TECHNICAL WORKING PAPER SERIES

A SIMPLE FRAMEWORK FOR NONPARAMETRIC SPECIFICATION TESTING

Glenn Ellison
Sara Fisher Ellison

Technical Working Paper 234
http://www.nber.org/papers/T0234

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1998

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-95-15094 to the NBER 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." Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

A Simple Framework for Nonparametric
Specification Testing
Glenn Ellison and Sara Fisher Ellison
NBER Technical Working Paper No. 234
September 1998
JEL No. C14, C12

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 and an application to some parametric models of gasoline demand is presented.

Glenn Ellison
MIT Department of Economics
50 Memorial Drive
Cambridge, MA 02139
and NBER
gellison@mit.edu

Sara Fisher Ellison
MIT Department of Economics
50 Memorial Drive
Cambridge, MA 02139
sellison@mit.edu

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,[1] 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 concern a large recent literature has examined the behavior of specification tests which exploit nonparametric techniques.[2] 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, explore the finite sample performance of specific implementations of the test via Monte Carlo simulations, and present an application.

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.[3] Our testing framework can also been 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

[1] See, for example, Ramsey (1969), Hausman (1978), Davidson and MacKinnon (1981), Newey (1985), Tauchen (1985), and White (1987).

[2] Work in this vein includes Azzalini, Bowman, and Härdle (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).

[3] This, for example, is the approach of Hong and White (1995), Eubank and Spiegelman (1990), and Wooldridge (1992).

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.[4]

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 which 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.[5] Section 4 presents the results of several sets

[4] A previous version of this paper showed the equivalence between a quadratic form test and an othogonality test in some (but only some) situations.

[5] 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 $N^{-1/2}$.

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. Simulations also suggest that the degradation in performance when one uses a bin implementation of the test is moderate, so that such a test may be a reasonable choice for some applications. Section 5 presents an application involving a model of gasoline demand, which serves to illustrate the performance of the test, the finite sample correction, and binning in a practical setting.

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 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 $\mathcal{T}_N$ 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 $\mathrm{r}(A)$ for the spectral radius of A defined by

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

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

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

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

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 $D \subset \Re^d$, and $\{u_i\}$ is a sequence of independent random variables with u_i independent of x_j for $j \neq i$, $E(u_i|x_i) = 0$, $0 < \underline{\sigma}^2 \leq Var(u_i|x_i) \leq \overline{\sigma}^2 < \infty$, and $E(u_i^4|x_i) \leq m < \infty$ for all i and x_i. Assume also that $f : D \times \Re^\ell \rightarrow \Re$ is twice continuously differentiable. Let $\tilde{\alpha}_N$ be a $\sqrt{N}$-consistent estimate for α_0. Define $\tilde{u}_{iN} \equiv y_i - f(x_i; \tilde{\alpha}_N)$. Write $\tilde{u}^N$ for the N-vector $(\tilde{u}_{1N}, \ldots, \tilde{u}_{NN})$, and $\tilde{U}^N$ for the $N \times N$ diagonal matrix with iith element $\tilde{u}_{iN}$. Let $W_N : D^N \rightarrow \Re^{N^2}$ be a function associating a symmetric $N \times N$ matrix to each realization of $(x_1 \ldots x_N)$. Suppose that $w_{ii} = 0$ for $i = 1, 2, \ldots, N$,* $\mathrm{r}(W_N)/\mathrm{s}(W_N) \xrightarrow{p} 0$ as $N \rightarrow \infty$, *and that $FSC_N \xrightarrow{p} 0$ as $N \rightarrow \infty$. Let*

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

Then, $\mathcal{T}_N \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1)$.[6]

Remarks:

1. Proposition 1 shows that $\mathcal{T}_N$ converges to a standard normal when $\mathrm{r}(W_N)/\mathrm{s}(W_N) \xrightarrow{p} 0$. 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 by

[6] 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 $W_N(x_1, \ldots, x_N)$. 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.

just computing $r(W_N)/s(W_N)$ and seeing if it approaches zero. 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. For the particular implementations of the test reported on in Table 1 of section 4.1, the true size of a test with 5% asymptotic critical values is between 4% and 6% in each of the parameter combinations for which $r(W_N)/s(W_N)$ is 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

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

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.[7] 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 and discuss its performance in simulations in section 4.1.

3. Several of the restrictions on W_N in the proposition should not in practice limit its applicability. For example, if one wants to base a test on a matrix W which is asymmetric,

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

one can instead base a test on $W_N^s = (W_N + W_N')/2$, which is equivalent because $\tilde{u}'W_N\tilde{u}$ and $\tilde{u}'W_N^s\tilde{u}$ are numerically equal. If one wants to base a test on a matrix whose diagonal elements are not all zero, one can set $\overline{W}_N$ equal to zero on the diagonal and equal to W_N off the diagonal and instead base a test on $\tilde{u}'\overline{W}_N\tilde{u}$. In either case, of course, applying the theorem requires that one check that the appropriate ratio ($\mathrm{r}(W_N^s)/\mathrm{s}(W_N^s)$ or $r(\overline{W}_N)/s(\overline{W}_N)$) converges in probability to 0.

Proof

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 $\mathrm{r}(W_N)/\mathrm{s}(W_N) \to 0$ as $N \to \infty$. To see this, apply Lemma 1 (which is presented in the appendix) with $a_N(x^N) = \mathrm{r}(W_N)/\mathrm{s}(W_N)$ and $t_N(x^N, u^N) = \mathcal{T}_N$.

To show that $\mathcal{T}_N \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}\ (0,1)$ when $\{x_i\}$ is nonstochastic and $\mathrm{r}(W_N)/\mathrm{s}(W_N) \to 0$ we use three additional lemmas which are presented in the appendix. First, a central limit theorem for quadratic forms, which is similar to a result in de Jong (1987) and is presented as Lemma 2 in the appendix, implies that

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

where Σ^N is an $N \times N$ diagonal matrix with the iith element equal to the standard deviation of u_i.

Two additional lemmas then let us conlude that $\mathcal{T}_N$ also converges in distribution to a standard normal. The conclusion of Lemma 3 is that

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

which implies that

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

Lemma 4 establishes that

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

which implies that

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

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

QED.

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 Λ diagonal and Φ the orthogonal matrix of eigenvectors of A. If u is a vector of independent random variables we then have

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

where $v = \Phi u$ is a vector of uncorrelated random variables. We thus have that $u'Wu$ 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 $\mathrm{r}(A)^2$) becomes arbitrarily small compared to the sum of the squares of the weights (which is equal to $\mathrm{s}(A)^2$).

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

$$\begin{aligned} E(\tilde{u}'W\tilde{u}) &= E(u'(I-P_X)W(I-P_X)u) \\ &= E(u'Wu) - E(u'P_XWu) - E(u'WP_Xu) + E(u'P_XWP_Xu) \\ &= -\sigma^2 Tr(P_XW) = -\sigma^2 Tr((X'X)^{-1}X'WX). \end{aligned}$$

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

Writing $\hat{X}$ for WX, the k^{th} diagonal element of the matrix $(X'X)^{-1}X'WX$ 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 $1+rank(X)$ such regression coefficents 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 $\hat{y} = Wy$. 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 $p(x) \geq \underline{p} > 0$ on D.*

Let the kernel $K(x)$ be a nonnegative function satisfying $\int_{\Re^d} K(x)dx = 1$ and $\int_{\Re^d} K(x)^2dx < \infty$, 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 $W_N^s = (W_N + W_N')/2$. If $h_N \to 0$ and $Nh_N^d \to \infty$ then

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

<u>Proof</u>

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

$$\mathrm{r}(W_N^s) = \sup_{v\neq 0} \frac{1}{2}\frac{\|(W_N + W_N')v\|}{\|v\|} \leq \frac{1}{2}\sup_{v\neq 0}\frac{\|W_N v\|}{\|v\|} + \frac{1}{2}\sup_{v\neq 0}\frac{\|W_N' v\|}{\|v\|} = \frac{1}{2}(\mathrm{r}(W_N) + \mathrm{r}(W_N')) = \mathrm{r}(W_N).$$

Also, because K is nonnegative

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

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 $s(W_N) \xrightarrow{p} \infty$.

We show more precisely that $s(W_N) = O_p(h^{-d/2})$.

$$\begin{aligned} h^d s(W_N)^2 &= \frac{1}{N}\sum_i \frac{\sum_{j\neq i}\frac{1}{Nh^d}K((x_i - x_j)/h)^2}{[\sum_{j\neq i}\frac{1}{Nh^d}K((x_i - x_j)/h)]^2} \\ &= \frac{1}{N}\sum_i p(x_i)^{-1}\int_{\Re^d} K(s)^2 ds + \frac{1}{N}\sum_i \tilde{z}_{iN} \end{aligned}$$

where $\tilde{z}_{iN} = \frac{\tilde{a}_{iN}}{\tilde{b}_{iN}^2} - p(x_i)^{-1}\int_{\Re^d} K(s)^2 ds$ with $\tilde{a}_{iN} \equiv \sum_{j\neq i}\frac{1}{Nh^d}K((x_i - x_j)/h)^2$ and $\tilde{b}_{iN} \equiv \sum_{j\neq i}\frac{1}{Nh^d}K((x_i - x_j)/h)$. The first term on the RHS of the equation for $h^d s(W_N)^2$ has plim $\int_D p(x)p(x)^{-1}dx \int_{\Re^d} K(s)^2 ds = \mu(D)\int_{\Re^d} K(s)^2 ds$ (where $\mu(D)$ is the Lesbegue measure of D). This is finite and nonzero, so it will suffice to complete the proof to show that $\frac{1}{N}\sum_i \tilde{z}_{iN} \xrightarrow{p} 0$.

To see this, note first that $\tilde{z}_{1N}, \ldots, \tilde{z}_{NN}$ are (nonindependent) identically distributed random variables. Hence, $E\left(\frac{1}{N}\sum_i \tilde{z}_{iN}\right) = E(\tilde{z}_{1N})$ and $Var\left(\frac{1}{N}\sum_i \tilde{z}_{iN}\right) \leq Var(\tilde{z}_{1N})$. It thus suffices to show that $E(\tilde{z}_{1N}) \longrightarrow 0$ and $Var(\tilde{z}_{1N}) \longrightarrow 0$ as $N \rightarrow \infty$. Next, observe that $\tilde{b}_{1N}$ is $(N-1)/N$ times a standard kernel density estimate of p evaluated at x_1. Given that K is nonnegative and $\int_{\Re^d} K(x)dx = 1$ it is a standard result (c.f. Theorem 3.1 in Devroye (1987)) that $\int_D |\tilde{b}_{1N}(x_1) - p(x_1)|dx_1 \xrightarrow{p} 0$. Similarly, $\tilde{a}_{1N}$ is like a kernel density estimate, but with the function K^2 playing the role of the kernel integrating to $\int_{\Re^d} K(s)^2 ds$ instead of to one. The assumption that this integral is finite thus allows us to conclude that $\int_D |\tilde{a}_{1N}(x_1) - p(x_1)\int_{\Re^d} K(s)^2 ds|dx_1 \xrightarrow{p} 0$. Because $p(x_1)$ is bounded above and bounded away from zero on D, these two conditions clearly imply that $\tilde{z}_{1N} \xrightarrow{p} 0$. (Given an ϵ we can choose N so that on a set of x_1's of high Lesbegue measure, there is a high probability both that $\tilde{a}_{1N}(x_1)$ is close enough to $p(x_1)\int_{\Re^d} K(s)^2 ds$ and that $\tilde{b}_{1N}(x_1)^2$ is close enough to $p(x_1)^2$ to make $\tilde{z}_{1N}$ close to zero.) Finally, observe that by construction $\tilde{z}_{1N}$ is bounded by $-\underline{p}^{-1}\int_{\Re^d} K(s)^2 ds \leq \tilde{z}_{1N} \leq 1$. Hence $\tilde{z}_{1N} \xrightarrow{p} 0$ implies $E(\tilde{z}_{1N}) \longrightarrow 0$ and $Var(\tilde{z}_{1N}) \longrightarrow 0$ as desired.

QED.

Remarks:

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. The form of Corollary 1 reflects a desire to provide a straightforward example of how Proposition 1 might be applied to establish the asymptotic normality of a class of test statistics and thereby eliminate the need to examine the weight matrices directly. 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 $\mathrm{r}(W_N)$ remains bounded as $N \to \infty$. If one wanted to use a kernel which was sometimes negative, a variety of other assumptions could be added to obtain these conclusions.[8]

3. The calculations in the proof illustrate our earlier comment that nonparametric test statistics may converge to their asymptotic distributions very slowly. The fact that $\mathrm{s}(W_N) = O_p(h^{-d/2})$ 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 $N^{-1/10}$; with two explanatory variables and $h_N = h_0 N^{-1/6}$, it tends to zero like $N^{-1/6}$.[9] Finite sample corrections may thus be important not only when working with small datasets, but even when hundreds of thousands of observations are available.

2.4 "Binning" and other tests

The testing framework can accomodate a wide variety of weight matrices. The orthogonality test motivation points to a number of possible choices: W could be the weight matrix corresponding to any smoother of the form $W_N y$. This includes kernel regression estima-

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

[9]Local power calculations presented later will suggest using window widths which shrink even more slowly.

tors, as well as k-nearest neighbor estimators, splines, orthogonal series estimators, and convolution smoothing.[10] The notion of a quadratic form test detecting spatial correlation in the residuals suggests other possibilities. For instance, 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 is not consistent of course—misspecifications in other X variables could go undetected—but it may have better power in detecting that particular form of misspecification.

One simple alternate implementation is a "bin" version of the test.[11] 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 $O(N^2h)$ 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 ν on D.*

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

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

Then $\mathcal{T}_N \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1)$.

The proof of Corollary 2 is presented in the appendix.[12]

[10] See Härdle (1990) for a discussion of the weight matrices corresponding to the estimators mentioned above and others.

[11] 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.

[12] 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 $N \to \infty$, *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 $\mathrm{r}(W_N) = 1$ and $\mathrm{s}(W_N) \xrightarrow{p} \infty$, 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.

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 $g_N : D \to \Re$ a sequence of functions. Write X for the matrix $(x_1 \ldots x_N)$. Let $\tilde{\alpha}(y, X)$ be an estimator for which there exists a sequence α_N^* such that $\sqrt{N}(\tilde{\alpha}(y,X) - \alpha_N^*) \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,\Omega)$ and $\alpha_N^* \to \alpha_0$ as $N \to \infty$ for some α_0. Let $W_N(X)$ be a sequence of matrices with $w_{ii} = 0\ \forall i$, $\mathrm{r}(W_N) \xrightarrow{p} 1$, and $1/\mathrm{s}(W_N^s) \xrightarrow{p} 0$ as $N \to \infty$. Suppose $FSC_N \xrightarrow{p} 0$ as $N \to \infty$. 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 $\bar{\xi} \equiv \sup\{\xi | N^{2\xi-1}\mathrm{s}(\Sigma^N W_N^s \Sigma^N) \xrightarrow{p} 0\}$, and suppose there exists a constant ξ, $0 \leq \xi < \bar{\xi}$ and a bounded function $e(x)$ with $\int_D e(x)^2 p(x)dx \neq 0$ such that $N^{\xi}(g_N(x) - f(x;\alpha_N^*)) \to e(x)$ uniformly in x. Suppose also that W_N is such that

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

Let

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

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

Proof

Given the bounds on the error moments and on $e(x)$, $N^{2\xi-1}\mathrm{s}(\Sigma^N W_N^s \Sigma^N) \xrightarrow{p} 0$ implies that $N^{2\xi-1}\mathrm{s}(\tilde{U}^N W_N^s \tilde{U}^N) \xrightarrow{p} 0$. Hence, to show that $\mathcal{T}_N \xrightarrow{p} \infty$ it suffices to show that $N^{2\xi-1}\tilde{u}' W_N \tilde{u}$ is bounded away from zero (in probability) as $N \to \infty$. 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

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

From the proof of Lemma 3 and $N^{2\xi-1}\mathrm{s}(\Sigma^N W_N^s \Sigma^N) \xrightarrow{p} 0$, we know that $N^{2\xi-1}u' W_N u$, $N^{2\xi-1}u' W_N (f^* - \tilde{f})$, and $N^{2\xi-1}(f^* - \tilde{f})' W_N (f^* - \tilde{f})$ each have plim zero.

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

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

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

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

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

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

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

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

as desired.

QED.

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 $\bar{\xi} = \frac{1}{2} + \lim_{N\to\infty} \frac{d}{2} \log_N h_N$. For example, if $d = 1$ and the kernel is chosen to be the standard "optimal"[13] 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, *e.g.*, $h_N = h_0/\log N$, the test has power against local alternatives of order $N^{-\xi}$ for all $\xi < \frac{1}{2}$. Similarly, the bin test has $s(W_N^s) = O_p(m(N))$ so $\bar{\xi} = \frac{1}{2} + \lim_{N\to\infty} -\frac{1}{2} log_N m(N)$. We again get $\bar{\xi} = \frac{1}{2}$ if we let the number of bins grow slowly, *e.g.*, $m(N) = O(log(N))$.

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 $N^{-\frac{1}{2}}$ as well.

4 Simulation Results

Here we present Monte Carlo studies of the finite sample power of our tests and the reliability of asymptotic critical values. The first two subsections explore the size and power of a kernel implementation of our test in a variety of settings, with the benchmark of "good" performance being judged by looking numerically at the accuracy of the asymptotic critical values, and by comparing the power of the nonparametric test to the power of an optimal parametric test for each particular form of misspecification. The third subsection attempts to evaluate whether the performance of our test is "good" by comparing its size and power in finite samples with those of several other nonparametric test statistics. In a fourth subsection, we discuss also the extent to which one can maintain adequate performance while reducing the computational burden of performing a nonparametric test by using a

[13]Here, again, we mean optimal kernel in the context of estimation, not testing.

bin weight matrix.

4.1 Distribution

To examine the degree to which the asymptotic critical values may be considered reliable in small samples, we first examined the null distribution of a kernel version of our test statistic when applied to datasets containing 100 observations on one, two, and three independent variables. In each case, x was taken to consist of independent draws from a uniform distribution on $[-1,1]^d$, and y was generated from a linear null $y = x \cdot \mathbf{1} + u$, with $\mathbf{1}$ a vector of ones and u a vector of independent draws from a standard normal distribution.[14] A linear model (with a constant term) was estimated on the data by OLS, and the residuals were used to form the kernel test statistic like that described in Corollary 1 using the finite sample correction $FSC_N = (1+d)/\sqrt{2}s(W_N^s)$. The weight matrices are generated using the Epanechnikov kernel with window widths of the form $h_N = h_{100}(N/100)^{-1/(4+d)}$ for various h_{100}.[15] Table 1 presents descriptive statistics from 10,000 simulations each with several window widths. We report estimates of each test statistic's mean, standard deviation, and 5% critical value, and the size of a test performed with the asymptotic critical values (ACV). The standard errors on the critical values are approximately 0.02.

The true critical values of our test statistics are well approximated by their asymptotic values in the simulations. Our power simulations with 100 observations will suggest the use of kernel window widths near 0.2 in one dimension and 0.4 in two dimensions. For each of these window widths the estimated critical value is about 1.61, and a test relying on 5% asymptotic critical values would falsely reject the null about 4.7% of the time. Such values will be regarded by practitioners as being very good approximations. The null distributions are somewhat skewed to the right — in one dimension with a window width

[14]Note that under a linear null the distribution of our test statistic is independent of the variance of a normally distributed error. Note also that the asymptotic normality in Lemma 1 comes not from a weighted sum of the u's but instead from a weighted sum of squared residuals. Hence, using nonnormal errors would not be expected to increase the departure from normality of the test statistic.

[15]We base window widths for different N on h_{100} in this way in order to compare power across different size data sets. As we have noted, there is no compelling argument for this choice of h_N. In fact, given that these simulation results will suggest that one does not need to use small window widths to obtain tests with an appropriate size while our theoretical results show power to be greater with more slowly shrinking window widths, we would recommend using window widths which shrink more slowly.

of 0.2 we estimate the coefficient of skewness to be 0.87.

Recall that the proof of Proposition 1 shows that the distribution of $\mathcal{T}_N$ converges in distribution to a standard normal whenever $\mathrm{r}(W_N)/\mathrm{s}(W_N)$ approaches 0. While one should not generalize too far from our simulations given that magnitudes of departures from normality in finite samples depend on the form of heteroskedasticity, higher moments of the error distribution, etc., it may be useful to note that $\mathrm{r}(W_N^s)/\mathrm{s}(W_N^s)$ is usually about 0.4 when $N = 100$ and $h = 0.2$ in our one dimensional simulations and is also about 0.4 when $N = 100$ and $h = 0.6$ in our two dimensional simulations.

4.2 Power

We have already seen that our test like some others is capable asymptotically of detecting local alternatives which shrink at a rate slower than $N^{-\frac{1}{2}}$. In this subsection we try to provide some rough intuition for how our test will work in finite samples by looking at its performance in a variety of situations and judging how well it does by comparing its power to that of the optimal Lagrange multiplier tests which one could use to detect each particular misspecifications if one knew which misspecification to look for.

To assess the power of our test in each of the situations described below, we constructed 1000 simulated datasets, with the x's consisting of between 50 and 200 draws from an uniform distribution on $[-1,1]^d$. y was generated from a model $y_i = x_i \cdot 1 + 0.5e(x_{i1}) + u_i$, with $e(x)$ a nonlinear function, and u a vector of independent draws from a standard normal distribution. A misspecified linear null (with a constant term) was estimated on the data by OLS, and the residuals were used to form the kernel test statistics described in Corollary 1. The performance of the nonparametric tests was measured by their ability to detect eight different forms of misspecification $e(x)$, the second through seventh Legendre polynomials (denoted $p_k(x)$), $\sin(2\pi x)$, and $\sin(10\pi x)$. In each case, $e(x)$ is normalized so that $\inf_{a,b} \int_{-1}^{1}(e(x) - (ax + b))^2 dx = 1$. To assess the true power of the tests, we used estimated critical values obtained from separate simulations (some of which are reported in Table 1).[16]

[16] Note that because the estimated critical values are very close to their asymptotic levels (except perhaps for the largest window width) the apparent power of a test using asymptotic critical values would be quite

Table 2 presents the results. The rows of the table give the performance of kernel tests with a window width of $h_{100}(N/100)^{-1/5}$. The numbers in the body of the tables are the proportion of the 1000 simulations which reject the null at the five percent level. The last row of the table includes for comparison the performance of the separate Lagrange multiplier tests designed to detect each misspecification.

While the nonparametric tests are, of course, less powerful than the alternative-specific Lagrange multiplier tests, the results are encouraging. For the less wavy quadratic and cubic alternatives the nonparametric tests can be nearly as powerful. For the more wavy alternatives the nonparametric tests may require more than twice as many data points to obtain the same power as the optimal tests, but can still be reasonably regarded as being able to detect alternatives of a similar magnitude. Of course, the great advantage of nonparametric tests is their flexibility. While the optimal test against one Legendre polynomial would have zero power against the other alternatives (*i.e.* would have a 5% rejection rate) the nonparametric tests show power against a variety of alternatives. This is particularly striking for the tests with the smallest window widths, where the power is nearly identical for the various alternatives.

The nonparametric tests show good power for a range of window widths. For the particular set of alternatives considered, it appears that a window width h_{100} of about 0.2 (perhaps more generally about one tenth the range over which the data have a nontrivial density?) gives a good mix of performance. Larger window widths are more successful in detecting the quadratic and cubic alternatives, while smaller window widths are more successful in detecting the high frequency $\sin(10\pi x)$ misspecification.

Table 3 compares the power of the tests with one, two, and three dimensional X's ($d = 1, 2, 3$). The true model in two and three dimensions is of the form $y_i = x_i \cdot \mathbf{1} + 0.5e(x_{i1}) + u_i$ so that it depends only on the first component of x_i. We should expect a significant diminution in the power of the nonparametric tests, because there is a much richer set of possible misspecifications in higher dimensions. In comparing the one, two, and three dimensional results, this loss of power is evident. It is most severe for the high frequency

similar.

alternatives, because the sparseness of the data in higher dimensions forces one to use a larger window width.

4.3 Comparison with other nonparametric tests

In this subsection, we present simulation results which compare the finite sample performance of our tests to others, both in terms of size and power.

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. By doing so we are able to compare our tests with those of Bierens (1990), Eubank and Spiegelman (1990) and Jayasuriya (1996), Hong and White (1995), Wooldridge (1992), and Yatchew (1988).

Our first comparative table, Table 4, speaks to the ability of applied researchers to rely on the asymptotic critical values of the various tests. For this table, we computed 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 $y_i = 1 + x_{1i} + x_{2i} + \epsilon_i$, with x_{1i} and x_{2i} being generated by setting $x_{1i} = (v_i + v_{1i})/2$ and $x_{2i} = (v_i + v_{2i})/2$ with $\{v_i\}$, $\{v_{1i}\}$, and $\{v_{2i}\}$ being sequences of i.i.d. uniform $[0, 2\pi]$ random variables and errors being independent standard normal random variables. The table compares the size of two particular versions of our test as computed from 10,000 simulations with the sizes of other tests as reported in Hong and White (1995).[17] 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 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).

From the results in the table, we would argue that our test performs extremely well.

[17] We would like to stress that for all tests other than our own we are merely reprinting here numbers which appeared in Hong and White's tables. We did not repeat their simulations.

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 percent, figures which are much closer to 5% than are those for any of the other tests. Several of the other tests often have an ACV size subtantially in excess of 5%. We noted earlier that nonparametric test statistics tend to converge to their asymptotic distributions relatively slowly. This is apparent in the table in that many of the test statistics do no better with 500 observations than they do with 100 observations.[18] Because ACV sizes improve so slowly, finite sample performance is of great importance.

As an extreme demonstration of the degree to which our critical values are reliable in small samples, we also computed the ACV size of our test when applied to this null hypothesis on datasets containing 30 observations. In doing so we find the sizes for the two window widths to be 3.0 and 4.7 percent, which looks quite good in comparison with the performance of the other tests on the much larger datasets.

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 5 again piggybacks on the work of Hong and White (1995) to compare 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). Each has the same distribution of the X's as above, with $y_i = 1 + x_{1i} + x_{2i} + 0.1(v_{1i} - \pi)(v_{2i} - \pi) + \epsilon_i$ in alternative 1, $y_i = (1 + x_{1i} + x_{2i})(1 + exp(-0.01(1 + x_{1i} + x_{2i})^2) + \epsilon_i$ in alternative 2, and $y_i = (1 + x_{1i} + x_{2i})^{-0.5} + \epsilon_i$ in alternative 3.[19] For each alternative we report rejection rates from simulations involving 100 and 300 observations.[20]

Our test performs relatively well in the simulations. Looking at each of the alternatives

[18] We would guess that they would probably show little improvement even with ten thousand observations.

[19] In each case the ϵ_i are again normally distributed. 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).

[20] The rejection rates for the tests other than ours are drawn from Hong and White's tables 3, 4, and 5.

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.

4.4 Bin tests

We have previously noted that it is possible to greatly reduce the computational burden of carrying out our test by using a bin weight matrix. To help assess the degree to which performance is degraded when one uses a bin rather than a kernel weight matrix Table 6 reports rejection rates of bin tests for the same set of alternatives as in Table 3. A comparison of the tables reveals that the power of the bin tests is somewhat less than that of the kernel tests, but is still of a comparable magnitude. For example, with 100 observations the rejection rates of a bin test with a bin width of 0.3 are 0.67, 0.57, 0.43, 0.30, 0.23, and 0.26 against the 2^{nd} through 7^{th} Legendre Polynomials, while those of a kernel test with a window width of 0.3 are 0.82, 0.73, 0.61, 0.49, 0.36, and 0.28.[21] We would conclude that the bin tests are clearly less powerful than kernel tests with comparable window widths, but not so much so as to make performing bin tests an unreasonable choice if computational concerns are important.

5 Application

Broadly speaking, there are two main ways in which this and other consistent tests could be fruitfully used: as a direct omnibus test of an economic theory and as a regression diagnostic. Demand analysis provides a rich empirical setting both for testing theories and

[21] These bin and window widths are those which appear to do best against the lower order polynomials in our tests.

for the use of regression diagnostics. There is a long history of empirical demand studies where specifications are guided by the theory and theoretical implications are then tested empirically. See, for example, Deaton and Muellbauer (1980) for a summary. We will focus here instead on using this test as a regression diagnostic.

In practice, ease of computation and interpretation have most often led to use of models like our (first) model of gasoline demand.

$$\log(q_i) = \beta_0 + \beta_1 \log(y_i) + \beta_2 \log(p_i) + \epsilon_i,$$

where q_i is the quantity of gasoline demanded, y_i is the income (or expenditure) of the consumer, and p_i is the price of the gasoline. This specification dates back as early as Schultz (1938), Wold (1953), and Stone (1954), and perhaps earlier. Its primary appeal is probably that it can be estimated using simple linear regression techniques, and that elasticity estimates can be read off from the parameter estimates.

In testing this and other specifications, the datasets we use are the U.S. Department of Energy's Residential Energy Consumption Survey from 1979-81 and its Residential Transportation Energy Consumption Survey from 1983, 1985, and 1988. We use them as a cross section with indicator variables for the years, yielding 18113 observations. The data are treated as in Hausman and Newey (1995), including the nonparametric "partialling out" of the demographic variables. We are thus left with a data set with price and income as the explanatory variables and gasoline demand as the dependent variable, all with the effect of the demographic variables taken out.

To test the logarithmic specification of demand, we estimated the model by OLS and computed a version of our test with a kernel weight matrix with the window width set to 0.2 after the data had been normalized so that each explanatory variable had unit variance (and with the general finite sample correction). Such a test very clearly rejects the logarithmic specification, with the value of the test statistic being 11.00. Upon rejecting a null specification, a practitioner will often add higher order terms to improve the fit, which leads us to our second and third null models:

$$\log(q_i) = \sum_{m=0}^{d} \sum_{n=0}^{d-m} \beta_{mn} \log(y_i)^m \log(p_i)^n + \epsilon_i$$

for $d = 2, 3$. Constructing similar kernel tests of these specifications (with the weight matrices being based simply on differences in y_i and p_i) we obtain test statistics of 10.14 and 5.12 for the quadratic and cubic specifications, respectively. We would interpret these results as indicating again that enough data are available to use semiparametric techniques and that such an approach should be considered if the shape of the demand functions is potentially relevant to questions in which one is interested.[22]

This application is also a good example of one of the situations (involving a very large number of datapoints and multiple regressors) in which there may be practical reasons to choose to perform a bin test. For comparison with the results above, we computed also versions of our test based on a bin weight matrix with the bin width equal to 0.3. The test statistics we obtained for the logarithmic, quadratic-logarithmic, and cubic-logarithmic specifications are 9.95, 10.92, and 6.79, respectively. Note that these statistics are of comparable magnitudes to the ones obtained from the kernel tests (which have the same asymptotic distribution under the null), and that the test is still easily sufficiently powerful to reject all of the parametric specifications. In this example the convenience of the bin test is a great advantage—running the regressions and performing all three tests took a total of 44 seconds on our Pentium PC, compared with many hours for our (not very efficiently programmed) kernel test.

Because it is so easy to compute, we were also able to conduct Monte Carlo simulations to analyze the null distribution of the bin test. While such simulations are clearly not necessary for the application given the values of the test statistics, they provide a nice illustration of the importance of finite-sample corrections even in what would ordinarily be regarded as a "large" sample. With the bin width we chose, the finite sample correction terms in the test statistics for the three null hypotheses were 0.17, 0.33, and 0.54. In 10000 simulations we estimated the true 5% critical values (when the corrections are used) to be 1.66, 1.56, and 1.44. The finite-sample corrections are thus making the true critical values much closer to the asymptotic prediction of 1.64 than they otherwise would be.

[22]This point is made forcefully in Hausman and Newey (1995), who argue that the choice of functional form can have a great effect on estimates of consumer surplus and deadweight loss, and hence that one must be very cautious in relying on parametric models in addressing such policy-relevant issues as the potential welfare losses which would result from a gasoline tax increase.

6 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, and in the application the finite sample corrections are of great practical importance despite the large sample sizes. Finally, while a version of our test exploiting standard "binning" techniques is not as strong a performer as the kernel version, the fact that such a test can be performed in a matter of seconds even on a large multidimensional dataset makes it attractive in certain circumstances.

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 the simulations 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.

References

Azzalini, A., A. Bowman, W. Härdle (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., and W. Ploberger (1997). "Asymptotic Theory of Integrated Conditional Moment Tests." *Econometrica*, **65**, 1129-1151.

Davidson, R.J., and J. MacKinnon (1981). "Several Tests for Model Specification in the Presence of Alternative Hypotheses." *Econometrica*, **49**, 781-793.

Deaton, A., and J. Muellbauer (1980). *Economics and Consumer Behavior.* Cambridge University Press.

Delgado, M.A., and T. Stengos (1994). "Semiparametric Testing of Non-Nested Models." *Review of Economic Studies*, **61**, 291-303.

deJong, P. (1987). "A Central Limit Theorem for Generalized Quadratic Forms." *Probability Theory and Related Fields*, **75**, 261-277.

deJong, P., and H.J. Bierens (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.

Devroye, L. (1987). *A Course in Density Estimation.* Birkhauser.

Duncan, A., and A. Jones (1994). "On the Specification of Labour Supply Models: A Nonparametric Evaluation." Discussion Paper No. 94/6, University of York.

Chevalier, J., and G. Ellison (1997). "Risk-Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy*, **105**, 1167-1200.

Eubank, R., and C. Spiegelman (1990). "Testing the Goodness of Fit of a Linear Model Via Nonparametric Regression Techniques." *Journal of the American Statistical Association*, **85**, 387-392.

Fan, Y., and Q. Li (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.

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

Härdle, W., and E. Mammen (1993) "Comparing Nonparametric Versus Parametric Regression Fits." *Annals of Statistics*, **21**, 1926-1947.

Hausman, J. (1978). "Specification Tests in Econometrics." *Econometrica*, **46**, 1251-1271.

Hausman, J., and W. Newey (1995). "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss." *Econometrica*, **63**, 1445-1476.

Hidalgo, J. (1992). "A Nonparametric Test for Nested and Non-nested Models." Mimeo, London School of Economics.

Hong, Y., and H. White (1995). "Consistent Specification Testing Via Nonparametric Series Regression." *Econometrica*, **63**, 1133-1159.

Horowitz, J., and W. Härdle (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., and G. Quesenberry (1965): "A Nonparametric Estimate of a Multivariate Density Function," *Annals of Mathematical Statistics*, **36**, 1049-1051.

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., and T. Stoker (1993). "Regression Test of Semiparametric Index Model Specification." Working Paper 9200-6, Center for Energy and Environmental Policy.

Schultz, H. (1938). *The Theory and Measurement of Demand.* University of Chicago Press.

Stone, J.R.N. (1954). "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand." *Economic Journal*, **64**, 511-527.

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-Verlag.

Tukey, J.W. (1961). "Curves as Parameters and Touch Estimation." *Proc 4th Berkeley Symposium*, 681-694.

White, H. (1987). "Specification Testing in Dynamic Models." *Advances in Econometrics, Fifth World Congress*, **I**, 1-58. Cambridge University Press.

White, H., and Y. Hong (1993). "M Testing Using Finite and Infinite Dimensional Parametric Estimators." University of California at San Diego Department of Economics Working

Paper 93-01.

Wold, H. (1953). *Demand Analysis: A Study in Econometrics.* Greenwood Press.

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.

Appendix

Lemma 1 *Let $\{u_i\}$ and $\{x_i\}$ be as above. Write u^N for $(u_1, \ldots, u_N)$, x^N for $(x_1 \ldots x_N)$ 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*

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

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 $a_N(x^N) \overset{\text{a.s.}}{\to} 0$ using the Dominated Convergence Theorem. Under those assumptions

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

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 $G_N(z) = \text{Prob}\{t_N(x^N, u^N) \le z\}$. 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 $a_N(x^N) \xrightarrow{\text{a.s.}} 0$ and hence we know $G_N(z) \to \Phi(z)$.

(To see that such a joint distribution exists, write H_N for the cumulative distribution function of $a_N(x^N)$, let $h_N : D^N \to D^{N+1}$ 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 $\{x^1, h_1(x^1), h_2(h_1(x^1)), \ldots\}$, with the u_i^N being related to the x_i^N in the obvious way. With this construction, any realization $\underline{x}^1, \underline{x}^2, \underline{x}^3, \ldots$, has $a_N(\underline{x}^N) = H_N^{-1}(H_1(\underline{x}^1)) \to 0$.)

QED.

Lemma 2 *Let $\{u_i\}$ be a sequence of independent random variables with $E(u_i) = 0$, $0 < \underline{\sigma}^2 \le Var(u_i) \le \overline{\sigma}^2 < \infty$, and $E(u_i^4) \le m < \infty$. Write u^N for $(u_1, \ldots, u_N)$ and Σ^N be the $N \times N$ diagonal matrix with $\Sigma_{ii}^N = Var(u_i)^{1/2}$. Let $\{W_N\}$ be a sequence of symmetric matrices with $\mathrm{r}(W_N)/\mathrm{s}(W_N) \to 0$ as $N \to \infty$. Then*

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

Proof of Lemma 2

Given a sequence $\{v_i\}$ of independent random variables with $E(v_i) = 0$ and $Var(v_i) = 1$, and a sequence of symmetric matrices $\{A_N\}$ (with A_N being $N \times N$), Theorem 5.2 of deJong (1987) states that

$$\frac{v^{N\prime} A_N v^N}{\sqrt{2}\mathrm{s}(A_N)} \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1)$$

if there exists a function $k(N)$ such that three conditions hold:

(i) $$\frac{k(N)^4}{2\mathrm{s}(A_N)^2} \max_{1 \le i \le N} \sum_{j=1}^{N} a_{ijN}^2 \to 0 \text{ as } N \to \infty$$

(ii) $$\max_{1 \le i \le N} E v_i^2 1_{|v_i| > k(N)} \to 0 \text{ as } N \to \infty$$

(iii) $$\frac{\mathrm{r}(A_N)}{\mathrm{s}(A_N)} \to 0 \text{ as } N \to \infty$$

Let $\overline{u}^N = \Sigma^{-1} u^N$. We then have

$$\frac{u^{N\prime} W_N u^N}{\sqrt{2}\mathrm{s}(\Sigma^N W_N \Sigma^N)} = \frac{\overline{u}^{N\prime} (\Sigma^N W_N \Sigma^N) \overline{u}^N}{\sqrt{2}\mathrm{s}(\Sigma^N W_N \Sigma^N)},$$

so it will suffice to show that $\overline{u}_i$ and $\Sigma^N W_N \Sigma^N$ satisfy the conditions on v_i and A_N in the theorem.

Condition (iii) clearly holds whenever $\mathrm{r}(W_N)/\mathrm{s}(W_N) \to 0$ as $\mathrm{r}(\Sigma^N W_N \Sigma^N) \le \overline{\sigma}^2 \mathrm{r}(W_N)$ and $\mathrm{s}(\Sigma^N W_N \Sigma^N) > \underline{\sigma}^2 \mathrm{s}(W_N)$.

Condition (ii) holds for any function $k(N)$ with $k(N) \to \infty$ as $N \to \infty$ because

$$E(\overline{u}_i^2 1_{|\overline{u}_i| > k(N)}) \le \frac{E(\overline{u}_i^4)}{k(N)^2} = \frac{E(u_i^4)}{Var(u_i)^2 k(N)^2} \le \frac{m}{\underline{\sigma}^4 k(N)^2}.$$

Finally, note that $\sum_{j=1}^N a_{ijN}^2 = \sum_{j=1}^N a_{jiN}^2 = \|A_N e_i\|^2 \le \mathrm{r}(\mathrm{A_N})^2$. Hence,

$$\frac{k(N)^4}{2\mathrm{s}(A_N)^2} \max_{1 \le i \le N} \sum_{j=1}^{N} a_{ijN}^2 \le \frac{\mathrm{r}(A_N)^2}{\mathrm{s}(A_N)^2} \frac{k(N)^4}{2}.$$

Therefore, if we choose $k(N) = \frac{\mathrm{s}(\Sigma^N W_N \Sigma^N)^{1/8}}{\mathrm{r}(\Sigma^N W_N \Sigma^N)^{1/8}}$, we will have both that $k(N) \to \infty$ (so that condition (ii) holds) and that $\frac{\mathrm{r}(\Sigma^N W_N \Sigma^N)^2}{\mathrm{s}(\Sigma^N W_N \Sigma^N)^2} k(N)^4 \to 0$ (so condition (i) holds).

QED.

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

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

Proof of Lemma 3

To begin, note that

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

for some $\overline{\alpha}(x_i, \tilde{\alpha})$ between α_0 and $\tilde{\alpha}$. Let

$$\begin{aligned} B_1 &= \sup_x |\frac{\partial f}{\partial \alpha}(x; \alpha_0)| \\ B_2(\alpha) &= \sup_{t \in [0,1], x \in D, k, j} |\frac{\partial^2 f}{\partial \alpha_k \partial \alpha_j}(x; \alpha_0 t + \alpha(1-t))| \frac{\ell^2}{2} \end{aligned}$$

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

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

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

$$\mathrm{Var}(\frac{1}{\sqrt{N}\mathrm{s}(W_N)} v_1' W_N u) \leq \frac{\overline{\sigma}^2}{N\mathrm{s}(W_N)^2} v_1' W_N W_N v_1.$$

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 $\mathrm{r}(W_N)\sqrt{N} B_1$. Each element of $v_1' W_N W_N v_1$ is then at most $N\mathrm{r}(W_N)^2 B_1^2$, and the variance-covariance matrix thus goes to the zero matrix.

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

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

3. $$\frac{1}{\mathrm{s}(W_N)}(\tilde{\alpha}-\alpha_0)'v_1'W_Nv_1(\tilde{\alpha}-\alpha_0) = \sqrt{N}(\tilde{\alpha}-\alpha_0)'\frac{v_1'W_Nv_1}{N\mathrm{s}(W_N)}\sqrt{N}(\tilde{\alpha}-\alpha_0)$$

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

4. The fourth term has

$$\begin{aligned}|\frac{1}{\mathrm{s}(W_N)^2}(\tilde{\alpha}-\alpha_0)'v_1'W_N\tilde{v}_2|^2 &\leq \|\sqrt{N}(\tilde{\alpha}-\alpha_0)\|^2\|v_1'W_N\tilde{v}_2\|^2\frac{1}{N\mathrm{s}(W_N)^2} \\ &\leq \|\sqrt{N}(\tilde{\alpha}-\alpha_0)\|^2\frac{1}{N\mathrm{s}(W_N)^2}\,\ell N B_1^2\,\mathrm{r}(W_N)^2\frac{N\|\tilde{v}_2\|^2}{N} \\ &\xrightarrow{p} 0.\end{aligned}$$

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

QED.

Lemma 4 *Let $\{W_N\}$ be a sequence of symmetric matrices (with W_N being $N \times N$) such that $\mathrm{r}(W_N)/\mathrm{s}(W_N) \to 0$ as $N \to \infty$. Let u^N and $\tilde{u}^N$ be as in Proposition 1 for a given sequence $\{x_i\}$. Then,*

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

Proof of Lemma 4

We show this in two steps: showing first that

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

and then that

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

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

$$\begin{aligned}\sum_{ij} w_{ij}^2u_i^2u_j^2 - \mathrm{s}(\Sigma^N W_N\Sigma^N)^2 &= (z+v)'W_2(z+v) - vW_2v \\ &= z'W_2z + 2z'W_2v\end{aligned}$$

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 $z_1, \ldots, z_N$ are independent random variables with $E(z_i) = 0$ and $Var(z_i) < m$.[23] For any symmetric nonnegative matrix with zeros on the diagonal we then have $E(z'Az) = 0$ and

$$Var(z'Az) = Var(\sum_{ij} a_{ij} z_i z_j) = \sum_{ijk\ell} a_{ij} a_{k\ell} Cov(z_i z_j, z_k z_\ell) = 2 \sum_{ij} a_{ij}^2 Var(z_i z_j) \leq 2m^2 \mathrm{s}(\mathrm{A})^2.$$

Hence, $E(z'W_2 z) = 0$ and $Var(z'W_2 z) < 2m^2 \sum_{ij} w_{ij}^4$, and

$$\begin{aligned} Var\left(\frac{z'W_2 z}{\mathrm{s}(\Sigma^N W_N \Sigma^N)^2}\right) &< \frac{2m^2}{\underline{\sigma}^8} \frac{\sum_{ij} w_{ij}^4}{\left(\sum_{ij} w_{ij}^2\right)^2} \leq \frac{2m^2}{\underline{\sigma}^8} \frac{(\max_{ij} w_{ij}^2) \sum_{ij} w_{ij}^2}{\left(\sum_{ij} w_{ij}^2\right)^2} \\ &\leq \frac{2m^2}{\underline{\sigma}^8} \frac{\mathrm{r}(W_N)^2}{\mathrm{s}(W_N)^2} \to 0. \end{aligned}$$

Second, we similarly have $E(2v'W_2 z) = 0$ and

$$\begin{aligned} Var\left(\frac{2v'W_2 z}{\mathrm{s}(\Sigma^N W_N \Sigma^N)^2}\right) &\leq \frac{4}{\underline{\sigma}^8 \mathrm{s}(W_N)^4} Var\left(\sum_{ij} w_{ij}^2 v_i z_j\right) < \frac{4\overline{\sigma}^4 m}{\underline{\sigma}^8 \mathrm{s}(W_N)^4} \sum_j \left(\sum_i w_{ij}^2\right)^2 \\ &\leq \frac{4\overline{\sigma}^4 m}{\underline{\sigma}^8 \mathrm{s}(W_N)^4} \left(\max_j \sum_i w_{ij}^2\right) \sum_j \left(\sum_i w_{ij}^2\right) \\ &\leq \frac{4\overline{\sigma}^4 m}{\underline{\sigma}^8 \mathrm{s}(W_N)^4} \mathrm{r}(W_N)^2 \mathrm{s}(W_N)^2 \to 0. \end{aligned}$$

Turning now to the second main step, note that

$$\frac{\sum_{ij} w_{ij}^2 \tilde{u}_{iN}^2 \tilde{u}_{jN}^2 - \sum_{ij} w_{ij}^2 u_i^2 u_j^2}{\mathrm{s}(\Sigma^N W_N \Sigma^N)^2} = \frac{\tilde{u}^{2\prime} W_2 \tilde{u}^2 - u^{2\prime} W_2 u^2}{\mathrm{s}(\Sigma^N W_N \Sigma^N)^2} = \frac{2(\tilde{u}^2 - u^2)' W_2 u^2}{\mathrm{s}(\Sigma^N W_N \Sigma^N)^2} + \frac{(\tilde{u}^2 - u^2)' W_2 (\tilde{u}^2 - u^2)}{\mathrm{s}(\Sigma^N W_N \Sigma^N)^2},$$

where we've written $\tilde{u}^2$ for the vector with ith element $\tilde{u}_i^2$ and u^2 for the vector with ith 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) $\mathrm{r}(W_2)/\mathrm{s}(W_N)^2 \to 0$, and (iii) $\|W_2 u^2 / \mathrm{s}(W_N)^2\| \xrightarrow{p} 0$.

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.

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

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

[23] Here as in Lemma 3 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.

Finally, to derive result (iii) note that

$$\|W_2u^2/\mathrm{s}(W_N)^2\| \leq \|W_2v/\mathrm{s}(W_N)^2\| + \|W_2z/\mathrm{s}(W_N)^2\|,$$

where again we've written v for the vector with ith element $v_i = Var(u_i)$ 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.

$$\|\frac{W_2v}{\mathrm{s}(W_N)^2}\|^2 \leq \frac{\overline{\sigma}^4 \sum_i \left(\sum_j w_{ij}^2\right)^2}{\mathrm{s}(W_N)^4} \leq \frac{\overline{\sigma}^4 \mathrm{s}(W_N)^2 \mathrm{r}(W_N)^2}{\mathrm{s}(W_N)^4} \to 0,$$

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

$$\|W_2z/\mathrm{s}(W_N)^2\|^2 = \mathrm{s}(W_N)^{-4} z' W_2' W_2 z = \mathrm{s}(W_N)^{-4} z' B_N z + \sum_{i=1}^{N} c_{iN} z_i^2,$$

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

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

with the second to last conclusion following from the fact that for any symmetric matrix A, $\mathrm{s}(A'A)^2 = \sum_i \lambda_i^4 \leq (\sum_i \lambda_i^2)^2 = \mathrm{s}(A)^4$. Finally, to see that $\sum_{i=1}^N c_{iN} z_i^2 \xrightarrow{p} 0$ as $N \to \infty$ 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) $\max_i |c_{iN}| \to 0$ as $N \to \infty$; (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 $Prob(|e_i| \geq t) \leq Prob(|e| \geq t)$ 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, 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 $c_{iN} = \mathrm{s}(W_N)^{-4} \sum_j w_{ijN}^4$ is nonnegative and $\sum_{i=1}^N c_{iN} = \mathrm{s}(W_N)^{-4} \mathrm{s}(W_2)^2 \to 0$ as $N \to 0$. Applying the theorem gives that

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

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

QED.

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

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

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

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

$$\mathrm{E}\Big(\frac{\sum_k a_{kN}}{m(N)}\Big) \geq \frac{m(N)}{m(N)}\mathrm{inf}_k \mathrm{E}(a_{kN}) = \mathrm{inf}_k(1 - ((1-\nu_k)^N + N\nu_k(1-\nu_k)^{N-1})) \to 1,$$

where we have written ν_k for $\nu(P_{kN})$.

$$\begin{aligned}\mathrm{Var}\Big(\frac{\sum_k a_{kN}}{m(N)}\Big) &\leq \mathrm{sup}_k \mathrm{Var}(a_{kN}) \leq \mathrm{sup}_k \mathrm{E}[(a_{kN}-1)^2] \\ &= \mathrm{sup}_k((1-\nu_k)^N + N\nu_k(1-\nu_k)^{N-1}) \to 0.\end{aligned}$$

Therefore, $\mathrm{r}(W_N)/\mathrm{s}(W_N) = \mathrm{O}_p(1/\sqrt{m(N)})$, as desired.

QED.

Table 1: Null Distributions of Kernel Tests: 100 observations, 10,000 simulations

One Dimension

Window width	mean	standard deviation	95^{th} percentile	ACV size
h_{100}=0.05	0.00	0.98	1.68	5.4
h_{100}=0.1	0.00	0.94	1.70	5.4
h_{100}=0.2	0.00	0.86	1.61	4.7
h_{100}=0.3	0.00	0.77	1.48	3.8
h_{100}=0.4	0.01	0.68	1.34	2.9

Two Dimensions

Window width	mean	standard deviation	95^{th} percentile	ACV size
h_{100}=0.2	0.01	0.98	1.68	5.5
h_{100}=0.4	0.00	0.92	1.62	4.8
h_{100}=0.6	0.04	0.79	1.49	4.0
h_{100}=0.8	0.10	0.64	1.32	2.4

Three Dimensions

Window width	mean	standard deviation	95^{th} percentile	ACV size
h_{100}=0.4	0.06	1.00	1.75	6.1
h_{100}=0.6	0.03	0.94	1.67	5.2
h_{100}=0.8	0.08	0.84	1.58	4.3

Table 2: Power in One Dimension: $y_i = x_i + 0.5e(x_i) + \epsilon_i$, 1000 simulations

50 observations

Test		Alternative $e(x)$							
		$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
kernel	h_{100}=0.05	0.19	0.18	0.18	0.17	0.19	0.16	0.19	0.12
	h_{100}=0.1	0.26	0.26	0.24	0.20	0.19	0.18	0.26	0.05
	h_{100}=0.2	0.42	0.36	0.29	0.24	0.19	0.18	0.33	0.05
	h_{100}=0.3	0.49	0.36	0.27	0.20	0.14	0.10	0.34	0.05
	h_{100}=0.4	0.53	0.36	0.23	0.15	0.10	0.06	0.30	0.05
optimal		0.66	0.64	0.65	0.65	0.65	0.69	0.60	0.69

100 observations

Test		Alternative $e(x)$							
		$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
kernel	h_{100}=0.05	0.42	0.42	0.41	0.39	0.38	0.36	0.38	0.29
	h_{100}=0.1	0.60	0.59	0.54	0.50	0.47	0.43	0.51	0.13
	h_{100}=0.2	0.76	0.71	0.65	0.56	0.47	0.38	0.63	0.04
	h_{100}=0.3	0.82	0.73	0.61	0.49	0.36	0.23	0.65	0.06
	h_{100}=0.4	0.84	0.72	0.55	0.37	0.22	0.12	0.62	0.04
optimal		0.94	0.92	0.93	0.93	0.92	0.94	0.88	0.96

200 observations

Test		Alternative $e(x)$							
		$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
kernel	h_{100}=0.05	0.77	0.76	0.76	0.76	0.76	0.72	0.69	0.66
	h_{100}=0.1	0.92	0.89	0.91	0.87	0.86	0.81	0.85	0.47
	h_{100}=0.2	0.98	0.95	0.95	0.92	0.88	0.81	0.93	0.02
	h_{100}=0.3	0.99	0.97	0.95	0.89	0.81	0.68	0.95	0.06
	h_{100}=0.4	0.99	0.97	0.93	0.81	0.64	0.42	0.94	0.04
optimal		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3: Power in Higher Dimensions: $y_i = x_i \cdot \mathbf{1} + 0.5e(x_{i1}) + \epsilon_i$, 100 observations, 1000 simulations

One Dimension

Test		Alternative $e(x)$							
		$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
kernel	h_{100}=0.05	0.42	0.42	0.41	0.39	0.38	0.36	0.38	0.29
	h_{100}=0.1	0.60	0.59	0.54	0.50	0.47	0.43	0.51	0.13
	h_{100}=0.2	0.76	0.71	0.65	0.56	0.47	0.38	0.63	0.04
	h_{100}=0.3	0.82	0.73	0.61	0.49	0.36	0.23	0.65	0.06
	h_{100}=0.4	0.84	0.72	0.55	0.37	0.22	0.12	0.62	0.04
optimal		0.94	0.92	0.93	0.93	0.92	0.94	0.88	0.96

Two Dimensions

Test		Alternative $e(x)$							
		$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
kernel	h_{100}=0.2	0.26	0.21	0.21	0.16	0.16	0.15	0.20	0.05
	h_{100}=0.4	0.48	0.33	0.24	0.18	0.14	0.09	0.27	0.05
	h_{100}=0.6	0.60	0.33	0.17	0.11	0.07	0.05	0.20	0.05
	h_{100}=0.8	0.64	0.25	0.09	0.04	0.04	0.05	0.09	0.05
optimal		0.92	0.92	0.93	0.93	0.92	0.94	0.88	0.91

Three Dimensions

Test		Alternative $e(x)$							
		$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
kernel	h_{100}=0.4	0.20	0.15	0.13	0.10	0.08	0.07	0.14	0.04
	h_{100}=0.6	0.31	0.16	0.12	0.09	0.06	0.05	0.12	0.05
	h_{100}=0.8	0.38	0.15	0.10	0.06	0.06	0.04	0.08	0.05
optimal		0.93	0.92	0.93	0.92	0.93	0.91	0.89	0.92

Table 4: 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).

Table 5: Comparison of Power

	Rejection rates with 5% empirical critical values					
	Alternative 1		Alternative 2		Alternative 3	
Test statistic	$N = 100$	$N = 300$	$N = 100$	$N = 300$	$N = 100$	$N = 300$
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, 4, and 5 of Hong and White (1995).

Table 6: Power of Bin Tests $y_i = x_i \cdot \mathbf{1} + 0.5e(x_{i1}) + \epsilon_i$, 100 observations, 1000 simulations

One dimension

	Test	$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
bin	h_{100}=0.05	0.28	0.27	0.27	0.26	0.26	0.24	0.26	0.22
	h_{100}=0.1	0.44	0.43	0.38	0.38	0.32	0.29	0.37	0.33
	h_{100}=0.2	0.61	0.57	0.44	0.36	0.26	0.25	0.48	0.05
	h_{100}=0.3	0.67	0.57	0.43	0.30	0.23	0.26	0.47	0.07
	h_{100}=0.4	0.72	0.50	0.21	0.26	0.39	0.05	0.42	0.06

Two Dimensions

	Test	$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
bin	h_{100}=0.2	0.19	0.14	0.14	0.10	0.09	0.11	0.13	0.06
	h_{100}=0.4	0.32	0.19	0.10	0.11	0.15	0.06	0.15	0.06
	h_{100}=0.6	0.33	0.19	0.17	0.05	0.06	0.08	0.07	0.04
	h_{100}=0.8	0.41	0.15	0.09	0.08	0.06	0.05	0.11	0.06

Three Dimensions

	Test	$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$	$p_6(x)$	$p_7(x)$	$\sin(2\pi x)$	$\sin(10\pi x)$
bin	h_{100}=0.4	0.13	0.12	0.07	0.07	0.10	0.05	0.09	0.05
	h_{100}=0.6	0.15	0.13	0.11	0.06	0.07	0.05	0.06	0.04
	h_{100}=0.8	0.19	0.12	0.08	0.08	0.07	0.07	0.09	0.08