

Reprinted from

# JOURNAL OF Econometrics

---

Journal of Econometrics 96 (2000) 1–23

A simple framework for nonparametric  
specification testing

Glenn Ellison\*, Sara Fisher Ellison

*MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02139, USA*





ELSEVIER

Journal of Econometrics 96 (2000) 1–23

JOURNAL OF  
Econometrics

www.elsevier.nl/locate/econbase

# A simple framework for nonparametric specification testing

Glenn Ellison\*, Sara Fisher Ellison

*MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02139, USA*

Received 1 November 1993; received in revised form 1 April 1999

---

## Abstract

This paper presents a simple framework for testing the specification of parametric conditional means. The test statistics are based on quadratic forms in the residuals of the null model. Under general assumptions the test statistics are asymptotically normal under the null. With an appropriate choice of the weight matrix, the tests are shown to be consistent and to have good local power. Specific implementations involving matrices of bin and kernel weights are discussed. Finite sample properties are explored in simulations. © 2000 Elsevier Science S.A. All rights reserved.

*JEL classification:* C14; C12

*Keywords:* Consistent testing; Specification testing; Nonparametric; Quadratic form

---

## 1. Introduction

Specification testing has become commonplace in econometrics, both as a means of testing economic theories which predict specific functional forms and as a regression diagnostic. Early specification tests,<sup>1</sup> although useful in many settings, are not consistent, i.e., there are alternatives which they will fail to detect regardless of the amount of data available. Partly in response to this

---

\* Corresponding author.

<sup>1</sup> See, for example, Ramsey (1969), Hausman (1978), Davidson and MacKinnon (1981), Newey (1985), Tauchen (1985), and White (1987).

concern a large recent literature has examined the behavior of specification tests which exploit nonparametric techniques.<sup>2</sup> The literature considers a variety of techniques including series estimation, spline estimation, and kernel estimation to test a null (parametric) model, with some of the tests having been shown to be consistent against all alternatives.

Our approach to testing a null model  $y_i = f(x_i; \alpha) + u_i$  employs test statistics based on quadratic forms in the model's residuals,  $\sum w_{ij} \tilde{u}_i \tilde{u}_j$ . One intuition is straightforward: quadratic forms can detect a spatial correlation in the residuals which would result from a functional form misspecification. We provide general conditions sufficient to ensure that the test statistics will be asymptotically normally distributed under the null and will be consistent and explore the finite sample performance of specific implementations of the test via Monte Carlo simulations.

Much of the previous literature on nonparametric specification testing has been motivated as testing the orthogonality between a model's residuals and an alternative nonparametric model.<sup>3</sup> Our testing framework can also be seen in this light. Consider a nonparametric estimator  $\hat{y} = Wy$ , e.g., kernel, spline, series, or other linear smoother. A Davidson–MacKinnon style test of orthogonality with  $\hat{y}$  as the misspecification indicator would be of the form:

$$\begin{aligned} T &= \frac{1}{c_N} \sum_i \hat{y}_i \tilde{u}_i \\ &= \frac{1}{c_N} \sum_i \left( \sum_j w_{ij} (x_j \tilde{\beta} + \tilde{u}_j) \right) \tilde{u}_i \\ &= \frac{1}{c_N} \sum_{ij} w_{ij} \tilde{u}_i \tilde{u}_j + \frac{1}{c_N} \sum_i \tilde{u}_i \left( \sum_j w_{ij} x_j \right) \tilde{\beta}. \end{aligned}$$

The first term is a quadratic form in the residuals. The second measures the orthogonality between the residuals and something that is of the form of an estimate of  $X$  and should be small. Hence, we can think of a quadratic form test with a weight matrix  $W$  as similar to an orthogonality test with  $Wy$  as the misspecification indicator.<sup>4</sup>

<sup>2</sup>Work in this vein includes Azzalini et al. (1989), Bierens (1982, 1984, 1990), Bierens and Ploberger (1997), de Jong and Bierens (1994), Delgado and Stengos (1994), Eubank and Spiegelman (1990), Fan and Li (1996), Gozalo (1993), Härdle and Mammen (1993), Hidalgo (1992), Hong and White (1995), Horowitz and Härdle (1994), Rodriguez and Stoker (1993), White and Hong (1993), Wooldridge (1992), Yatchew (1992), and Zheng (1996).

<sup>3</sup>This, for example, is the approach of Hong and White (1995), Eubank and Spiegelman (1990) and Wooldridge (1992).

<sup>4</sup>A previous version of this paper showed the equivalence between a quadratic form test and an orthogonality test in some (but only some) situations.

We hope  
construction  
a variety of  
detect various  
power and i  
framework is  
and thus allo

The remai  
class of qua  
establishes t  
general envi  
cluding one  
which was i  
corrections  
establishing  
tation the lo  
contemporar  
Hong and V  
Section 4 pre  
the power of  
samples. In a  
finite sample  
accuracy of  
power.

## 2. The test s

We first  
regression m  
the null. We  
utility of the

### 2.1. General

We will be  
model of the  
steps. The r

<sup>5</sup>The one te  
Ploberger (1997  
than  $N^{-1/2}$ , bu

We hope that our approach may be seen as useful for a few reasons. The construction is general and transparent, allowing researchers to base tests on a variety of nonparametric estimation techniques and to easily tailor tests to detect various types of misspecification, if desired. The tests have good local power and in simulations appear to have reliable size in small samples. The framework is also well-suited to the application of standard binning techniques and thus allows for tests which are undemanding computationally.

The remainder of the paper is structured as follows. Section 2 introduces the class of quadratic form test statistics with which we shall be concerned and establishes their asymptotic normality with correct specification in a fairly general environment. Several specific implementations are then discussed, including one based on a kernel regression estimator that is similar to the test which was independently proposed by Zheng (1996). Potential finite sample corrections are also discussed. Section 3 contains a fairly general theorem establishing the consistency of the tests. With an appropriate kernel implementation the local power of the tests is equal to that of the best of the prior and contemporaneously proposed tests, e.g. Eubank and Spiegelman (1990) and Hong and White (1995), and is superior to that of many other approaches.<sup>5</sup> Section 4 presents the results of a set of Monte Carlo simulations which examine the power of our tests and the reliability of the asymptotic critical values in finite samples. In a comparison with several other tests, our tests (with the suggested finite sample corrections) appear to have an advantage in the finite-sample accuracy of asymptotic critical values, and perform fairly well also in terms of power.

## 2. The test statistic: Definition and asymptotic distribution

We first introduce a general class of specification tests for a nonlinear regression model and prove that the statistics are asymptotically normal under the null. We then discuss several special cases in more detail to illustrate the utility of the framework.

### 2.1. General definition and asymptotic normality

We will be concerned with testing the specification of a nonlinear regression model of the form  $E(y_i | X) = f(x_i; \alpha)$ . We explore procedures consisting of two steps. The null model is first estimated by some  $\sqrt{N}$ -consistent procedure

---

<sup>5</sup>The one test we are aware of which has slightly better local power is that of Bierens and Ploberger (1997) which is consistent not only against local alternatives which shrink at a rate slower than  $N^{-1/2}$ , but also against local alternatives which shrink at a rate of exactly  $N^{-1/2}$ .

producing residuals  $\tilde{u}$ . Test statistics based on a quadratic form  $\tilde{u}'W\tilde{u}$  are then formed. Large positive values of the test statistic indicate misspecification.

We begin with a very general proposition which defines a class of test statistics  $\mathcal{F}_N$  and gives their asymptotic distribution. Given conditions on the eigenvalues of the weight matrix, the quadratic form test statistics are asymptotically normal. We would like to emphasize that Proposition 1 is applicable not only to consistent tests, but also to quadratic form tests tailored to detect particular forms of misspecification.

To state the proposition, we need a few definitions. The matrix  $A$  is said to be nonnegative if each of its elements is nonnegative. For an  $N \times N$  matrix  $A$ , we write  $r(A)$  for the spectral radius of  $A$  defined by

$$r(A) \equiv \sup_{v \in \mathbb{R}^N, v \neq 0} \frac{\|Av\|}{\|v\|}.$$

When  $A$  is symmetric with eigenvalues  $|\gamma_1| \geq |\gamma_2| \geq \dots \geq |\gamma_N|$ , it is well known that  $r(A) = |\gamma_1|$ . Define

$$s(A) \equiv \left( \sum_{i,j} a_{ij}^2 \right)^{1/2}.$$

When  $A$  is symmetric, it is easy to see that  $s(A) = (\sum \gamma_i^2)^{1/2}$ .

*Proposition 1.* Suppose  $y_i = f(x_i; \alpha_0) + u_i$ , where  $\{x_i\}$  is a sequence of i.i.d. random variables having compact support  $D \subset \mathbb{R}^d$ , and  $\{u_i\}$  is a sequence of independent random variables with  $u_i$  independent of  $x_j$  for  $j \neq i$ ,  $E(u_i | x_i) = 0$ ,  $0 < \sigma^2 \leq \text{Var}(u_i | x_i) \leq \bar{\sigma}^2 < \infty$ , and  $E(u_i^4 | x_i) \leq m < \infty$  for all  $i$  and  $x_i$ . Assume also that  $f: D \times \mathbb{R}^r \rightarrow \mathbb{R}$  is twice continuously differentiable. Let  $\tilde{\alpha}_N$  be a  $\sqrt{N}$ -consistent estimate for  $\alpha_0$ . Define  $\tilde{u}_{iN} \equiv y_i - f(x_i; \tilde{\alpha}_N)$ . Write  $\tilde{u}^N$  for the  $N$ -vector  $(\tilde{u}_{1N}, \dots, \tilde{u}_{NN})$ , and  $\tilde{U}^N$  for the  $N \times N$  diagonal matrix with  $i$ th element  $\tilde{u}_{iN}$ . Let  $W_N: D^N \rightarrow \mathbb{R}^N$  be a function associating a symmetric  $N \times N$  matrix to each realization of  $(x_1 \dots x_N)$ . Suppose that  $w_{ii} = 0$  for  $i = 1, 2, \dots, N$ ,  $r(W_N)/s(W_N) \xrightarrow{P} 0$  as  $N \rightarrow \infty$ , and that  $FSC_N \xrightarrow{P} 0$  as  $N \rightarrow \infty$ . Let

$$\mathcal{F}_N = \frac{\tilde{u}^N W_N \tilde{u}^N}{\sqrt{2s(\tilde{U}^N W_N \tilde{U}^N)}} + FSC_N.$$

Then,  $\mathcal{F}_N \xrightarrow{\mathcal{L}} \mathcal{L} \sim \mathcal{N}(0,1)$ .<sup>6</sup>

<sup>6</sup> We take this opportunity to explain a few of the notational liberties we will take. First, we will normally write  $W_N$  for the matrix  $W_N(x_1, \dots, x_N)$ . Note that  $W_N$  is stochastic when  $\{x_i\}$  is stochastic. Second, we will often drop the  $N$  subscript or superscript when no confusion will arise. For example, the elements of the matrix  $W_N$  will be written  $w_{ij}$ , and the vector  $u^N$  will often be simply  $u$ . Finally, to refer to the matrix of explanatory variables in our models, we will sometimes use the notation  $x^N$  and sometimes  $X$ , depending on the context.

Rem  
r(W  
aut  
a m  
can  
com  
idea  
sim  
with  
spec  
2.  
stra  
a cc  
stan  
vari  
estim  
Mo

W  
a de  
use  
For  
or i  
for

whe  
in a  
a re  
in S

T  
H  
nor

7  
a reg

*Remark 1.* Proposition 1 shows that  $\mathcal{T}_N$  converges to a standard normal when  $r(W_N)/s(W_N) \xrightarrow{P} 0$ . We will see in Sections 2.3 and 2.4 that this condition is automatically satisfied given appropriate regularity conditions when  $W$  is a matrix of bin or kernel weights. For any other sequence of weight matrices, one can always try to apply Proposition 1 directly. In any case, it may be useful to compute the ratio  $r(W_N)/s(W_N)$  and see how close it is to zero to get a rough idea of how far the statistics may be from normal in the given finite sample. In simulations described in Ellison and Ellison (1998), we find the true size of a test with 5% asymptotic critical values to be between 4% and 6% in each of the specifications we tried for which  $r(W_N)/s(W_N)$  was less than 0.4.

2. The term  $FSC_N$  in the test statistic is a finite-sample correction. The most straightforward application of Proposition 1 would be to the development of a consistent test for misspecification of a linear regression model (with a constant). For such an application with  $X$  the matrix of nonconstant explanatory variables and with  $W_N$  being a weight matrix such that  $W_N y$  is a consistent estimator of  $f$  (such as a matrix of kernel weights) we recommend and use in our Monte Carlo study the correction

$$FSC_N = \frac{1 + \text{rank}(X)}{\sqrt{2 s(W_N)}}.$$

While the use of nonparametric specification tests is often motivated by a desire for consistency, at other times an applied econometrician might want to use a nonparametric test designed to detect specific forms of misspecification. For example, one might be particularly interested in nonlinearity in one variable or in the presence of an omitted variable.<sup>7</sup> Our more general recommendation for such cases would be to use the correction

$$FSC_N = \frac{\sum_{k=0}^d \hat{\beta}_k}{\sqrt{2 s(W_N)}},$$

where  $\hat{\beta}_k$  is the coefficient on  $X_{\cdot k}$  (the  $k$ th explanatory variable in the null model) in a regression of  $W_N X_{\cdot k}$  on  $X$  (and a constant) and  $\hat{\beta}_0$  is the constant term from a regression of  $W_N 1_N$  on  $X$ . We provide motivation for our suggested correction in Section 2.2.

The proof of Proposition 1, as well as all other proofs, is in the Appendix.

Intuitively, the reason why a quadratic form in the residuals is asymptotically normal is that any symmetric matrix  $A$  can be written as  $A = \Phi' \Lambda \Phi$  with

<sup>7</sup> Chevalier and Ellison (1997), for example, use a quadratic form test to investigate whether a regression is nonlinear in one of several independent variables.

$\Lambda$  diagonal and  $\Phi$  the orthogonal matrix of eigenvectors of  $A$ . If  $u$  is a vector of independent random variables we then have

$$u' Au = u' \Phi' \Lambda \Phi u = v' A v = \sum_i \lambda_i v_i^2,$$

where  $v = \Phi u$  is a vector of uncorrelated random variables. We thus have that  $u' W u$  is a weighted sum of the squares of a set of uncorrelated random variables, and this is asymptotically normal provided the square of the largest weight (which is equal to  $r(A)^2$ ) becomes arbitrarily small compared to the sum of the squares of the weights (which is equal to  $s(A)^2$ ).

### 2.2. Motivation for a finite-sample correction

The form of our suggested finite-sample correction,  $FSC_N$ , is motivated by an analysis of the finite sample mean of the numerator of the test statistic in the simplest case – the parametric null being a linear regression with homoskedastic errors estimated by OLS. In this case we have

$$\begin{aligned} E(\tilde{u}' W \tilde{u}) &= E(u'(I - P_X)W(I - P_X)u) \\ &= E(u' W u) - E(u' P_X W u) - E(u' W P_X u) + E(u' P_X W P_X u) \\ &= -\sigma^2 \text{Tr}(P_X W) = -\sigma^2 \text{Tr}((X' X)^{-1} X' W X). \end{aligned}$$

(The last line follows from repeatedly applying the identities  $E(u' A u) = \sigma^2 \text{Tr}(A)$  and  $\text{Tr}(AB) = \text{Tr}(BA)$ .)

Writing  $\hat{X}$  for  $W X$ , the  $k$ th diagonal element of the matrix  $(X' X)^{-1} X' W X$  is simply the coefficient on  $X_{\cdot k}$  in a regression of  $\hat{X}_{\cdot k}$  on  $X$ . This motivates our more general suggested finite sample correction. When  $W$  is the weight matrix corresponding to a consistent estimator,  $\hat{X}_{\cdot k}$  will approach  $X_{\cdot k}$  as  $N \rightarrow \infty$  (under appropriate conditions), and hence each of these regression coefficients should approach one. The simpler finite sample correction we recommend for such  $W$  is motivated by there being  $1 + \text{rank}(X)$  such regression coefficients when  $X$  is augmented by a column of ones.

### 2.3. A special case: The kernel test

The proposition of the previous section identifies the asymptotic distribution of a broad class of test statistics. Recall that one motivation for looking at a statistic of the form  $\tilde{u}' W \tilde{u}$  is that it is similar to a test of orthogonality between  $\tilde{u}$  and the nonparametric estimate  $\hat{y} = W y$ . A typical application of our framework suggested by this motivation is to take the matrix  $W$  to be the weight matrix from a kernel regression of  $y$  on  $X$ .

An easy application of Proposition 1 shows that a test statistic formed from kernel weights is asymptotically normal given very minimal restrictions on the kernel and the rate at which the window width shrinks to zero.

*Corollary 1.* Let  $y, f, \{x_i\}, \{u_i\}, \tilde{\alpha}, \tilde{u}, \tilde{U}, D$  be as in Proposition 1. Suppose also that the distribution of  $x_i$  has a twice continuously differentiable density  $p(x) \geq p > 0$  on  $D$ .

Let the kernel  $K(x)$  be a nonnegative function satisfying  $\int_{\mathbb{R}^d} K(x) dx = 1$  and  $\int_{\mathbb{R}^d} K(x)^2 dx < \infty$ , and define  $W_N$  by

$$w_{ijN} = \begin{cases} \frac{K(1/h_N \cdot (x_i - x_j))}{\sum_{k \neq i} K(1/h_N \cdot (x_i - x_k))} & \text{if } j \neq i \text{ and } \sum_{k \neq i} K(1/h_N \cdot (x_i - x_k)) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let  $W_N^s = (W_N + W_N')/2$ . Suppose that for some  $\delta > 0$ ,  $N^\delta h_N \rightarrow 0$  and  $Nh_N^d h_N^{-\delta} \rightarrow \infty$ . Then

$$\mathcal{T}_N = \frac{\tilde{u}' W_N \tilde{u}}{\sqrt{2s(\tilde{U}^N W_N^s \tilde{U}^N)}} + \frac{1+d}{\sqrt{2s(W_N^s)}} \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1).$$

*Remark 1.* The test described in Corollary 1 is quite similar to the test which was independently proposed in Zheng (1996). The principal difference is that the density weighting of the various terms in the quadratic form differs because we have made each row of the weight matrix sum to one rather than using raw kernel weights.

2. Some assumptions have been made purely for convenience. For example, the assumption that the kernel is nonnegative is used only because it makes it easy to conclude that  $\sum_{ij} w_{ij} w_{ji} \geq 0$  and that  $r(W_N)$  remains bounded as  $N \rightarrow \infty$ . If one wanted to use a kernel which was sometimes negative, a variety of other assumptions could be added to obtain these conclusions.<sup>8</sup>

3. The calculations in the proof illustrate our earlier comment that non-parametric test statistics may converge to their asymptotic distributions very slowly. In the proof we note that  $s(W_N) = O_p(h^{-d/2})$ . This implies that our recommended finite-sample correction only tends to zero like  $h^{d/2}$ . If, for example, one is testing a model with one explanatory variable and chooses  $h_N = h_0 N^{-1/5}$ , our finite sample correction term will only vanish at the rate of

<sup>8</sup> Without the assumption that the kernel is nonnegative, an additional assumption, e.g.  $\int_{\mathbb{R}^d} |K(x)| dx < \infty$ , would also be needed to ensure that the kernel density estimates used in the proof are consistent.



$N^{-1/10}$ .<sup>9</sup> Finite sample corrections may thus be important, even when hundreds of thousands of observations are available.

#### 2.4. 'Binning' and other tests

The testing framework can accommodate a wide variety of weight matrices. For example,  $W$  could be the weight matrix corresponding to any smoother of the form  $W_N y$ , such as  $k$ -nearest neighbor estimators, splines, orthogonal series estimators, and convolution smoothing.<sup>10</sup> If one suspects nonlinearity in one of several  $X$  variables, one could choose weights which depend only on differences in that one variable. Such a test would not be consistent of course – misspecifications in other  $X$  variables could go undetected – but it might have better power in detecting that particular form of misspecification.

One simple alternate implementation is a 'bin' version of the test.<sup>11</sup> It may be obtained by dividing the data into  $m(N)$  bins and setting all nondiagonal weights equal to each other inside the bins and equal to zero outside the bins. In contrast to a kernel test statistic, which requires  $O(N^2 h)$  computations, a bin test statistic requires  $O(N)$  computations. An  $O(N)$  computation of the test statistic and general conditions sufficient to ensure its asymptotic normality are described by the following corollary.

*Corollary 2. Let  $y, f, \{x_i\}, \{u_i\}, \tilde{\alpha}, \tilde{u}, D$  be as in Proposition 1. Suppose also that each  $x_i$  is drawn from a distribution with measure  $\nu$  on  $D$ .*

*Consider a sequence of partitions  $\{P_{kN}\}$  with  $D = P_{1N} \cup P_{2N} \cup \dots \cup P_{m(N)N}$ ,  $P_{kN} \cap P_{jN} = \emptyset$ ,  $k \neq j$ ,  $m(N) \rightarrow \infty$ , and  $N \inf_k \nu(P_{kN}) \rightarrow \infty$ . Write  $C_{kN}$  for the random variable giving the number of elements of  $\{x_1, \dots, x_N\}$  which lie in  $P_{kN}$ ,  $S_N$  for  $(\sum_{k:s.t. C_{kN} \geq 2} C_{kN} / (C_{kN} - 1))^{1/2}$ , and  $V_{kN}(n)$  for  $\sum_{i:s.t. x_i \in P_{kN}} \tilde{u}_{iN}^n$ . Define*

$$\mathcal{T}_N = \frac{\sum_{k:s.t. C_{kN} \geq 2} \frac{V_{kN}(1)^2 - V_{kN}(2)}{C_{kN} - 1}}{\left(2 \sum_{k:s.t. C_{kN} \geq 2} \frac{V_{kN}(2)^2 - V_{kN}(4)}{(C_{kN} - 1)^2}\right)^{1/2}} + \frac{1 + d}{\sqrt{2S_N}}$$

*Then  $\mathcal{T}_N \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1)$ .*<sup>12</sup>

<sup>9</sup> Local power calculations presented later will suggest using window widths which shrink even more slowly.

<sup>10</sup> See Härdle (1990) for a discussion of the weight matrices corresponding to the estimators mentioned above and others.

<sup>11</sup> See Tukey (1961) and Loftsgaarden and Quesenberry (1965) for early discussions of computationally simpler smoothing estimators, such as the regressogram. These estimators would result in a weight matrix similar to that which we suggest for the bin test.

<sup>12</sup> If one wishes to use bins which are based on finer and finer divisions of only one, or more generally  $z$ , of the  $X$  variables, the  $(1 + d)$  term in the finite sample correction could be replaced by  $(1 + z)$ .

### 3. Consistency and local power

An important motivation for nonparametric specification testing is that parametric tests will fail to detect departures from the null in certain directions. In this section we verify that, given fairly general conditions on the choice of a weight matrix, the tests described in the previous chapter are indeed consistent.

As is standard, we consider whether the test can detect alternatives  $g_N(x)$  which approach  $f(x; \alpha_0)$  as  $N \rightarrow \infty$ , e.g.,  $g_N(x) = f(x; \alpha_0) + N^{-\xi}e(x)$  with  $\xi \geq 0$  and  $e(x)$  orthogonal to the space of null functions  $f$ . The proposition shows that if the alternatives do not approach the null too quickly, i.e., if  $\xi < \bar{\xi}$ , then the test will detect the alternative with probability one. Only two fairly weak conditions on the weight matrix are required: that the eigenvalues satisfy  $r(W_N) = 1$  and  $s(W_N) \xrightarrow{p} \infty$ , and that the nonparametric estimator  $\hat{e}_N(x) \equiv W_N e(x)$  have a mean squared error smaller than the function being estimated. These conditions are satisfied for bin and kernel weight matrices, among others, if  $e(x)$  is piecewise continuous.

*Proposition 2. Suppose  $y_i = g_N(x_i) + u_i$  with  $\{x_i\}$ ,  $\{u_i\}$ ,  $\tilde{u}$ ,  $\tilde{U}$ ,  $D$  as in Corollary 1, and  $g_N: D \rightarrow \mathcal{R}$  a sequence of functions. Write  $X$  for the matrix  $(x_1 \dots x_N)$ . Let  $\tilde{\alpha}(y, X)$  be an estimator for which there exists a sequence  $\alpha_N^*$  such that  $\sqrt{N}(\tilde{\alpha}(y, X) - \alpha_N^*) \xrightarrow{L} \mathcal{Z} \sim \mathcal{N}(0, \Omega)$  and  $\alpha_N^* \rightarrow \alpha_0$  as  $N \rightarrow \infty$  for some  $\alpha_0$ . Let  $W_N(X)$  be a sequence of matrices with  $w_{ii} = 0 \forall i$ ,  $r(W_N) \xrightarrow{p} 1$ , and  $1/s(W_N^s) \xrightarrow{p} 0$  as  $N \rightarrow \infty$ . Suppose  $FSC_N \xrightarrow{p} 0$  as  $N \rightarrow \infty$ . Let  $\tilde{f}$  be the vector whose  $i$ th element is  $f(x_i; \tilde{\alpha})$  and similarly for other functions of  $x_i$ .*

*Let  $\bar{\xi}$  be defined by  $\bar{\xi} \equiv \sup\{\xi | N^{2\xi-1} s(\Sigma^N W_N^s \Sigma^N) \xrightarrow{p} 0\}$ , and suppose there exists a constant  $\xi$ ,  $0 \leq \xi < \bar{\xi}$  and a bounded function  $e(x)$  with  $\int_D e(x)^2 p(x) dx \neq 0$  such that  $N^\xi(g_N(x) - f(x; \alpha_N^*)) \rightarrow e(x)$  uniformly in  $x$ . Suppose also that  $W_N$  is such that*

$$\Pr\{\|W_N e - e\| \leq (1 - \underline{\delta})\|e\|\} \rightarrow 1 \text{ as } N \rightarrow \infty \text{ for some } \underline{\delta} > 0.$$

*Let*

$$\mathcal{T}_N = \frac{\tilde{u}' W_N \tilde{u}}{\sqrt{2s(\tilde{U}^N W_N^s \tilde{U}^N)}} + FSC_N.$$

*Then,  $\mathcal{T}_N \xrightarrow{p} \infty$  as  $N \rightarrow \infty$ .*

It is instructive here to comment on the rate at which the local alternatives may approach the null. Recall that in the case of the kernel test we saw (in the proof of Corollary 1) that  $s(W_N^s) = O_p(h_N^{-d/2})$ . Hence, the definition of  $\bar{\xi}$  gives  $\bar{\xi} = \frac{1}{2} + \lim_{N \rightarrow \infty} (d/2) \log_N h_N$ . For example, if  $d = 1$  and the kernel is chosen to

be the standard ‘optimal’<sup>13</sup> kernel with  $h_N = O(N^{-1/5})$  then the test is consistent against alternatives of order  $N^{-\xi}$  for  $\xi < \frac{2}{5}$ . With a more slowly shrinking window width, the test can be made to have power against local alternatives of order  $N^{-\xi}$  for any  $\xi < \frac{1}{2}$ . Similarly, the bin test has  $s(W_N^s) = O_p(m(N))$  so  $\bar{\xi} = \frac{1}{2} + \lim_{N \rightarrow \infty} -\frac{1}{2} \log_N m(N)$ . We again get  $\bar{\xi}$  close to  $\frac{1}{2}$  if we let the number of bins grow slowly.

This local power is equal to that of the best of the prior and contemporaneously proposed nonparametric tests, such as Eubank and Spiegelman (1990) and Hong and White (1995), and is superior to that of many other approaches. The only test of which we are aware that obtains slightly superior local power is the test of Bierens and Ploberger (1997) which is consistent against local alternatives which shrink at a rate of exactly  $N^{-1/2}$  as well.

#### 4. Simulation results

Here we present a Monte Carlo study of the finite sample power of our tests and the reliability of asymptotic critical values. A more comprehensive set of Monte Carlo results can be found in Ellison and Ellison (1998). For easy comparison (and to convince the reader that we had not designed the simulations to highlight our tests’ attributes), we have chosen to piggyback on the work done by Hong and White (1995) by simply adding statistics on the performance of our tests to tables containing the results of their Monte Carlo study.

Table 1 speaks to the ability of applied researchers to rely on the asymptotic critical values of the various tests. For this table, we used 10,000 simulations to estimate the size of two kernel implementations of our tests when they are performed using 5% asymptotic critical values on the null specification used in Hong and White (1995). The specification involves a linear model with two explanatory variables. The test labelled Ellison-Ellison1 is based on a kernel weight matrix with  $h_{100} = 1.0$ , and that labelled Ellison-Ellison2 is a kernel test with  $h_{100} = 1.5$ . The sizes of other tests are merely reprinted from Hong and White (1995). We did not repeat their simulations. The tests labelled Bierensi, ES&Ji, Hong–Whitei, Wooldridgei, and Yatchewi are versions of the tests of Bierens (1990), Eubank and Spiegelman (1990) and Jayasuriya (1996), Hong and White (1995), Wooldridge (1992), and Yatchew (1992), respectively, with the particular smoothing parameters, series expansions, etc. described in Hong and White (1995).

The true sizes of our tests in the six sample size-window width combinations are 3.8%, 4.1%, 4.4%, 4.8%, 4.9%, and 4.9%, figures which are much closer to 5% than are those for any of the other tests. Several of the other tests often have

<sup>13</sup> Here we mean optimal kernel in the context of estimation, not testing.

Table 1  
Comparison of finite-sample ACV size

Test statistic	Rejection rates with 5% ACV under null		
	$N = 100$	$N = 300$	$N = 500$
Ellison-Ellison1	4.9	4.9	4.8
Ellison-Ellison2	3.8	4.1	4.4
Bierens1	4.8	3.6	6.1
Bierens2	6.7	7.5	10.7
ES & J1	2.0	1.3	2.2
ES & J2	2.7	3.2	3.0
Hong-White1	1.6	2.0	2.0
Hong-White2	2.8	2.8	2.7
Wooldridge1	8.7	5.3	7.5
Wooldridge2	10.0	9.7	10.5
Yatchew1	7.2	6.2	8.5
Yatchew2	10.4	9.9	13.1

Source: Figures for Ellison-Ellison tests computed from 10,000 simulations. Figures for other tests taken from Table 2 of Hong and White (1995).

an ACV size substantially in excess of 5%. We noted earlier that nonparametric test statistics tend to converge to their asymptotic distributions relatively slowly. Here, many of the test statistics do no better with 500 observations than they do with 100 observations.<sup>14</sup> Because ACV sizes improve so slowly, finite sample performance is of great importance.

Duncan and Jones (1994) have in the course of their empirical work on labor supply also performed a Monte Carlo study which compares our test with those of Gozalo (1993) and Delgado and Stengos (1994). Their results also indicate that the asymptotic critical values of our test are more reliable than those of the other two tests in finite samples.

Table 2 compares the finite sample power of our test with that of the other nonparametric tests mentioned above. The table reports the rejection rates of our and other tests (using empirically estimated critical values) against the three other alternatives mentioned in Hong and White (1995).<sup>15</sup> For each alternative we report rejection rates from simulations involving 100 and 300 observations.<sup>16</sup>

<sup>14</sup> We would guess that they would probably show little improvement even with ten thousand observations.

<sup>15</sup> We have chosen to use the set of simulations with  $\sigma^2 = 1$  for alternatives 1 and 3, and those with  $\sigma^2 = 4$  for alternative 2 so as to make the power of the tests against the three alternatives more comparable. Note that in the case of alternative 2 the exponent within the exponential term is 2 not  $-2$ , which we believe is the proper correction to a misprint in the text of Hong and White (1995).

<sup>16</sup> The rejection rates for the tests other than ours are drawn from Hong and White's Tables 3–5.

Table 2  
Comparison of power

Test statistic	Rejection rates with 5% empirical critical values					
	Alternative 1		Alternative 2		Alternative 3	
	<i>N</i> = 100	<i>N</i> = 300	<i>N</i> = 100	<i>N</i> = 300	<i>N</i> = 100	<i>N</i> = 300
Ellison-Ellison1	31.2	81.2	21.8	59.0	5.5	5.1
Ellison-Ellison2	37.2	89.4	34.3	75.7	4.3	6.0
Bierens1	12.1	31.6	38.6	83.7	4.8	4.6
Bierens2	10.5	31.3	43.0	88.9	5.1	4.7
ES & J1	43.8	88.9	22.8	61.3	5.3	4.6
ES & J2	36.2	79.1	20.0	54.9	5.7	5.3
Hong-White1	46.5	91.6	28.0	71.1	5.4	5.0
Hong-White2	36.8	81.5	21.4	59.5	5.4	5.2
Wooldridge1	6.2	2.8	19.5	29.3	19.8	30.8
Wooldridge2	4.7	4.1	15.4	38.4	22.9	43.6
Yatchew1	8.6	9.3	7.5	8.2	6.5	5.0
Yatchew2	8.4	10.4	7.7	7.8	6.2	5.2

Source: Figures for Ellison-Ellison tests computed from 1000 simulations. Figures for other tests taken from Tables 3–5 of Hong and White (1995).

Looking at each of the alternatives in turn, it appears that the Hong-White test, Eubank-Spiegelman-Jayasuriya test, and our test do much better than the others against alternative 1. Bierens' test appears to be most powerful against alternative 2, followed by our test, Hong and White's, and Eubank and Spiegelman and Jayasuriya's. Only Wooldridge's test does at all well against alternative 3. It should be noted, of course, that assessing power from performance against so few alternatives may be misleading. This is particularly true here because the power of the tests based on series expansions is greatly affected by the degree to which the alternatives and the included series terms are collinear and because all of the alternatives in Hong and White's study are similar in that they involve low frequency misspecifications.

## 5. Conclusion

In this paper, we have presented a framework for specification testing which involves working directly with quadratic forms in a model's residuals. The framework allows one to construct asymptotically normal test statistics exploiting a variety of nonparametric techniques, and we have seen that these tests can be consistent and have good local power.

We hope that several factors may make our tests attractive to applied researchers. First, the tests are very intuitive, which we feel is important not only because one is always more comfortable with a test which one understands well, but because this understanding makes it easy to adapt the test to the particular situation one is facing. Second, because the null distributions of nonparametric tests tend to converge slowly to their asymptotic limits, and computational concerns make simulating null distributions undesirable, it is particularly important that the asymptotic approximation to the null distribution of a nonparametric test be accurate in small samples. Using the finite sample correction we suggest, our test does substantially better on this count than other tests in our Monte Carlo simulation. Finally, it is easy to construct computationally undemanding versions of the test.

As for future extensions, we see the greatest loose end in the current formulation as being the need for a choice of a smoothing parameter. While simulations can provide some guidance, it would be interesting to explore criteria for choosing the smoothing parameter automatically. A preliminary idea is to base a test on choosing the smoothing parameter to maximize a quadratic form test statistic. Given the success of Bierens and Ploberger (1997), we hope that such a construction might both eliminate an arbitrary choice and improve local power.

### Acknowledgements

We thank the following persons for assistance and helpful comments: Joseph Beaulieu, Jerry Hausman, Guido Imbens, Roger Klein, Whitney Newey, Danny Quah, Tom Stoker, Jeff Wooldridge, and anonymous referees. We also thank Emek Basker and Sean May for excellent research assistance. Financial support from National Science Foundation grant SBR-95-15076 and a Sloan Research Fellowship is acknowledged. This paper is a revised version of Sara Fisher Ellison's 1991 paper, "A Nonparametric Residual-based Specification Test: Asymptotic, Finite-sample, and Computational Properties".

### Appendix A

*Proof of Proposition 1.* To begin, we note that an elementary probabilistic argument shows that it suffices to prove that the result holds whenever  $\{x_i\}$  is nonstochastic and  $r(W_N)/s(W_N) \rightarrow 0$  as  $N \rightarrow \infty$ . To see this, apply Lemma A.1 below with  $a_N(x^N) = r(W_N)/s(W_N)$  and  $t_N(x^N, u^N) = \mathcal{T}_N$ .

To show that  $\mathcal{T}_N \xrightarrow{d} \mathcal{Z} \sim \mathcal{N}(0,1)$  when  $\{x_i\}$  is nonstochastic and  $r(W_N)/s(W_N) \rightarrow 0$ , note first that Theorem 1.1 of Mikosch (1991) (using

$A_N = \Sigma^N W_N \Sigma^N$  implies that

$$\frac{u^N W_N u^N}{\sqrt{2s(\Sigma^N W_N \Sigma^N)}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1),$$

where  $\Sigma^N$  is an  $N \times N$  diagonal matrix with  $\Sigma_{ii}^N = \text{Var}(u_i)^{1/2}$ . Two additional lemmas then let us conclude that  $\mathcal{F}_N$  also converges in distribution to a standard normal. The conclusion of Lemma A.2 is that

$$\frac{\tilde{u}^N W_N \tilde{u}^N - u^N W_N u^N}{\sqrt{2s(\Sigma^N W_N \Sigma^N)}} \xrightarrow{P} 0,$$

which implies that

$$\frac{\tilde{u}^N W_N \tilde{u}^N}{\sqrt{2s(\Sigma^N W_N \Sigma^N)}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1).$$

Lemma A.3 establishes that

$$\frac{\sum_{ij} w_{ij}^2 \tilde{u}_{iN}^2 \tilde{u}_{jN}^2}{s(\Sigma^N W_N \Sigma^N)^2} \xrightarrow{P} 1,$$

which implies that

$$\frac{\tilde{u}^N W_N \tilde{u}^N}{\sqrt{2s(\tilde{U}^N W_N \tilde{U}^N)}} \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1).$$

We have assumed that  $FSC_N \xrightarrow{P} 0$  so this implies  $\mathcal{F}_N \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1)$ .  $\square$

*Lemma A.1.* Let  $\{u_i\}$  and  $\{x_i\}$  be as above. Write  $u^N$  for  $(u_1, \dots, u_N)$ ,  $x^N$  for  $(x_1 \dots x_N)$  and  $\underline{x}^N$  for a realization of  $x^N$ . Let  $a_N(x^N)$  and  $t_N(x^N, u^N)$  be measurable functions. If

$$a_N(\underline{x}^N) \rightarrow 0 \Rightarrow t_N(\underline{x}^N, u^N) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1),$$

and

$$a_N(x^N) \xrightarrow{P} 0,$$

then  $t_N(x^N, u^N) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1)$ .

*Proof of Lemma 1.* As a first step we note that the result follows easily from the first condition in the lemma and  $a_N(x^N) \xrightarrow{a.s.} 0$  using the Dominated Convergence

The

with  
N  
onv  
that  
only  
if w  
mar  
dist  
(  
dist  
with

and  
 $h_2(l$   
Wit  
 $(x^1)$

Lem.  
 $N \times$   
1 fo

Pro

Theorem. Under those assumptions

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{Prob}\{t_N(x^N, u^N) \leq z\} &= \lim_{N \rightarrow \infty} \int_x \text{Prob}\{t_N(\underline{x}^N, u^N) \leq z\} d\mu(x) \\ &= \int_x \lim_{N \rightarrow \infty} \text{Prob}\{t_N(\underline{x}^N, u^N) \leq z\} d\mu(x) \\ &= \int_x \Phi(z) d\mu(x) = \Phi(z) \end{aligned}$$

with the last line following from the almost sure convergence of  $a_N(x^N)$ .

Next we show that only convergence in probability of  $a_N(x^N)$ , not almost sure convergence, is necessary. To see this, let  $G_N(z) = \text{Prob}\{t_N(x^N, u^N) \leq z\}$ . Note that  $G_N(z)$  depends on the joint distribution of the sequence of matrices  $\{x^N u^N\}$  only through the distribution of  $x^N u^N$ . Hence the sequence  $\{G_N(z)\}$  is unchanged if we choose any other joint distribution on the sequence  $\{x^N u^N\}$  with the same marginal distribution on each  $x^N u^N$ . It is always possible to choose such a joint distribution so that  $a_N(x^N) \xrightarrow{a.s.} 0$  and hence we know  $G_N(z) \rightarrow \Phi(z)$ .

(To see that such a joint distribution exists, write  $H_N$  for the cumulative distribution function of  $a_N(x^N)$ , let  $h_N: D^N \rightarrow D^{N+1}$  be a sequence of mappings with

$$H_{N+1}(h_N(x^N)) = H_N(x^N),$$

and let the joint distribution on  $\{x^N\}$  be the distribution of  $\{x^1, h_1(x^1), h_2(h_1(x^1)), \dots\}$ , with the  $u_i^N$  being related to the  $x_i^N$  in the obvious way. With this construction, any realization  $\underline{x}^1, \underline{x}^2, \underline{x}^3, \dots$ , has  $a_N(\underline{x}^N) = H_N^{-1}(H_1(\underline{x}^1)) \rightarrow 0$ .  $\square$

*Lemma A.2.* Let  $\{W_N\}$  be a sequence of symmetric matrices (with  $W_N$  being  $N \times N$ ) such that  $r(W_N)/s(W_N) \rightarrow 0$  as  $N \rightarrow \infty$ . Let  $u^N$  and  $\tilde{u}^N$  be as in Proposition 1 for a given sequence  $\{x_i\}$ . Then,

$$\frac{\tilde{u}^{N'} W_N \tilde{u}^N - u^{N'} W_N u^N}{\sqrt{2s(\Sigma^N W_N \Sigma^N)}} \xrightarrow{P} 0.$$

*Proof of Lemma 2.* To begin, note that

$$\begin{aligned} \tilde{u}_i - u_i &= -\frac{\partial f}{\partial \alpha} (x_i; \alpha_0)(\tilde{\alpha} - \alpha_0) - \frac{1}{2} \sum_{k,j=1}^{\ell} \frac{\partial^2 f}{\partial \alpha_k \partial \alpha_j} (x_i; \bar{\alpha}(x_i, \tilde{\alpha})) \\ &\quad (\tilde{\alpha}_k - \alpha_{0k})(\tilde{\alpha}_j - \alpha_{0j}), \end{aligned}$$



for some  $\bar{\alpha}(x_i, \tilde{\alpha})$  between  $\alpha_0$  and  $\tilde{\alpha}$ . Let

$$B_1 = \sup_x \left| \frac{\partial f}{\partial \alpha} (x; \alpha_0) \right|$$

$$B_2(\alpha) = \sup_{t \in [0,1], x \in D, k, j} \left| \frac{\partial^2 f}{\partial \alpha_k \partial \alpha_j} (x; \alpha_0 t + \alpha(1-t)) \right| \frac{\rho^2}{2}$$

Because  $f$  has two continuous derivatives and  $D$  is compact,  $B_1$  and  $B_2(\alpha)$  exist, and  $B_2(\alpha)$  is continuous. If we write

$$\tilde{u}_i - u_i = v_{1i}(\tilde{\alpha} - \alpha_0) + \tilde{v}_{2i},$$

we immediately have  $|v_{1i}| \leq B_1$  and  $|\tilde{v}_{2i}| \leq B_2(\tilde{\alpha})(\tilde{\alpha} - \alpha_0)(\tilde{\alpha} - \alpha_0)$ .

Now, consider the expansion,

$$\begin{aligned} \tilde{u}' W_N \tilde{u} - u' W_N u &= (u' W_N u + 2(\tilde{u} - u)' W_N u + (\tilde{u} - u)' W_N (\tilde{u} - u)) - u' W_N u \\ &= 2((\tilde{\alpha} - \alpha_0)' v_1' W_N u) + 2(\tilde{v}_2' W_N u) + (\tilde{\alpha} - \alpha_0)' v_1' W_N v_1 (\tilde{\alpha} - \alpha_0) \\ &\quad + 2(\tilde{\alpha} - \alpha_0)' v_1' W_N \tilde{v}_2 + \tilde{v}_2' W_N \tilde{v}_2. \end{aligned}$$

We now show that  $(\tilde{u}' W_N \tilde{u} - u' W_N u) / \mathbf{s}(\Sigma^N W_N \Sigma^N) \xrightarrow{p} 0$  by showing that each of the five terms on the right-hand side of the expression above have plim zero when divided by  $\mathbf{s}(W_N)$  (which suffices because  $\mathbf{s}(\Sigma^N W_N \Sigma^N) \geq \underline{\sigma}^2 \mathbf{s}(W_N)$ ).

$$1. (1/\mathbf{s}(W_N))(\tilde{\alpha} - \alpha_0)' v_1' W_N u = \sqrt{N}(\tilde{\alpha} - \alpha_0)' (1/\sqrt{N} \mathbf{s}(W_N)) v_1' W_N u.$$

$\sqrt{N}(\tilde{\alpha} - \alpha_0)$  has an asymptotic distribution. The vector it is multiplied by has

$$\text{Var} \left( \frac{1}{\sqrt{N} \mathbf{s}(W_N)} v_1' W_N u \right) \leq \frac{\bar{\sigma}^2}{N \mathbf{s}(W_N)^2} v_1' W_N W_N v_1.$$

Each column of  $v_1$  is an  $N \times 1$  vector of norm  $\leq \sqrt{N} B_1$ , so each column of  $W_N v_1$  has a norm of at most  $r(W_N) \sqrt{N} B_1$ . Each element of  $v_1' W_N W_N v_1$  is then at most  $N r(W_N)^2 B_1^2$ , and the variance-covariance matrix thus goes to the zero matrix.

$$2. |(1/\mathbf{s}(W_N)) \tilde{v}_2' W_N u|^2 \leq (1/\mathbf{s}(W_N)^2) \|\tilde{v}_2 W_N\|^2 \|u\|^2 \leq (r(W_N)^2 / \mathbf{s}(W_N)^2) \|\tilde{v}_2\|^2 \|u\|^2.$$

$\text{Var}(u_i) \leq \bar{\sigma}^2$  and  $E(u_i^4) \leq m$  implies that  $\|u\|/N = O_p(1)$ . Also,

$$N \|\tilde{v}_2\|^2 \leq N N B_2(\tilde{\alpha})^2 (\tilde{\alpha} - \alpha_0)' (\tilde{\alpha} - \alpha_0) \cdot (\tilde{\alpha} - \alpha_0)' (\tilde{\alpha} - \alpha_0)$$

As  $N \rightarrow \infty$ ,  $B_2(\tilde{\alpha})^2 \xrightarrow{p} B_2(\alpha_0)^2$  and  $(\tilde{\alpha} - \alpha_0)' (\tilde{\alpha} - \alpha_0) N$  is bounded in probability so  $N \|\tilde{v}_2\|^2 = O_p(1)$ . Hence,  $r(W_N) / \mathbf{s}(W_N) \rightarrow 0$  implies that term 2 has plim 0.

$$3. (1/\mathbf{s}(W_N))(\tilde{\alpha} - \alpha_0)' v_1' W_N v_1 (\tilde{\alpha} - \alpha_0) = \sqrt{N}(\tilde{\alpha} - \alpha_0)' (v_1' W_N v_1 / N \mathbf{s}(W_N)) \sqrt{N}(\tilde{\alpha} - \alpha_0).$$

T  
B<sub>1</sub><sup>2</sup> n  
4.

5.

Lem  
N ×  
1 fo

Pro.

and

F  
that  
ijth  
hav

We  
has

The middle term is a  $\ell \times \ell$  matrix, and as in 1, each term is bounded by  $B_1^2 r(W_N)/s(W_N)$  so the matrix converges in probability to zero.

4. The fourth term has

$$\begin{aligned} \left| \frac{1}{s(W_N)^2} (\tilde{\alpha} - \alpha_0)' v_1' W_N \tilde{v}_2 \right|^2 &\leq \|\sqrt{N}(\tilde{\alpha} - \alpha_0)\|^2 \|v_1' W_N \tilde{v}_2\|^2 \frac{1}{N s(W_N)^2} \\ &\leq \|\sqrt{N}(\tilde{\alpha} - \alpha_0)\|^2 \frac{1}{N s(W_N)^2} \ell N B_1^2 r(W_N)^2 \frac{N \|\tilde{v}_2\|^2}{N} \end{aligned}$$

$$\xrightarrow{p} 0.$$

$$5. (1/s(W_N)^2) |\tilde{v}_2' W_N \tilde{v}_2|^2 \leq (1/N^2 s(W_N)^2) \|\tilde{v}_2\|^2 N r(W_N)^2 N \|\tilde{v}_2\|^2 \xrightarrow{p} 0. \quad \square$$

*Lemma A.3.* Let  $\{W_N\}$  be a sequence of symmetric matrices (with  $W_N$  being  $N \times N$ ) such that  $r(W_N)/s(W_N) \rightarrow 0$  as  $N \rightarrow \infty$ . Let  $u^N$  and  $\tilde{u}^N$  be as in Proposition 1 for a given sequence  $\{x_i\}$ . Then,

$$\frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij}^2 \tilde{u}_{iN}^2 \tilde{u}_{jN}^2}{s(\Sigma^N W_N \Sigma^N)^2} \xrightarrow{p} 1.$$

*Proof.* We show this in two steps: showing first that

$$\frac{\sum_{ij} w_{ij}^2 u_i^2 u_j^2 - s(\Sigma^N W_N \Sigma^N)^2}{s(\Sigma^N W_N \Sigma^N)^2} \xrightarrow{p} 0$$

and then that

$$\frac{\sum_{ij} w_{ij}^2 \tilde{u}_{iN}^2 \tilde{u}_{jN}^2 - \sum_{ij} w_{ij}^2 u_i^2 u_j^2}{s(\Sigma^N W_N \Sigma^N)^2} \xrightarrow{p} 0.$$

For the first step, let  $z$  be a vector with  $i$ th element  $z_i = u_i^2 - \text{Var}(u_i)$ . Note that  $E(z_i) = 0$  and  $\text{Var}(z_i) = E(u_i^4) - \text{Var}(u_i)^2 < m$ . Let  $W_2$  be a matrix with  $ij$ th element equal to  $w_{ij}^2$ . Let  $v$  be a vector with  $i$ th element  $\text{Var}(u_i)$ . We then have

$$\begin{aligned} \sum_{ij} w_{ij}^2 u_i^2 u_j^2 - s(\Sigma^N W_N \Sigma^N)^2 &= (z + v)' W_2 (z + v) - v' W_2 v \\ &= z' W_2 z + 2z' W_2 v. \end{aligned}$$

We now show that each of the terms on the right-hand side of this expression has plim zero when divided by  $s(\Sigma^N W_N \Sigma^N)^2$ .

First, note that  $z_1, \dots, z_N$  are independent random variables with  $E(z_i) = 0$  and  $\text{Var}(z_i) < m$ .<sup>17</sup> For any symmetric nonnegative matrix with zeros on the diagonal we then have  $E(z'Az) = 0$  and

$$\begin{aligned} \text{Var}(z'Az) &= \text{Var}\left(\sum_{ij} a_{ij}z_iz_j\right) = \sum_{ijk\ell} a_{ij}a_{k\ell} \text{Cov}(z_iz_j, z_kz_\ell) \\ &= 2\sum_{ij} a_{ij}^2 \text{Var}(z_iz_j) < 2m^2 \mathbf{s}(A)^2. \end{aligned}$$

Hence,  $E(z'W_2z) = 0$  and  $\text{Var}(z'W_2z) < 2m^2 \sum_{ij} w_{ij}^4$ , and

$$\begin{aligned} \text{Var}\left(\frac{z'W_2z}{\mathbf{s}(\Sigma^N W_N \Sigma^N)^2}\right) &\leq \frac{2m^2}{\underline{\sigma}^8} \frac{\sum_{ij} w_{ij}^4}{(\sum_{ij} w_{ij}^2)^2} \leq \frac{2m^2 (\max_{ij} w_{ij}^2) \sum_{ij} w_{ij}^2}{\underline{\sigma}^8 (\sum_{ij} w_{ij}^2)^2} \\ &\leq \frac{2m^2 r(W_N)^2}{\underline{\sigma}^8 \mathbf{s}(W_N)^2} \rightarrow 0. \end{aligned}$$

Second, we similarly have  $E(2v'W_2z) = 0$  and

$$\begin{aligned} \text{Var}\left(\frac{2v'W_2z}{\mathbf{s}(\Sigma^N W_N \Sigma^N)^2}\right) &\leq \frac{4}{\underline{\sigma}^8 \mathbf{s}(W_N)^4} \text{Var}\left(\sum_{ij} w_{ij}^2 v_iz_j\right) < \frac{4\bar{\sigma}^4 m}{\underline{\sigma}^8 \mathbf{s}(W_N)^4} \sum_j \left(\sum_i w_{ij}^2\right)^2 \\ &\leq \frac{4\bar{\sigma}^4 m}{\underline{\sigma}^8 \mathbf{s}(W_N)^4} \left(\max_j \sum_i w_{ij}^2\right) \sum_j \left(\sum_i w_{ij}^2\right) \\ &\leq \frac{4\bar{\sigma}^4 m}{\underline{\sigma}^8 \mathbf{s}(W_N)^4} r(W_N)^2 \mathbf{s}(W_N)^2 \rightarrow 0. \end{aligned}$$

Turning now to the second main step, note that

$$\begin{aligned} \frac{\sum_{ij} w_{ij}^2 \tilde{u}_i^2 \tilde{u}_j^2 - \sum_{ij} w_{ij}^2 u_i^2 u_j^2}{\mathbf{s}(\Sigma^N W_N \Sigma^N)^2} &= \frac{\tilde{u}'W_2\tilde{u}^2 - u'W_2u^2}{\mathbf{s}(\Sigma^N W_N \Sigma^N)^2} \\ &= \frac{2(\tilde{u}^2 - u^2)'W_2u^2}{\mathbf{s}(\Sigma^N W_N \Sigma^N)^2} + \frac{(\tilde{u}^2 - u^2)'W_2(\tilde{u}^2 - u^2)}{\mathbf{s}(\Sigma^N W_N \Sigma^N)^2}, \end{aligned}$$

where we have written  $\tilde{u}^2$  for the vector with  $i$ th element  $\tilde{u}_i^2$  and  $u^2$  for the vector with  $i$ th element  $u_i^2$ . To show that both of the terms on the right-hand side of this expression have plim 0 it will suffice to show that (i)  $\|\tilde{u}^2 - u^2\| = O_p(1)$ , (ii)  $r(W_2)/\mathbf{s}(W_N)^2 \rightarrow 0$ , and (iii)  $\|W_2u^2/\mathbf{s}(W_N)^2\| \xrightarrow{p} 0$ .

Result (i) is standard:  $\|\tilde{u}^2 - u^2\|$  is just the difference between the sum of squared residuals and the sum of squared errors.

<sup>17</sup> Here as in Lemma A.2 the sequence  $\{x_i\}$  is taken to be fixed and thus we write  $E(z_i)$  rather than  $E(z_i|x_i)$  and similarly for other expectations and variances.

with  $E(z_i) = 0$   
 h zeros on the

To derive result (ii) note that for any symmetric matrix  $A$ ,  $r(A) \leq s(A)$ . Hence,  $r(W_2)/s(W_N)^2 \leq s(W_2)/s(W_N)^2$ , and (ii) follows from

$$\frac{s(W_2)^2}{s(W_N)^4} = \frac{\sum_{ij} w_{ij}^4}{s(W_N)^4} \leq \frac{(\max_{ij} w_{ij}^2) \sum_{ij} w_{ij}^2}{s(W_N)^4} \leq \frac{r(W_N)^2 s(W_N)^2}{s(W_N)^4} \rightarrow 0.$$

Finally, to derive result (iii) note that

$$\|W_2 u^2 / s(W_N)^2\| \leq \|W_2 v / s(W_N)^2\| + \|W_2 z / s(W_N)^2\|,$$

where again we have written  $v$  for the vector with  $i$ th element  $v_i = \text{Var}(u_i)$  and  $z$  for  $u^2 - v$ . Using calculations similar to those above it is easy to see that the first of these terms converges to zero.

$$\left\| \frac{W_2 v}{s(W_N)^2} \right\|^2 \leq \frac{\bar{\sigma}^4 \sum_i (\sum_j w_{ij}^2)^2}{s(W_N)^4} \leq \frac{\bar{\sigma}^4 s(W_N)^2 r(W_N)^2}{s(W_N)^4} \rightarrow 0,$$

To see that the second term has plim 0 also we write

$$\|W_2 z / s(W_N)^2\|^2 = s(W_N)^{-4} z' W_2' W_2 z = s(W_N)^{-4} z' B_N z + \sum_{i=1}^N c_{iN} z_i^2,$$

where  $B_N$  is the  $N \times N$  matrix with  $b_{iN} = 0$  and  $b_{ijN} = (W_2' W_2)_{ij}$  for  $i \neq j$  and  $c_{iN}$  is the  $i$ th element of  $W_2' W_2 / s(W_N)^4$ . To see that  $s(W_N)^{-4} z' B_N z \xrightarrow{p} 0$  we note as before that  $E(z' B_N z) = 0$  and

$$\text{Var}\left(\frac{z' B_N z}{s(W_N)^4}\right) \leq \frac{2m^2 s(B_N)^2}{s(W_N)^8} \leq \frac{2m^2 s(W_2' W_2)^2}{s(W_N)^8} \leq \frac{2m^2 s(W_2)^4}{s(W_N)^8} \rightarrow 0,$$

with the second to last conclusion following from the fact that for any symmetric matrix  $A$ ,  $s(A'A)^2 = \sum_i \lambda_i^4 \leq (\sum_i \lambda_i^2)^2 = s(A)^4$ . Finally, to see that  $\sum_{i=1}^N c_{iN} z_i^2 \xrightarrow{p} 0$  as  $N \rightarrow \infty$  we note that Theorem 3.4.9 of Taylor (1978) concludes that a weighted sum of independent random variables of the form  $\sum_{i=1}^N c_{iN} e_i$  has plim zero provided that five conditions hold: (1)  $E(e_i) = 0$ ; (2)  $E(|e_i|) < \infty$ ; (3)  $\max_i |c_{iN}| \rightarrow 0$  as  $N \rightarrow \infty$ ; (4) there exists a  $C$  such that  $\sum_{i=1}^N |c_{iN}| \leq C$  for all  $N$ ; and (5) there exists a random variable  $e$  such that  $\text{Prob}(|e_i| \geq t) \leq \text{Prob}(|e| \geq t)$  for all  $t \geq 0$  and all  $N$ . Each of these hypotheses holds for  $e_i = z_i^2 - E(z_i^2)$ : the first is trivial; the second and fifth are immediate consequences of  $z_i^2$  being nonnegative and the assumed uniform bounds on the second and fourth moments of the  $u_i$  conditional on  $x_i$ ; the third and fourth follow from the facts that  $c_{iN} = s(W_N)^{-4} \sum_j w_{ijN}^4$  is nonnegative and  $\sum_{i=1}^N c_{iN} = s(W_N)^{-4} s(W_2)^2 \rightarrow 0$  as  $N \rightarrow \infty$ . Applying the theorem gives that

$$\sum_{i=1}^N c_{iN} (z_i^2 - E(z_i^2)) \xrightarrow{p} 0,$$

$\sum_{i=1}^N c_{iN} z_i^2$

$u^2$  for the vector  
 t-hand side of this  
 $- u^2\| = O_p(1)$ , (ii)

ween the sum of

write  $E(z_i)$  rather than

and the desired result follows from noting that  $\sum_{i=1}^N c_{iN} E(z_i^2) \leq m \sum_i c_{iN} \rightarrow 0$  as well.  $\square$

*Proof of Corollary 1.* Clearly  $w_{ii}^s = 0$ , so it suffices to show that  $r(W_N^s)/s(W_N^s) \xrightarrow{p} 0$  and  $(1+d)/\sqrt{2s(W_N^s)} \xrightarrow{p} 0$ . Note first that because  $W$  and  $W'$  have the same eigenvalues,

$$\begin{aligned} r(W_N^s) &= \sup_{v \neq 0} \frac{1 \|(W_N + W'_N)v\|}{2 \|v\|} \leq \frac{1}{2} \sup_{v \neq 0} \frac{\|W_N v\|}{\|v\|} + \frac{1}{2} \sup_{v \neq 0} \frac{\|W'_N v\|}{\|v\|} \\ &= \frac{1}{2} (r(W_N) + r(W'_N)) = r(W_N). \end{aligned}$$

Also, because  $K$  is nonnegative

$$s(W_N^s)^2 = \sum_{ij} \left( \frac{w_{ij} + w_{ji}}{2} \right)^2 = \frac{1}{2} s(W_N)^2 + \frac{1}{2} \sum_{ij} w_{ij} w_{ji} \geq \frac{1}{2} s(W_N)^2.$$

Hence it suffices to show that  $r(W_N)/s(W_N) \xrightarrow{p} 0$  and  $1/s(W_N) \xrightarrow{p} 0$ . Let  $W_N^*$  be the  $N \times N$  matrix with  $w_{ii}^* = 1$  if the  $i$ th row of  $W_N$  is identically zero and  $w_{ij}^* = w_{ij}$  for all other  $i, j$ .  $W_N^*$  is a Markov transition matrix so  $r(W_N) \leq r(W_N^*) = 1$ . Hence, we need only show  $s(W_N) \xrightarrow{p} \infty$ . This follows directly from standard results.

To see this, note that when the model  $y = m(x) + \varepsilon$  (with  $m(x)$  continuous and  $\varepsilon$  and *i.i.d.*  $\mathcal{N}(0, \sigma^2)$  error) is estimated by  $\hat{y} = W_N y$ , the average conditional variance of  $\hat{y}$  is

$$\begin{aligned} \frac{1}{N} \sum_i \text{Var}(\hat{y}_i | x_i, \dots, x_N) &= \frac{1}{N} \sum_i \text{Var} \left( \sum_j w_{ij} y_j \right) \\ &= \frac{\sigma^2}{N} \sum_i \sum_j w_{ij}^2 = \frac{\sigma^2}{N} s(W_N)^2. \end{aligned}$$

The fact that the average conditional variance of the kernel estimator in this environment is  $O_p(N^{-1}h^{-d})$  thus implies that  $s(W_N) = O_p(h^{-d/2})$ .  $\square$

*Proof of Corollary 2.* Note that  $\mathcal{T}_N$  is of the same form as the test statistic in Proposition 1 where  $W_N$  is the symmetric matrix given by

$$w_{ij} = \begin{cases} \frac{1}{C_{kN} - 1} & \text{if } i \neq j, x_i, x_j \in P_{kN}, \\ 0 & \text{otherwise.} \end{cases}$$

Again  $W_N$  is non-negative with all rows summing to at most one, so  $r(W_N) \leq 1$  with equality holding provided that not all bins are empty. Hence, we need only show  $1/s(W_N) \xrightarrow{p} 0$ .

To see this, note that  $\mathbf{s}(W_N)^2 \geq \sum_k a_{kN}$ , where  $a_{kN}$  is a random variable given by  $a_{kN} = 0$  if  $C_{kN} \leq 1$ ,  $a_{kN} = 1$  if  $C_{kN} \geq 2$ .

$$E\left(\frac{\sum_k a_{kN}}{m(N)}\right) \geq \frac{m(N)}{m(N)} \inf_k E(a_{kN}) = \inf_k (1 - ((1 - v_k)^N + Nv_k(1 - v_k)^{N-1})) \rightarrow 1,$$

where we have written  $v_k$  for  $v(P_{kN})$ .

$$\begin{aligned} \text{Var}\left(\frac{\sum_k a_{kN}}{m(N)}\right) &\leq \sup_k \text{Var}(a_{kN}) \leq \sup_k E[(a_{kN} - 1)^2] \\ &= \sup_k ((1 - v_k)^N + Nv_k(1 - v_k)^{N-1}) \rightarrow 0. \end{aligned}$$

Therefore,  $r(W_N)/\mathbf{s}(W_N) = O_p(1/\sqrt{m(N)})$ , as desired.  $\square$

*Proof of Proposition 2.* Given the bounds on the error moments and on  $e(x)$ ,  $N^{2\xi-1}\mathbf{s}(\Sigma^N W_N^s \Sigma^N) \xrightarrow{p} 0$  implies that  $N^{2\xi-1}\mathbf{s}(\tilde{U}^N W_N^s \tilde{U}^N) \xrightarrow{p} 0$ . Hence, to show that  $\mathcal{F}_N \xrightarrow{p} \infty$  it suffices to show that  $N^{2\xi-1}\tilde{u}'W_N\tilde{u}$  is bounded away from zero (in probability) as  $N \rightarrow \infty$ . To see this note that

$$\tilde{u}_i = y_i - f(x_i; \tilde{\alpha}) = u_i + g_N(x_i) - f(x_i; \alpha_N^*) + f(x_i; \alpha_N^*) - f(x_i; \tilde{\alpha}).$$

Writing  $e_N$  for  $g_N - f^*$  we have

$$\begin{aligned} N^{2\xi-1}\tilde{u}'W_N\tilde{u} &= N^{2\xi-1}[u'W_Nu + 2u'W_N(f^* - \tilde{f}) + 2u'W_Ne_N \\ &\quad + (f^* - \tilde{f})'W_N(f^* - \tilde{f}) \\ &\quad + 2(f^* - \tilde{f})'W_Ne_N + e_N'W_Ne_N] \end{aligned}$$

From the proof of Lemma A.2 and  $N^{2\xi-1}\mathbf{s}(\Sigma^N W_N^s \Sigma^N) \xrightarrow{p} 0$ , we know that  $N^{2\xi-1}u'W_Nu$ ,  $N^{2\xi-1}u'W_N(f^* - \tilde{f})$ , and  $N^{2\xi-1}(f^* - \tilde{f})'W_N(f^* - \tilde{f})$  each have plim zero.

To see that  $N^{2\xi-1}(f^* - \tilde{f})'W_Ne_N \xrightarrow{p} 0$  we write  $(f^* - \tilde{f})' = (\tilde{\alpha} - \alpha_N^*)'v_1' + \tilde{v}_2'$  as in the proof of Lemma A.2, and first note that  $N^{2\xi-1}(\tilde{\alpha} - \alpha_N^*)'v_1'W_Ne_N = \sqrt{N}(\tilde{\alpha} - \alpha_N^*)'v_1'W_NN^\xi e_N N^{\xi-3/2}$ , which has plim zero because  $\sqrt{N}(\tilde{\alpha} - \alpha_N^*)$  has an asymptotic distribution and

$$\|v_1'W_NN^\xi e_N N^{\xi-3/2}\|^2 \leq \ell^2 B_1^2 \frac{1}{N} \|N^\xi e_N\|^2 N^{2\xi-1}.$$

$N^{2\xi-1}\tilde{v}_2'W_Ne_N$  has plim zero because

$$\|\tilde{v}_2'W_Ne_N\| N^{2\xi-1} \leq N^{2\xi-1}\|\tilde{v}_2\|\|e_N\| = N^{2\xi-1}O_p(1/\sqrt{N})O_p(\sqrt{N} \cdot N^{-\xi}).$$

Finally,

$$N^{2\xi-1} e_N' W_N e_N = N^{2\xi-1} \|e_N\|^2 + N^{2\xi-1} e_N' (W_N e_N - e_N).$$

The first term is

$$N^{2\xi-1} \|e_N\|^2 = \frac{1}{N} \sum_i (N^\xi e_N(x_i))^2 \rightarrow \int_D e(x)^2 p(x) dx.$$

The second term has magnitude at most  $N^{2\xi-1} \|e_N\| \|W_N e_N - e_N\| \leq N^{2\xi-1} (1 - \underline{\delta}) \|e_N\|^2$  with probability approaching one. Hence,

$$\Pr\{N^{2\xi-1} \tilde{u}' W \tilde{u} > \underline{\delta}/2 \int_D e(x)^2 p(x) dx\} \rightarrow 1$$

as desired.  $\square$

## References

- Azzalini, A., Bowman, A., Härdle, W., 1989. On the use of nonparametric regression for model checking. *Biometrika* 76, 1–11.
- Bierens, H., 1982. Consistent model specification tests. *Journal of Econometrics* 20, 105–134.
- Bierens, H., 1984. Model specification testing for time series regression. *Journal of Econometrics* 26, 323–353.
- Bierens, H., 1990. A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.
- Bierens, H., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.
- Chevalier, J., Ellison, G., 1997. Risk-taking by mutual funds as a response to incentives. *Journal of Political Economy* 105, 1167–1200.
- Davidson, R.J., MacKinnon, J., 1981. Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49, 781–793.
- Delgado, M.A., Stengos, T., 1994. Semiparametric testing of non-nested models. *Review of Economic Studies* 61, 291–303.
- deJong, P., Bierens, H.J., 1994. On the limiting behavior of a Chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory* 10, 53–90.
- Duncan, A., Jones, A., 1994. On the specification of labour supply models: A nonparametric evaluation. Discussion Paper No. 94/6, University of York.
- Ellison, G., Ellison, S.F., 1998. A simple framework for nonparametric specification testing. NBER Technical Working Paper 234.
- Eubank, R., Spiegelman, C., 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., Li, Q., 1996. Consistent model misspecification tests: omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.
- Gozalo, P., 1993. A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* 9, 451–477.

Härd  
Härd  
o  
Haus  
Hida  
o  
Hon  
E  
Horc  
E  
Jayas  
J  
Lofts  
fu  
Mikk  
th  
New  
n  
Ram.  
J  
Rodr  
W  
Tauc  
E  
Tayl  
S  
Tuke  
S  
Whit  
C  
Whit  
U  
Woo  
T  
Yatc  
P  
Zhen  
J

$e_N$  |||  $W_N e_N - e_N$  |||  
ce,

regression for model

rics 20, 105–134.

al of *Econometrics* 26,

m. *Econometrica* 58,

tional moment tests.

incentives. *Journal of*

on in the presence of

odels. *Review of Eco-*

e test if the number of  
, 53–90.

els: A nonparametric

gication testing. NBER

del via nonparametric  
, 387–392.

es and semiparametric

imation of regression

- Härdle, W., 1990. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.
- Hausman, J., 1978. Specification tests in econometrics. *Econometrica* 46, 1251–1271.
- Hidalgo, J., 1992. A nonparametric test for nested and non-nested models. Mimeo, London School of Economics.
- Hong, Y., White, H., 1995. Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.
- Horowitz, J., Härdle, W., 1994. Testing a parametric model against a semiparametric alternative. *Econometric Theory* 10, 821–848.
- Jayasuriya, B., 1996. Testing for polynomial regression using nonparametric regression techniques. *Journal of the American Statistical Association* 96, 1626–1631.
- Loftsgaarden, D., Quesenberry, G., 1965. A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics* 36, 1049–1051.
- Mikosch, T., 1991. Functional limit theorems for random quadratic forms. *Stochastic Processes and their Applications* 37, 81–98.
- Newey, W., 1985. Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53, 1047–1070.
- Ramsey, J.B., 1969. Tests for specification errors in classical least squares regression analysis. *Journal of the Royal Statistical Society, Series B* 31, 350–371.
- Rodriguez, D., Stoker, T., 1993. Regression test of semiparametric index model specification. Working paper 9200-6, Center for Energy and Environmental Policy.
- Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415–443.
- Taylor, R.L., 1978. *Stochastic Convergence of Weighted Sums of Random Elements in Linear Spaces*, Springer, Berlin.
- Tukey, J.W., 1961. Curves as parameters and touch estimation. *Proceedings of the Fourth Berkeley Symposium*, pp. 681–694.
- White, H., 1987. Specification testing in dynamic models. *Advances in Econometrics Fifth World Congress*, vol. I. Cambridge University Press, Cambridge, pp. 1–58.
- White, H., Hong, Y., 1993. M testing using finite and infinite dimensional parametric estimators. University of California at San Diego, Department of Economics Working paper 93-01.
- Wooldridge, J., 1992. A test for functional form against nonparametric alternatives. *Econometric Theory* 8, 452–475.
- Yatchew, A., 1992. Nonparametric regression tests based on an infinite dimensional least squares procedure. *Econometric Theory* 8, 435–451.
- Zheng, J., 1996. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.