, where it adds the interaction term $x_1 x_2$ into model 1,

$$y_i = 1 + x_{1i} + x_{2i} + x_{1i}x_{2i} + \varepsilon_i. \tag{5.4}$$

To investigate the power of the test against a general nonlinear regression model, we consider a concave and a convex alternative, models 3 and 4. In model 3, the dependent variable y is defined as

$$y_i = (1 + x_{1i} + x_{2i})^{1/3} + \varepsilon_i. \tag{5.5}$$

In model 4, the dependent variable y is

$$y_i = (1 + x_{1i} + x_{2i})^{5/3} + \varepsilon_i. \tag{5.6}$$

Obviously, H_0 is true for model 1 and false for models 2, 3, and 4. The simulation is conducted to sample sizes 100, 200, ..., 700. Each experiment is based on 1000 replications. The kernel function K is chosen to be the bivariate standard normal density function

$$K(u_1, u_2) = \frac{1}{2\pi}\exp\left(-\frac{u_1^2 + u_2^2}{2}\right). \tag{5.7}$$

The bandwidth h is chosen to be $c \cdot n^{-2/5}$ where c is a constant. The bandwidth satisfies $nh^m \to \infty$ where $m = 2$ in this case. To investigate whether the test is sensitive to the choice of bandwidth, we calculate the test statistic for c equal to 0.5, 1.0, and 2.0. The critical values for the test are from the standard normal table: for 1% significance level the critical value is 2.576, for 5% level the critical value is 1.960, and for 10% level it is 1.645.

The results of the size study are shown in Table 1. As can be seen, the test has adequate size in most cases. The sizes get closer to the asymptotic sizes when n gets larger. The size seems not very sensitive to the choice of bandwidth.

The results of the power study are presented in Tables 2, 3, and 4. The powers of the test are very high in most cases and quickly converge to 1. The power increases with sample size in all cases for any chosen bandwidth.

To know better how the test behaves in finite samples, we compare the test with a classical test, the t-test. To illustrate this, we test the linear model (5.2)

Table 1
Proportion of rejections in model 1: $y_i = 1 + x_{1i} + x_{2i} + \varepsilon_i$

Parameter c	Sample size	1% significance	5% significance	10% significance
	100	0.2	3.4	8.2
	200	0.3	4.2	9.2
	300	0.5	4.7	10.3
0.5	400	0.5	5.2	10.5
	500	0.7	4.8	9.6
	600	0.9	4.8	10.8
	700	1.0	5.4	11.9
	100	0.6	3.9	8.9
	200	0.4	4.4	9.7
	300	0.5	4.6	10.5
1.0	400	1.2	5.1	9.7
	500	0.9	4.1	8.4
	600	0.7	4.3	9.3
	700	1.6	5.4	10.8
	100	0.5	2.4	7.5
	200	0.3	2.9	8.1
	300	0.4	3.7	9.5
2.0	400	0.6	3.4	7.9
	500	0.5	3.3	8.7
	600	0.2	3.1	8.1
	700	1.1	4.2	9.4
	100	5.4	8.0	11.2
	200	4.4	6.5	10.2
	300	2.7	4.6	7.7
t-test	400	4.8	7.1	9.9
	500	2.9	5.6	7.9
	600	3.7	5.6	9.4
	700	2.6	5.0	7.2

against a high-order linear model,

$$\mathrm{H}_1:\quad \mathrm{E}(y_i \mid x_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{1i} x_{2i}$$
$$\text{for some}\quad \alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3) \in R^4. \tag{5.8}$$

In other words, we test if $\alpha_3 = 0$ in (5.8). Denote the OLS estimator of α by $\hat{\alpha}$. The asymptotic variance–covariance matrix Ω of $\hat{\alpha}$ can be estimated by White's (1980) heteroskedasticity-robust covariance matrix $\hat{\Omega} \equiv (\omega_{i,j})$,

$$\hat{\Omega} = \left(\sum_{i=1}^{n} x_i' x_i\right)^{-1} \left(\sum_{i=1}^{n} (y_i - x_i'\hat{\alpha})^2\, x_i' x_i\right) \left(\sum_{i=1}^{n} x_i' x_i\right)^{-1} \tag{5.9}$$

Table 2
Proportion of rejections in model 2: $y_i = 1 + x_{1i} + x_{2i} + x_{1i}x_{2i} + \varepsilon_i$

Parameter c	Sample size	1% significance	5% significance	10% significance
	100	25.2	50.0	64.9
	200	64.2	84.8	90.3
	300	88.6	97.0	98.7
0.5	400	96.4	99.2	99.5
	500	99.5	99.8	99.8
	600	99.8	99.9	100
	700	99.8	100	100
	100	80.2	90.4	94.2
	200	98.8	99.8	99.9
	300	100	100	100
1.0	400	100	100	100
	500	100	100	100
	600	100	100	100
	700	100	100	100
	100	97.7	99.1	99.7
	200	100	100	100
	300	100	100	100
2.0	400	100	100	100
	500	100	100	100
	600	100	100	100
	700	100	100	100
	100	97.1	97.7	98.1
	200	99.7	99.7	99.7
	300	99.8	100	100
t-test	400	100	100	100
	500	100	100	100
	600	100	100	100
	700	100	100	100

Hence we can use the t-statistic $t = \sqrt{n}\hat{\alpha}_3/\hat{\omega}_{4,4}$ to test if $\alpha_3 = 0$. The critical values are the same as in our consistent test. The results of the t-test are given in the lower panels of Table 1 through Table 4. The sizes of the t-test at 1% significance level are very larger than the corresponding sizes of the standard normal distribution. Since the t-test is directed to test model 2, it is not surprising that the t-test reaches its best power in model 2. In general situations, we can expect that our consistent test does better than the t-test. Indeed, our consistent test does better than the t-test in model 3, and it also does better than the t-test in model 4 for sample sizes over 300.

Table 3
Proportion of rejections in model 3: $y_i = (1 + x_{1i} + x_{2i})^{1/3} + \varepsilon_i$

Parameter c	Sample size	1% significance	5% significance	10% significance
	100	3.7	13.6	22.6
	200	10.7	26.0	36.7
	300	21.3	40.1	51.0
0.5	400	33.3	54.2	65.6
	500	42.3	65.3	76.1
	600	53.0	72.0	81.0
	700	63.0	80.6	86.7
	100	17.8	31.0	39.0
	200	42.1	60.5	70.4
	300	65.4	80.7	87.0
1.0	400	83.6	92.9	94.8
	500	93.7	97.8	98.7
	600	96.7	99.2	99.7
	700	98.3	99.7	99.8
	100	36.3	48.5	57.4
	200	79.5	88.8	91.9
	300	95.2	97.9	99.1
2.0	400	99.3	99.7	99.9
	500	100	100	100
	600	100	100	100
	700	100	100	100
	100	45.1	55.9	61.4
	200	43.9	56.6	62.0
	300	41.9	54.5	60.4
t-test	400	43.2	54.2	62.7
	500	44.8	56.2	61.7
	600	43.5	53.3	62.8
	700	46.6	59.7	66.5

6. Conclusion

This paper has provided a consistent specification test of parametric nonlinear regression models against general alternatives. The idea of the test exploits the close connection between the conditional moment test and nonparametric tests. The test is more powerful than the Bierens' (1990) consistent test and many of the nonparametric tests. The test also has more power than many alternative tests in detecting local misspecifications. The test is shown to be consistent against local alternatives approaching the null at rates slower than the parametric rate $1/\sqrt{n}$. The test can no longer distinguish the local alternatives from the

Table 4
Proportion of rejections in model 4: $y_i = (1 + x_{1i} + x_{2i})^{5/3} + \varepsilon_i$

Parameter c	Sample size	1% significance	5% significance	10% significance
	100	69.0	87.5	92.5
	200	96.8	99.1	99.3
	300	99.9	100	100
0.5	400	99.9	100	100
	500	100	100	100
	600	100	100	100
	700	100	100	100
	100	98.5	99.3	99.6
	200	100	100	100
	300	100	100	100
1.0	400	100	100	100
	500	100	100	100
	600	100	100	100
	700	100	100	100
	100	99.7	100	100
	200	100	100	100
	300	100	100	100
2.0	400	100	100	100
	500	100	100	100
	600	100	100	100
	700	100	100	100
	100	78.3	82.9	85.5
	200	76.8	81.8	84.5
	300	74.2	79.3	82.3
t-test	400	75.0	80.7	83.9
	500	73.7	79.5	83.3
	600	71.5	77.6	81.1
	700	73.5	79.1	81.7

null if their rate of convergence to zero is at least $1/\sqrt{n}$. The test is also easier to compute than Bierens' consistent moment test. The simulation results demonstrate that the test has adequate size and the size is not very sensitive to the choices of the bandwidth. In most cases, the power is very high and approaches to one quickly as sample size increases. But a further extensive study on the choice of bandwidth is necessary and this is left for future research.

The test can be applied to testing the underlying distribution assumptions in the parametric binary choice, censored regression, truncated regression, and sample selection models. The testing procedure can be immediately applied to testing omitted variables. It may be extended to testing a semiparametric model against a nonparametric model, where a semiparametric estimator is available.

It is also possible to extend the test to heterogeneous, autocorrelation, and some other time series cases.

Appendix: Proofs

Proof of Lemma 3.3

The proof of Lemma 3.3 is broken into proofs of Lemma 3.3a through Lemma 3.3e.

The statistic V_n can be decomposed into three parts:

$$
\begin{aligned}
V_n &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) e_i e_j \\
&= \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i \varepsilon_j \right\} \\
&\quad - 2\left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i [f(x_j, \hat{\theta}) - f(x_j, \theta_0)] \right\} \\
&\quad + \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \right. \\
&\quad \left. \times [f(x_i, \hat{\theta}) - f(x_i, \theta_0)][f(x_j, \hat{\theta}) - f(x_j, \theta_0)] \right\} \\
&\equiv V_{1n} - 2V_{2n} + V_{3n}.
\end{aligned} \tag{A.1}
$$

Under the null, we will show that $nh^{m/2}V_{1n}$ is normally distributed, $nh^{m/2}V_{2n} = o_p(1)$, and $nh^{m/2}V_{3n} = o_p(1)$.

The following lemma establishes the asymptotic distribution of V_{1n} under the null hypothesis:

Lemma 3.3a. Given Assumptions 1–5, if $h \to 0$ and $nh^m \to \infty$, then under the null hypothesis (2.1),

$$nh^{m/2} V_{1n} \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Sigma), \tag{A.2}$$

where Σ is the asymptotic variance of $nh^{m/2}V_{1n}$,

$$\Sigma = 2 \int K^2(u)\, \mathrm{d}u \cdot \int [\sigma^2(x)]^2 p^2(x)\, \mathrm{d}x. \tag{A.3}$$

Proof of Lemma 3.3a

V_{1n} can be written in a U-statistic form with

$$H_n(z_i, z_j) = \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i \varepsilon_j. \tag{A.4}$$

Under null hypothesis, since $\mathrm{E}[H_n(z_1, z_2) | z_1] = (1/h^m)\varepsilon_1 \mathrm{E}[K((x_1 - x_2)/h) \times \mathrm{E}(\varepsilon_2 | x_2)] = 0$, V_{1n} is a degenerate statistic.

To apply Lemma 3.2, we need to verify its condition. Since

$$\begin{aligned}
&\mathrm{E}[G_n^2(z_1, z_2)] \\
&= \mathrm{E}\{\mathrm{E}[H_n(z_3, z_1)\, H_n(z_3, z_2) | z_1, z_2]\}^2 \\
&= \mathrm{E}\left\{\mathrm{E}\left[\frac{1}{h^{2m}} K\left(\frac{x_3 - x_1}{h}\right) K\left(\frac{x_3 - x_2}{h}\right) \varepsilon_1 \varepsilon_2 \varepsilon_3^2 | z_1, z_2\right]\right\}^2 \\
&= \frac{1}{h^{4m}} \mathrm{E}\left\{\varepsilon_1 \varepsilon_2 \mathrm{E}\left[K\left(\frac{x_3 - x_1}{h}\right) K\left(\frac{x_3 - x_2}{h}\right) \sigma^2(x_3) | x_1, x_2\right]\right\}^2 \\
&= \frac{1}{h^{4m}} \mathrm{E}\left\{\varepsilon_1 \varepsilon_2 \int K\left(\frac{x_3 - x_1}{h}\right) K\left(\frac{x_3 - x_2}{h}\right) \sigma^2(x_3)\, p(x_3)\, \mathrm{d}x_3\right\}^2 \\
&= \frac{1}{h^{4m}} \mathrm{E}\left\{\varepsilon_1 \varepsilon_2 \int K(u)\, K\left(u + \frac{x_1 - x_2}{h}\right) \sigma^2(x_1 + hu) p(x_1 + hu)\, h^m \mathrm{d}u\right\}^2 \\
&= \frac{1}{h^{2m}} \mathrm{E}\left\{\mathrm{E}\left\{\varepsilon_1^2 \varepsilon_2^2 \left[\int K(u)\, K\left(u + \frac{x_1 - x_2}{h}\right)\right.\right.\right. \\
&\qquad \left.\left.\left. \times \sigma^2(x_1 + hu)\, p(x_1 + hu)\, \mathrm{d}u\right]^2 \middle| x_1, x_2\right\}\right\} \\
&= \frac{1}{h^{2m}} \mathrm{E}\left\{\sigma^2(x_1)\, \sigma^2(x_2) \left[\int K(u)\, K\left(u + \frac{x_1 - x_2}{h}\right)\right.\right. \\
&\qquad \left.\left. \times \sigma^2(x_1 + hu)\, p(x_1 + hu)\, \mathrm{d}u\right]^2\right\} \\
&= \frac{1}{h^{2m}} \int \sigma^2(x_1)\, \sigma^2(x_2) \left[\int K(u)\, K\left(u + \frac{x_1 - x_2}{h}\right)\right. \\
&\qquad \left. \times \sigma^2(x_1 + hu) p(x_1 + hu)\, \mathrm{d}u\right]^2 p(x_1) p(x_2)\, \mathrm{d}x_1 \mathrm{d}x_2
\end{aligned}$$

$$= \frac{1}{h^{2m}} \int \sigma^2(x_1)\, \sigma^2(x_1 - hv) \left[\int K(u)\, K(u+v) \right.$$

$$\left. \times \sigma^2(x_1 + hu) p(x_1 + hu)\, \mathrm{d}u \right]^2 h^m p(x_1)\, p(x_1 - hv)\, \mathrm{d}x_1 \mathrm{d}v$$

$$= \frac{1}{h^m} \int \left[\int K(u)\, K(u+v)\, \mathrm{d}u \right]^2 \mathrm{d}v \int [\sigma^2(x)]^4 p^4(x) \mathrm{d}x + \mathrm{o}(1/h^m)$$

$$= \mathrm{O}(1/h^m), \tag{A.5}$$

where $\sigma^4(x) = \mathrm{E}[\varepsilon_i^4 | x_i = x]$,

$$\begin{aligned} \mathrm{E}[H_n^2(z_1, z_2)] &= \mathrm{E}\{\mathrm{E}[H_n^2(z_1, z_2) \,|\, x_1, x_2]\} \\ &= \int \frac{1}{h^{2m}} K^2\left(\frac{x_1 - x_2}{h}\right) \sigma^2(x_1)\, \sigma^2(x_2)\, p(x_1)\, p(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\ &= \frac{1}{h^{2m}} \int K^2(u)\, \sigma^2(x)\, \sigma^2(x - hu)\, p(x)\, p(x - hu)\, \mathrm{d}x h^m\, \mathrm{d}u \\ &= \frac{1}{h^m} \int K^2(u)\, \mathrm{d}u \cdot \int [\sigma^2(x)]^2 p^2(x)\, \mathrm{d}x + \mathrm{o}(1/h^m) \qquad \text{(A.6)} \\ &= \mathrm{O}(1/h^m), \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}[H_n^4(z_1, z_2)] &= \int \frac{1}{h^{4m}} K^4\left(\frac{x_1 - x_2}{h}\right) \sigma^4(x_1)\, \sigma^4(x_2)\, p(x_1)\, p(x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \frac{1}{h^{4m}} \int K^4(u)\, \sigma^4(x)\, \sigma^4(x - hu)\, p(x)\, p(x - hu)\, \mathrm{d}x h^m \mathrm{d}u \\ &= \mathrm{O}(1/h^{3m}), \end{aligned} \tag{A.7}$$

we have

$$\begin{aligned} \frac{\mathrm{E}[G_n^2(z_1, z_2)] + n^{-1}\mathrm{E}[H_n^4(z_1, z_2)]}{\{\mathrm{E}[H_n^2(z_1, z_2)]\}^2} &= \frac{\mathrm{O}(1/h^m) + n^{-1}\mathrm{O}(1/h^{3m})}{\mathrm{O}(1/h^{2m})} \\ &= \mathrm{O}(h^m) + \mathrm{O}(1/nh^m) \to 0 \quad \text{as} \quad n \to \infty, \end{aligned} \tag{A.8}$$

since $h \to 0$ and $nh^m \to \infty$ as $n \to \infty$.

Thus the condition in Lemma 3.2 is satisfied and it follows that

$$n \cdot V_{1n} / \{2\mathrm{E}[H_n^2(z_i, z_j)]\}^{1/2} \xrightarrow{\mathrm{d}} \mathrm{N}(0, 1).$$

By (A.6), we have

$$nh^{m/2} V_{1n} \xrightarrow{\mathrm{d}} \mathrm{N}(0, 2 \int K^2(u)\, \mathrm{d}u \cdot \int [\sigma^2(x)]^2 p^2(x)\, \mathrm{d}x). \quad \text{Q.E.D.}$$

Lemma 3.3b. *Given Assumptions 1–5, if $h \to 0$ and $nh^m \to \infty$, then under the null hypothesis (2.1),*

$$W_n \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i M(x_j) = \mathrm{O}_p(1/\sqrt{n}), \tag{A.9}$$

whenever M is continuously differentiable and $\|M(x)\| \leqslant b(x)$ for $x \in R^m$ and $\mathrm{E}[b^2(x_i)] < \infty$.

Proof of Lemma 3.3b

The lemma can be proved by modifying the proof of Theorem 3.1 in Powell, Stock, and Stoker (1989).

W_n can be written in a U-statistic form with

$$H_n(z_i, z_j) = \frac{1}{2h^m} K\left(\frac{x_i - x_j}{h}\right) [\varepsilon_i M(x_j) + \varepsilon_j M(x_i)]. \tag{A.10}$$

To apply Lemma 3.1, we need to verify the condition that $\mathrm{E}[\|H_n(z_i, z_j)\|^2] = \mathrm{o}(n)$. We have

$$\begin{aligned}
&\mathrm{E}[\|H_n(z_i, z_j)\|^2] \\
&\leqslant 2\mathrm{E}\left[\frac{1}{2h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i M(x_j)\right]^2 + 2\mathrm{E}\left[\frac{1}{2h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_j M(x_i)\right]^2 \\
&= \int \frac{1}{h^{2m}} K^2\left(\frac{x_i - x_j}{h}\right) \sigma^2(x_i)\, M^2(x_j) p(x_i) p(x_j)\, \mathrm{d}x_i \mathrm{d}x_j \\
&= \frac{1}{h^{2m}} \int K^2(u)\, \sigma^2(x_i) M^2(x_i - hu) p(x_i) p(x_i - hu)\, \mathrm{d}x_i h^m \mathrm{d}u \\
&= \mathrm{O}(1/h^m) = \mathrm{O}[n(nh^m)^{-1}] = \mathrm{o}(n) \quad \text{since} \quad nh^m \to \infty.
\end{aligned} \tag{A.11}$$

For the above $H_n(z_i, z_j)$, we have $\bar{r}_n = \mathrm{E}[H_n(z_i, z_j)] = 0$,

$$\begin{aligned}
r_n(z_i) &= \frac{1}{2h^m} \int K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i M(x_j)\, p(x_j)\, \mathrm{d}x_j \\
&= \frac{1}{2h^m} \int K(u)\, \varepsilon_i M(x_i - hu) p(x_i - hu)\, h^m \mathrm{d}u \\
&= \tfrac{1}{2} \varepsilon_i M(x_i) p(x_i) + t_n(z_i).
\end{aligned} \tag{A.12}$$

Thus

$$\sqrt{n}\, \hat{U}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i M(x_i)\, p(x_i) + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} t_n(z_i). \tag{A.13}$$

Since $t_n(z_i)$ has second moment equal to $o(h^2)$, the asymptotic distribution of $\sqrt{n}\hat{U}_n$ is the same as that of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i M(x_i)p(x_i). \tag{A.14}$$

By the Lindeberg-Levy central limit theorem, we have

$$\sqrt{n}\,U_n \xrightarrow{d} \mathrm{N}(0, \mathrm{E}[\sigma^2(x_i)M^2(x_i)p^2(x_i)]).$$

Thus $W_n = \mathrm{O}_p(1/\sqrt{n})$. Q.E.D.

From the standard least squares theory (cf Jennrich, 1969; White, 1981, 1982), it follows that:

Lemma 3.3c. Given Assumptions 1-4, under the null (2.1) or alternative (2.2),

$$\sqrt{n}(\hat{\theta}-\theta_0) = \mathrm{O}_p(1). \tag{A.15}$$

Applying Lemmas 3.3b and 3.3c, we have the following result:

Lemma 3.3d. Given Assumptions 1-5, if $h \to 0$ and $nh^m \to \infty$, then under the null hypothesis (2.1),

$$nh^{m/2}V_{2n} \xrightarrow{p} 0 \quad \textit{and} \quad nh^{m/2}V_{3n} \xrightarrow{p} 0. \tag{A.16}$$

Proof of Lemma 3.3d

V_{2n} can also be approximated by U-statistics. To see this, note that

$$\begin{aligned} V_{2n} &= \left\{\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{h^m}K\left(\frac{x_i-x_j}{h}\right)\varepsilon_i\frac{\partial f(x_j,\theta_0)}{\partial\theta'}(\hat{\theta}-\theta_0)\right\} \\ &\quad+\left\{(\hat{\theta}-\theta_0)'\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{h^m}K\left(\frac{x_i-x_j}{h}\right)\varepsilon_i\frac{\partial^2 f(x_j,\bar{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}-\theta_0)\right\} \\ &\equiv S_{1n}(\hat{\theta}-\theta_0)+(\hat{\theta}-\theta_0)'S_{2n}(\hat{\theta}-\theta_0), \end{aligned} \tag{A.17}$$

where $\bar{\theta}$, depending on x_i, lies between θ_0 and $\hat{\theta}$.

Applying Lemma 3.3b, we immediately have

$$S_{1n} = \mathrm{O}_p(1/\sqrt{n}). \tag{A.18}$$

For S_{2n}, denote $\sigma(x_i) = \mathrm{E}(|\varepsilon_i| \,|\, x_i)$, we have

$$\begin{aligned}
\mathrm{E}(\|S_{2n}\|) &\leqslant \mathrm{E}\left[\frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \sigma(x_i) \left\|\frac{\partial f(x_i, \theta_0)}{\partial \theta'}\right\|\right] \\
&= \int \frac{1}{h_m} K\left(\frac{x_i - x_j}{h}\right) \sigma(x_i) \left\|\frac{\partial f(x_i, \theta_0)}{\partial \theta'}\right\| p(x_i)\, p(x_j)\, \mathrm{d}x_i \mathrm{d}x_j \\
&= \int K(u)\, \sigma(x_i) \left\|\frac{\partial f(x_i, \theta_0)}{\partial \theta'}\right\| p(x_i)\, p(x_i - hu)\, \mathrm{d}x_i \mathrm{d}u \\
&= \mathrm{O}(1).
\end{aligned} \tag{A.19}$$

Thus we have

$$S_{2n} = \mathrm{O}_p(1). \tag{A.20}$$

Since $\hat{\theta} - \theta_0 = \mathrm{O}_p(1/\sqrt{n})$, we have

$$\begin{aligned}
V_{2n} &= \mathrm{O}_p(1/\sqrt{n}) \cdot \mathrm{O}_p(1/\sqrt{n}) + \mathrm{O}_p(1/\sqrt{n}) \cdot \mathrm{O}_p(1) \cdot \mathrm{O}_p(1/\sqrt{n}) \\
&= \mathrm{O}_p(1/n).
\end{aligned} \tag{A.21}$$

Thus

$$nh^{m/2} V_{2n} = \mathrm{O}_p(h^{m/2}) \xrightarrow{p} 0.$$

V_{3n} can be written as

$$\begin{aligned}
V_{3n} &= (\hat{\theta} - \theta_0)' \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \\
&\quad \times \frac{\partial f(x_i, \bar{\theta}_1)}{\partial \theta} \frac{\partial f(x_j, \bar{\theta}_2)}{\partial \theta'} (\hat{\theta} - \theta_0) \\
&= (\hat{\theta} - \theta_0)' S_{3n} (\hat{\theta} - \theta_0),
\end{aligned} \tag{A.22}$$

where $\bar{\theta}_1$, depending on x_i, lies between θ_0 and $\hat{\theta}$, and $\bar{\theta}_2$, depending on x_i, also lies between θ_0 and $\hat{\theta}$. Similar to the proof of $S_{2n} = \mathrm{O}_p(1)$, we can easily show that

$$S_{3n} = \mathrm{O}_p(1). \tag{A.23}$$

Thus we have $V_{3n} = \mathrm{O}_p(1/\sqrt{n}) \cdot \mathrm{O}_p(1) \cdot \mathrm{O}_p(1/\sqrt{n}) = \mathrm{O}_p(1/n)$. So

$$nh^{m/2} V_{3n} = \mathrm{O}_p(h^{m/2}) \xrightarrow{p} 0. \quad \text{Q.E.D.}$$

From Lemmas 3.3a to 3.3d, we have proved the first part of Lemma 3.3. To show that $\hat{\Sigma}$ is a consistent estimator of Σ, we apply Lemma 3.1.

Lemma 3.3e. Given Assumptions 1–5, if $h \to 0$ and $nh^m \to \infty$, then under the null hypothesis (2.1),

$$\hat{\Sigma} \xrightarrow{p} \Sigma. \tag{A.24}$$

Proof of Lemma 3.3e

Using the same method in the proof of Lemmas 3.3a–3.3d, we can show that

$$\begin{aligned}\hat{\Sigma} &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) e_i^2 e_j^2 \\ &= 2\left\{\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) \varepsilon_i^2 \varepsilon_j^2\right\} + o_p(1) \\ &\equiv 2S_{4n} + o_p(1).\end{aligned} \tag{A.25}$$

S_{4n} is a standard U-statistic with

$$H_n(z_i, z_j) = \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) \varepsilon_i^2 \varepsilon_j^2. \tag{A.26}$$

As in the proof of Lemma 3.3b, we can easily show that $\mathrm{E}[\| H_n(z_i, z_j) \|^2] = o(n)$. Thus the condition of Lemma 3.1 is satisfied.

For the above $H_n(z_i, z_j)$, by (A.6) we have

$$\begin{aligned}\bar{r}_n &= \mathrm{E}\left[\frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) \varepsilon_i^2 \varepsilon_j^2\right] \\ &= \int K^2(u)\, \mathrm{d}u \int [\sigma^2(x)]^2 p^2(x)\, \mathrm{d}x + o(1) \\ &= \Sigma/2 + o(1),\end{aligned} \tag{A.27}$$

Therefore by Lemma 3.1,

$$S_{4n} = \bar{r}_n + o_p(1) = \Sigma/2 + o_p(1). \tag{A.28}$$

So we have

$$\hat{\Sigma} = 2S_{4n} + o_p(1) = \Sigma + o_p(1). \quad \text{Q.E.D.}$$

Summarizing our results from Lemmas 3.3a to 3.3e, we have proved Lemma 3.3. Q.E.D.

Proof of Lemma 3.4

The proof is similar to the proof of Lemma 3.3e.
First we can show that

$$\begin{aligned} V_n &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) e_i e_j \\ &= \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i \varepsilon_j \right\} + o_p(1) \\ &\equiv S_{5n} + o_p(1). \end{aligned} \tag{A.29}$$

S_{5n} is again a standard U-statistic with

$$H_n(z_i, z_j) = \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i \varepsilon_j. \tag{A.30}$$

Similar to the proof of Lemma 3.3b, the condition of Lemma 3.1 can be shown to be satisfied.

For the above $H_n(z_i, z_j)$, since $\mathrm{E}(\varepsilon_i \mid x_i) = g(x_i) - f(x_i, \theta_0)$ under the alternative, we have

$$\begin{aligned} \bar{r}_n &= \mathrm{E}\{\mathrm{E}[H_n(z_i, z_j) \mid x_i, x_j]\} \\ &= \frac{1}{h^m} \mathrm{E}\left\{ K\left(\frac{x_i - x_j}{h}\right) [g(x_i) - f(x_i, \theta_0)] [g(x_j) - f(x_j, \theta_0)] \right\} \\ &= \frac{1}{h^m} \int K\left(\frac{x_i - x_j}{h}\right) [g(x_i) - f(x_i, \theta_0)][g(x_j) - f(x_j, \theta_0)] p(x_i)\, p(x_j)\, \mathrm{d}x_i\, \mathrm{d}x_j \\ &= \frac{1}{h^m} \int K(u)\, [g(x_i) - f(x_i, \theta_0)]\, [g(x_i - hu) - f(x_i - hu, \theta_0)] \\ &\qquad \times p(x_i)\, p(x_i - hu)\, \mathrm{d}x_i\, h^m \mathrm{d}u \\ &= \int [g(x) - f(x, \theta_0)]^2 p^2(x)\, \mathrm{d}x + o(1) \\ &= \mathrm{E}\{[g(x_i) - f(x_i, \theta_0)]^2 p(x_i)\} + o(1). \end{aligned} \tag{A.31}$$

Therefore, by Lemma 3.1,

$$\begin{aligned} V_n &= S_{5n} + o_p(1) = \bar{r}_n + o_p(1) \\ &= \mathrm{E}\{[g(x_i) - f(x_i, \theta_0)]^2 p(x_i)\} + o_p(1). \end{aligned} \tag{A.32}$$

So we have shown the first part of Lemma 3.4. The second part can be shown in a similar way as Lemma 3.3e.

Under the alternative we can show that

$$\hat{\Sigma} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) e_i^2 e_j^2$$

$$= 2\left\{\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) \varepsilon_i^2 \varepsilon_j^2\right\} + o_p(1)$$

$$\equiv 2S_{6n} + o_p(1). \tag{A.33}$$

S_{6n} is a standard U-statistic with

$$H_n(z_i, z_j) = \frac{1}{h^m} K^2\left(\frac{x_i - x_j}{h}\right) \varepsilon_i^2 \varepsilon_j^2. \tag{A.34}$$

The condition of Lemma 3.1 can be easily verified as in Lemma 3.3b.

For the above $H_n(z_i, z_j)$, note that under the alternative $E(\varepsilon_i^2 \mid x_i) = \sigma^2(x_i) + [g(x_i) - f(x_i, \theta_0)]^2$, we have

$$\bar{r}_n = \frac{1}{h^m} \int K^2\left(\frac{x_i - x_j}{h}\right) \{\sigma^2(x_i) + [g(x_i) - f(x_i, \theta_0)]^2\}$$

$$\times \{\sigma^2(x_j) + [g(x_j) - f(x_j, \theta_0)]^2\} p(x_i)\, p(x_j)\, dx_i dx_j$$

$$= \frac{1}{h^m} \int K^2(u) \{\sigma^2(x_i) + [g(x_i) - f(x_i, \theta_0)]^2\}$$

$$\times \{\sigma^2(x_i - hu) + [g(x_i - hu) - f(x_i - hu, \theta_0)]^2\}$$

$$\times p(x_i)\, p(x_i - hu)\, dx_i h^m du$$

$$= \int K^2(u)\, du \int \{\sigma^2(x) + [g(x_i) - f(x_i, \theta_0)]^2\}^2 p^2(x)\, dx + o(1). \tag{A.35}$$

Therefore by Lemma 3.1,

$$\hat{\Sigma} = 2S_{6n} + o_p(1) = 2\bar{r}_n + o_p(1)$$

$$= 2 \int K^2(u)\, du \int \{\sigma^2(x) + [g(x_i - f(x_i, \theta_0)]^2\}^2 p^2(x)\, dx$$

$$+ o(1). \quad \text{Q.E.D.} \tag{A.36}$$

Proof of Theorem 3

Like the previous proofs, we can show

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) e_i e_j$$

$$= \left\{\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \varepsilon_i \varepsilon_j\right\} + o_p(1)$$

$$\equiv S_{7n} + o_p(1). \tag{A.37}$$

Let $u_i = \varepsilon_i - \delta_n \cdot l(x_i)$. Then $\mathrm{E}(u_i \mid x_i) = 0$. S_{7n} can be decomposed as

$$
\begin{aligned}
S_{7n} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) [u_i + \delta_n \cdot l(x_i)]\,[u_j + \delta_n \cdot l(x_j)] \\
&= \left\{\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) u_i u_j\right\} \\
&\quad + \delta_n \cdot \left\{\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) u_i\, l(x_j)\right\} \\
&\quad + \delta_n^2 \cdot \left\{\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) l(x_i)\, l(x_j)\right\} \\
&\equiv Q_{1n} + \delta_n \cdot Q_{2n} + \delta_n^2 \cdot Q_{3n}. \qquad \text{(A.38)}
\end{aligned}
$$

Similar to the proof of Lemma 3.3, we can show

$$nh^{m/2} Q_{1n} \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Sigma). \qquad \text{(A.39)}$$

Like the proof of Lemma 3.3b, we can show

$$\sqrt{n}\, Q_{2n} \xrightarrow{\mathrm{d}} \mathrm{N}(0, \mathrm{E}[\sigma^2(x_i)\, l^2(x_i)\, p^2(x_i)]). \qquad \text{(A.40)}$$

By mimic the proof of Lemma 3.4, we have

$$Q_{3n} \xrightarrow{\mathrm{p}} \mathrm{E}[l^2(x_i)\, p(x_i)] > 0. \qquad \text{(A.41)}$$

If $\delta_n = n^{-1/2} h^{-m/4}$, then

$$nh^{m/2} \delta_n Q_{2n} = h^{m/4} \cdot \sqrt{n}\, Q_{2n} \xrightarrow{\mathrm{p}} 0,$$

$$nh^{m/2} \delta_n^2\, Q_{3n} = Q_{3n} \xrightarrow{\mathrm{p}} \mathrm{E}[l^2(x_i)\, p(x_i)].$$

Thus

$$nh^{m/2} V_n \xrightarrow{\mathrm{p}} \mathrm{N}(\mathrm{E}[l^2(x_i)\, p(x_i)], \Sigma). \qquad \text{(A.42)}$$

Following the proof of Lemma 3.3e, we can easily show that $\hat{\Sigma} \to_p \Sigma$. Thus

$$T_n = \sqrt{\frac{n-1}{n}} \cdot \frac{\mathrm{n}h^{m/2} V_n}{\sqrt{\hat{\Sigma}}} \xrightarrow{\mathrm{p}} \mathrm{N}(\mu, 1),$$

where

$$\mu = \mathrm{E}[l^2(x_i)\, p(x_i)] / \sqrt{\Sigma}. \quad \text{Q.E.D.}$$

References

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

Davidson, R. and J.G. MacKinnon, 1981, Several model specification tests in the presence of alternative hypotheses, Econometrica 49, 781–793.

Eubank, R.L. and 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

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 University Press, Cambridge).

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

Hausman, J.A., 1978, Specification testing in econometrics, Econometrica 46, 1251–1271.

Hoeffding, W., 1948, A class of statistics with asymptotically normal distribution, Annals of Mathematical Statistics 19, 293–325.

Jennrich, R.I., 1969, Asymptotic properties of nonlinear least squares estimators, Annals of Mathematical Statistics 40, 633–643.

Lee, B.-J., 1988, A nonparametric model specification test using a kernel regression method, Ph.D. dissertation (University of Wisconsin, Madison, WI).

Newey, W.K., 1985a, Maximum likelihood specification testing and conditional moment tests, Econometrica 53, 1047–1070.

Newey, W.K., 1985b, Generalized method of moments specification testing, Journal of Econometrics 29, 229–256.

Pitman, E.J.G., 1949, Notes on non-parametric statistical inference (Columbia University, New York, NY).

Powell, J.L., J.H. Stock, and T.M. Stoker, 1989, Semiparametric estimation of index coefficients, Econometrics 57, 1403–1430.

Prakasa Rao, B.L.S., 1983, Nonparametric functional estimation (Academic Press, New York, NY).

Ruud, P.A., 1984, Tests of specification in econometrics, Econometric Reviews 3, 211–242.

Silverman, B.W., 1986, Density estimation for statistics and data analysis (Chapman and Hall, London).

Tauchen, G., 1985, Diagnostic testing and evaluation of maximum likelihood models, Journal of Econometrics 30, 415–443.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817–838.

White, H., 1981, Consequences and detection of misspecified nonlinear regression models, Journal of the American Statistical Association 76, 419–433.

White, H., 1982, Maximum likelihood estimation of misspecified models, Econometrics 50, 1–25.

White, H., 1987, Specification testing in dynamic models, in: T.F. Bewley, ed., Advances in econometrics, Fifth world congress, Vol. I (Cambridge University Press, Cambridge).

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., 1993, A residual-based consistent test of parametric regression models, Unpublished paper (University of Texas, Austin, TX).