Journal of Econometrics 20 (1982) 105–134. North-Holland Publishing Company

CONSISTENT MODEL SPECIFICATION TESTS*

Herman J. BIERENS

Foundation for Economic Research of the University of Amsterdam, 1011 NH Amsterdam, The Netherlands

In this paper we propose two consistent tests for functional form of nonlinear regression models without employing specified alternative models. The null hypothesis is that the regression function equals the conditional expectation function, which is tested against the alternative hypothesis that the null is false. These tests are based on a Fourier transform characterization of conditional expectations.

1. Introduction

In the literature on model specification testing two trends can be distinguished. One trend consists of tests using one or more well-defined non-nested alternative specifications. See Cox (1961, 1962), Atkinson (1969, 1970), Quandt (1974), Pereira (1977, 1978), Pesaran and Deaton (1978), Davidson and MacKinnon (1981). Aguirre-Torres and Gallant (1983) present a distribution-free Cox test. These tests are, however, not real model specification tests but only model selection tests in the sense that they are only valid if either the null or one of the alternatives is true. See Bierens (1981b) for an example of the inconsistency of the P-test proposed by Davidson and MacKinnon (1981) in the case that all the hypotheses are false. The other trend consists of tests of the orthogonality condition, i.e., the condition that the expectation of the errors conditional on the regressors equals zero a.s., without employing a well-specified alternative hypothesis. Notable work on this problem has been done by Ramsey (1974, 1970), Hausman (1978) and White (1981). Although Ramsey's test (RESET) aims to be a pure orthogonality test, it is essentially a test of the hypothesis that the expectation of the errors conditional on the regression function is zero, i.e., if the specification is the standard linear model $y=X\beta+u$ then Ramsey's

*An earlier version of this paper was written while I was spending a three-month term at the University of Minnesota, whose kind hospitality is gratefully acknowledged. I am indebted to Christopher A. Sims and the participants of the econometrics workshops at the University of California at San Diego and Yale University for stimulating comments. Research in this paper was financed, in part, by a grant from the Netherlands Organisation for the Advancement of Pure Research (Z.W.O.).

RESET is not a test of the null hypothesis $\mathrm{E}(u \mid X)=0$ a.s., but only of the hypothesis $\mathrm{E}(u \mid X\beta_0)=0$ a.s., where β_0 is the probability limit of the OLS estimator $\hat{\beta}=(X'X)^{-1}X'y$. It is well-known that the latter hypothesis does not automatically imply the former. Hausman's test is based on the idea that for the true model the difference between an efficient estimator (under the null) and a consistent but non-efficient estimator, times the square root of the number of observations, will converge in distribution to the normal with zero mean, whereas in the case of misspecification the two estimators will have different probability limits. White's test is a variant of Hausman's test for the case of nonlinear regression models. The two estimators employed by White are the least-squares estimator as the efficient estimator (assuming normal errors) and the weighted least-squares estimator as the consistent but non-efficient estimator. The problem with this test is that the power and even the consistency heavily depends on the choice of the non-efficient estimator. (See the example in section 9.) So there is still need for alternative tests that are more watertight than those discussed in the literature. In this paper we will propose two consistent model specification tests, the first one being a simple but rather crude test, the second one a more sophisticated but also more laborious test. These tests are based on a Fourier transform characterization of conditional expectations.

2. The null hypothesis and its identification

Consider an i.i.d. stochastic process $(y_1,x_1),\ldots,(y_n,x_n),\ldots$ in $\boldsymbol{R}\times\boldsymbol{R}^k$, where the y_j satisfy

$$\mathrm{E}|y_j|<\infty. \tag{1}$$

The condition (1) is sufficient for the existence of a Borel measurable function $g(\cdot)$ on R^k such that

$$g(x_j)=\mathrm{E}(y_j \mid x_j) \quad \text{a.s.} \tag{2}$$

[see Chung (1974, theorem 9.1.2)]. Defining $u_j=y_j-g(x_j)$ we then get the following tautological regression model:

$$y_j=g(x_j)+u_j, \qquad j=1,2,\ldots,n,\ldots, \tag{3}$$

where obviously the u_j satisfy

$$\mathrm{E}(u_j \mid x_j)=0 \qquad \text{a.s.}, \qquad j=1,2,\ldots,n,\ldots \tag{4}$$

Suppose we have specified the regression function $g(x)$ as $f(x,\theta_0)$, where $f(x,\theta)$ is a known real-valued Borel measurable function on $\boldsymbol{R}^k \times \Theta$ and Θ is a parameter space containing the unknown parameter θ_0 if the specification is true. Ideally this specification is true if $g(x)=f(x,\theta_0)$ for all $x \in \boldsymbol{R}^k$ and some $\theta_0 \in \Theta$. However, this is too strong a condition, for the equality of $g(x)$ and $f(x,\theta_0)$ is only relevant for x in a set $S \subset \boldsymbol{R}^k$ such that $P(x_j \in S)=1$. For example, if the x_j's are uniform $[0,1]$ distributed random variables, then $g(x) \neq f(x,\theta_0)$ for $x \notin [0,1]$ does not matter as long as $g(x)=f(x,\theta_0)$ for $x \in [0,1]$. Thus testing the truth of the specification we have to test the null hypothesis

$$H_0 \colon P[f(x_j,\theta_0)=g(x_j)]=1 \quad \text{for some} \quad \theta_0 \in \Theta, \tag{5}$$

against the alternative hypothesis that H_0 is false, i.e.,

$$H_1 \colon P[f(x_j,\theta)=g(x_j)]<1 \quad \text{for all} \quad \theta \in \Theta. \tag{6}$$

But how can we identify H_0 versus H_1 from the distribution of the data, or in other words, how do H_0 and H_1, respectively, correspond to distinct characteristics of the distribution of (y_j, x_j)? The answer to this question is suggested by the following fundamental theorem.

Theorem 1. Let v be a real-valued random variable satisfying $\mathrm{E}|v|<\infty$ and let z be a random vector in $\boldsymbol{R}^k$. Then we have:

(I) $P[\mathrm{E}(v \mid z)=0]<1$ *if and only if* $\mathrm{E}(v\,\mathrm{e}^{\mathrm{i}t'z}) \neq 0$ *for some non-random vector* $t \in \boldsymbol{R}^k$.

(II) *If in addition z is bounded then $P[\mathrm{E}(v \mid z)=0]<1$ if and only if $\mathrm{E}(v\,\mathrm{e}^{\mathrm{i}t'z}) \neq 0$ for some non-random vector $t_0 \in \boldsymbol{R}^k$ in an arbitrarily small neighborhood of $t=0$.*

This theorem, together with Theorem 2 below, will be proved in the mathematical appendix.

Note that

$$\mathrm{E}(v\,\mathrm{e}^{\mathrm{i}t'z})=\int_{\boldsymbol{R}^k} \mathrm{e}^{\mathrm{i}t'z}\{\mathrm{E}(v \mid z)\}\,\mathrm{d}F(z),$$

where $F(z)$ is the distribution function of z, hence $\mathrm{E}(v\,\mathrm{e}^{\mathrm{i}t'z})$ is the Fourier transform of $\{\mathrm{E}(v \mid z)\}\,\mathrm{d}F(z)$. Now from part I of theorem 1 it follows that

$$\mathrm{E}[y_j - f(x_j, \theta_0) \mid x_j] = \mathrm{E}(u_j \mid x_j) + \mathrm{E}[g(x_j) - f(x_j, \theta_0) \mid x_j]$$
$$= E[g(x_j) - f(x_j, \theta_0) \mid x_j] = 0,$$

with probability 1 for some $\theta_0 \in \Theta$, if and only if

$$\mathrm{E}(y_j - f(x_j, \theta_0)) \mathrm{e}^{\mathrm{i}t'x_j} \equiv 0 \quad \text{for all} \quad t \in \boldsymbol{R}^k.$$

Hence:

Corollary 1. *H_0 is true if and only if for some $\theta_0 \in \Theta$,*

$$\mathrm{E}(y_j - f(x_j, \theta_0)) \mathrm{e}^{\mathrm{i}t'x_j} = 0 \quad \textit{for all} \quad t \in \boldsymbol{R}^k.$$

Also part II of Theorem 1 can be used for identifying the hypotheses, as we now show. Let Φ be a bounded Borel measurable mapping from $\boldsymbol{R}^k$ into $\boldsymbol{R}^k$ such that x_j and $\Phi(x_j)$ generate the same Euclidean Borel field, for example,

$$\Phi(x_j) = \begin{bmatrix} \operatorname{atan}(x_{1,j}) \\ \vdots \\ \operatorname{atan}(x_{k,j}) \end{bmatrix}. \tag{7}$$

Then

$$\mathrm{E}[g(x_j) - f(x_j, \theta) \mid x_j] = \mathrm{E}[g(x_j) - f(x_j, \theta) \mid \Phi(x_j)] \quad \text{a.s.} \tag{8}$$

Hence, applying part II of Theorem 1, we have:

Corollary 2. *Let Φ be any bounded Borel measurable mapping from $\boldsymbol{R}^k$ into $\boldsymbol{R}^k$ such that $\Phi(x_j)$ and x_j generate the same Euclidean Borel field. Then H_1 is true if and only if for every $\theta \in \Theta$ there is a t_0 in an arbitrarily small neighborhood of $t = 0$ such that*

$$\mathrm{E}(y_j - f(x_j, \theta)) \mathrm{e}^{\mathrm{i}t_0'\Phi(x_j)} \neq 0.$$

Of course, if H_0 is true then for some $\theta_0 \in \Theta$,

$$\mathrm{E}(y_j - f(x_j, \theta_0)) \mathrm{e}^{\mathrm{i}t'\Phi(x_j)} \equiv 0 \quad \text{for all} \quad t \in \boldsymbol{R}^k,$$

but it is sufficient to verify this only for an arbitrarily small neighborhood of $t = 0$.

A direct consequence of Corollary 2 is:

Corollary 3. *Let Φ be defined as in Corollary 2. Let*

$$N_0 = \mathop{\times}_{l=1}^{k} [-\varepsilon_l, \varepsilon_l] \quad \textit{where} \quad \varepsilon_l > 0, \quad l = 1, 2, \ldots, k,$$

are arbitrarily chosen. Put

$$\eta(\theta) = \int_{N_0} |\mathrm{E}(y_j - f(x_j, \theta))\mathrm{e}^{\mathrm{i}t'\Phi(x_j)}|^2 \, \mathrm{d}t. \tag{9}$$

Then H_0 is true if $\eta(\theta_0) = 0$ for some $\theta_0 \in \Theta$ and H_1 is true if

$$\inf_{\theta \in \Theta} \eta(\theta) > 0.$$

3. Model specification test 1

Corollary 3 suggests testing H_0 against H_1 by using a test statistic of the form

$$\hat{\eta} = \int_{N_0} \left| \frac{1}{n} \sum_{j=1}^{n} (y_j - f(x_j, \hat{\theta}))\, \mathrm{e}^{\mathrm{i}t'\Phi(x_j)} \right|^2 \mathrm{d}t, \tag{10}$$

where N_0 and Φ are as before and $\hat{\theta}$ is the nonlinear least-squares estimator of θ_0, i.e.,

$$\hat{\theta} \in \Theta \quad \text{a.s.,} \quad \sum_{j=1}^{n} (y_j - f(x_j, \hat{\theta}))^2 = \inf_{\theta \in \Theta} \sum_{j=1}^{n} (y_j - f(x_j, \theta))^2. \tag{11}$$

Note that if we put

$$z_j = \Phi(x_j), \qquad z_j' = (z_{1,j}, \ldots, z_{k,j}), \tag{12}$$

and

$$\hat{u}_j = y_j - f(x_j, \hat{\theta}), \tag{13}$$

$\hat{\eta}$ can be written as

$$\begin{aligned} \hat{\eta} &= \frac{1}{n^2} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \hat{u}_{j_1} \hat{u}_{j_2} \int_{N_0} \exp(\mathrm{i}t'(z_{j_1} - z_{j_2}))\, \mathrm{d}t \\ &= 2^k \frac{1}{n^2} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \hat{u}_{j_1} \hat{u}_{j_2} \prod_{l=1}^{k} \frac{\sin[\varepsilon_l(z_{l,j_1} - z_{l,j_2})]}{z_{l,j_1} - z_{l,j_2}}. \end{aligned} \tag{14}$$

In section 6 we show that if H_0 is true then under fairly general conditions $n\hat{\eta}$ converges in distribution to a non-negative random variable as $n\to\infty$, whereas $\text{plim}_{n\to\infty} n\hat{\eta}=\infty$ if H_0 is false. Thus $n\hat{\eta}$ may serve as a test statistic of a consistent model specification test. However, a disadvantage of this test is that the limiting distribution of $n\hat{\eta}$ under H_0 is of an unknown type. We shall propose two ways of getting around this problem. Our first approach is based on the fact that under H_0 the limiting distribution of $n\hat{\eta}$ has a first moment μ, say, which can be estimated consistently by

$$\hat{\mu}=\hat{\sigma}^2 2^k\left\{\prod_{l=1}^{k}\varepsilon_l-\operatorname{tr}(\hat{A}^{-1}\hat{B})\right\}, \tag{15}$$

where $\hat{\sigma}^2$ is the usual estimate of the variance of the u_j,

$$\hat{A}=\frac{1}{n}\sum_{j=1}^{n}\{(\partial/\partial\theta')f(x_j,\hat{\theta})\}\{(\partial/\partial\theta)f(x_j,\hat{\theta})\}, \tag{16}$$

and

$$\hat{B}=\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\{(\partial/\partial\theta')f(x_{j_1},\hat{\theta})\}\{(\partial/\partial\theta)f(x_{j_2},\hat{\theta})\}$$

$$\times\prod_{l=1}^{k}\frac{\sin[\varepsilon_l(z_{l,j_1}-z_{l,j_2})]}{z_{l,j_1}-z_{l,j_2}}. \tag{17}$$

By applying Chebishev's inequality for first moments we conclude that for any $\alpha\in(0,1)$,

$$\limsup_{n\to\infty} P\left[n\hat{\eta}>\frac{1}{\alpha}\hat{\mu}\right]\leqq\alpha \quad\text{if}\quad H_0 \text{ is true.} \tag{18}$$

Moreover, for every $\alpha>0$ we have

$$\lim_{n\to\infty} P\left[n\hat{\eta}>\frac{1}{\alpha}\hat{\mu}\right]=1 \quad\text{if}\quad H_1 \text{ is true,} \tag{19}$$

as also will be shown in section 6. Thus proceeding to test at the $(1-\alpha)\times 100$ percent confidence level we accept H_0 if $\hat{\mu}/n\hat{\eta}\geqq\alpha$, and we reject H_0 if not. However, this test will probably be rather crude in small samples, as it is based on Chebishev's inequality. This is not a very sharp inequality, but it has the advantage of computational simplicity.

The above test is only one example out of a larger class of tests that can be derived from Theorem 1. For example, we also may use a test statistic of the form

$$\hat{\eta}=\int_{R^k}\left|\frac{1}{n}\sum_{j=1}^{n}(y_j-f(x_j,\hat{\theta}))\mathrm{e}^{\mathrm{i}t'x_j}\right|^2 W(t)\,\mathrm{d}t,$$

where $W(t)$ is a positive weight function, for example a k-variate normal density, and for this test statistic we can, on basis of part I of Theorem 1, derive similar results as (18) and (19). Moreover, the same applies to test statistics of the form

$$\hat{\hat{\eta}}=\sum_{m_1=0}^{\infty}\sum_{m_2=0}^{\infty}\cdots\sum_{m_k=0}^{\infty}\left\{\frac{1}{n}\sum_{j=1}^{n}(y_j-f(x_j,\hat{\theta}))\prod_{l=1}^{k}z_{l,j}^{m_l}\right\}^2 W^*(m_1,\ldots,m_k),$$

where $W^*(m_1,\ldots,m_k)$ is a positive weight function such that the series converges (for example $W^*(m_1,\ldots,m_k)=\prod_{l=1}^{k}1/m_l!$), because of the following augmentation of part II of Theorem 1.

Theorem 2. Let the conditions of part II of Theorem 1 be satisfied and let $z'=(z_1,\ldots,z_k)$. Then

$$P[\mathrm{E}(v\,|\,z)=0]<1 \quad \textit{if and only if} \quad \mathrm{E}\left(v\prod_{l=1}^{k}z_l^{m_l}\right)\neq 0,$$

for some non-negative integers $m_1,\ldots,m_k$.

4. Model specification test 2

A second way of getting around the problem that the limiting distribution of $n\hat{\eta}$ under H_0 is unknown is the following. Define

$$\hat{R}_1(t\,|\,\beta_n)=\beta_n\hat{\eta}\frac{1}{n}\sum_{j=1}^{n}\hat{u}_j\{\cos(t'z_j)+\sin(t'z_j)\}-\tfrac{1}{2}t't, \tag{20}$$

$$\hat{R}_2(t\,|\,\beta_n)=-\beta_n\hat{\eta}\frac{1}{n}\sum_{j=1}^{n}\hat{u}_j\{\cos(t'z_j)+\sin(t'z_j)\}-\tfrac{1}{2}t't, \tag{21}$$

where $\hat{\eta}$, $\hat{\mu}_j$ and z_j are defined as before and (β_n) is a sequence of positive numbers satisfying

$$\beta_n=\mathrm{o}(\sqrt{n}),\qquad \beta_n\to\infty \quad \text{as} \quad n\to\infty. \tag{22}$$

If H_0 is true then

$$\operatorname{plim} \beta_n \hat{\eta} = 0,$$

and

$$\operatorname{plim} \sup_{t \in N_0} \left| \frac{1}{n} \sum_{j=1}^{n} \hat{u}_j (\cos(t'z_j) + \sin(t'z_j)) \right| = 0,$$

hence (20) and (21), converge in probability, uniformly on N_0, to $-\frac{1}{2}t't$. The term $-\frac{1}{2}t't$ in (20) and (21) therefore provides that the probability limits of $\hat{R}_1(t \mid \beta_n)$ and $\hat{R}_2(t \mid \beta_n)$ have a unique supremum at $t=0$. On the other hand, if H_0 is false then

$$\operatorname{plim} \beta_n \hat{\eta} = \infty,$$

and

$$\frac{1}{n} \sum_{j=1}^{n} \hat{u}_j (\cos(t'z_j) + \sin(t'z_j))$$

converge in probability, uniformly on N_0, to a function with extrema at t unequal to the zero vector, hence so do

$$\frac{1}{\beta_n \hat{\eta}} \hat{R}_1(t \mid \beta_n) \quad \text{and} \quad \frac{1}{\beta_n \hat{\eta}} \hat{R}_2(t \mid \beta_n).$$

Thus if we put

$$\hat{t}_l \in N_0, \quad \hat{R}_l(\hat{t}_l \mid \beta_n) = \sup_{t \in N_0} \hat{R}_l(t \mid \beta_n), \qquad l = 1, 2, \tag{23}$$

then

$$\operatorname{plim} \|\hat{t}_1 - \hat{t}_2\| = 0 \quad \text{under } H_0,$$

and

$$\operatorname{plim} \|\hat{t}_1 - \hat{t}_2\| > 0 \quad \text{under } H_1.$$

However, under H_0 we also have the stronger result

$$(\sqrt{n}(\hat{t}_1 - \hat{t}_2))/\beta_n \hat{\eta} \to N(0, 4\sigma^2 \Omega) \quad \text{in distr.,} \tag{24}$$

where σ^2 is the variance of u_j and Ω is a positive definite matrix.

Moreover, since under H_1 $\operatorname{plim}\hat{\eta}>0$, it follows from (22) that then $\operatorname{plim}\sqrt{n}/\hat{\eta}\beta_n=\infty$, and hence

$$\operatorname*{plim}_{n\to\infty}(\sqrt{n}\|\hat{t}_1-\hat{t}_2\|)/\beta_n\hat{\eta}=\infty \quad \text{if } H_1 \text{ is true.} \tag{25}$$

The matrix Ω is of the form

$$\Omega=D-CA^{-1}C', \tag{26}$$

where

$$A=\mathrm{E}\{(\partial/\partial\theta')f(x_j,\theta_0)\}\{(\partial/\partial\theta)f(x_j,\theta_0)\}, \tag{27}$$

$$C=\mathrm{E}z_j(\partial/\partial\theta)f(x_j,\theta_0), \tag{28}$$

$$D=\mathrm{E}z_jz_j'. \tag{29}$$

It can be shown that these matrices can be estimated consistently by $\hat{A}$ [defined by (16)],

$$\hat{C}=\frac{1}{n}\sum_{j=1}^{n}z_j(\partial/\partial\theta)f(x_j,\hat{\theta}), \tag{30}$$

and

$$\hat{D}=\frac{1}{n}\sum_{j=1}^{n}z_jz_j', \tag{31}$$

respectively. Denoting

$$\hat{\Omega}=\hat{D}-\hat{C}\hat{A}^{-1}\hat{C}', \tag{32}$$

we then have

$$\operatorname*{plim}_{n\to\infty}\hat{\Omega}=\Omega, \tag{33}$$

provided A is non-singular. Since also $\operatorname{plim}\hat{\sigma}^2=\sigma^2$, where $\hat{\sigma}^2$ is the usual estimate of the variance of the u_j's, we may conclude from (24) that under H_0

$$\tfrac{1}{2}\hat{\sigma}^{-1}\hat{\Omega}^{-\frac{1}{2}}[(\sqrt{n}(\hat{t}_1-\hat{t}_2))/\beta_n\hat{\eta}]\to\mathrm{N}(0,I) \quad \text{in distr.,} \tag{34}$$

provided Ω is non-singular. Denoting

$$\hat{T}=(n/4\beta_n^2\hat{\eta}^2\hat{\sigma}^2)(\hat{t}_1-\hat{t}_2)'\hat{\Omega}^{-1}(\hat{t}_1-\hat{t}_2), \tag{35}$$

we thus have

$$\hat{T} \to \chi^2(k) \quad \text{in distr.,} \quad \text{if } H_0 \text{ is true.} \tag{36}$$

Moreover, (25) implies that

$$\text{plim}\, \hat{T} = \infty \quad \text{if } H_1 \text{ is true.} \tag{37}$$

Thus using $\hat{T}$ as a test statistic and proceeding the test at the $(1-\alpha)\times 100$ percent confidence level, we accept H_0 if $\hat{T} \leqq T_\alpha$, and we reject H_0 if not, where T_α is such that $P(\chi^2(k) \leqq T_\alpha) = 1-\alpha$.

5. The assumptions and some preliminary results

In this section we shall set forth conditions such that (18), (19), (36) and (37) are true. These conditions are closely related to those for consistency and asymptotic normality of nonlinear least-squares estimators as can be found in Jennrich (1969) and Bierens (1981a, 1982). However, in this paper we shall limit our attention to the case of i.i.d. data, i.e.:

Assumption 1. The observations $(y_1, x_1), \ldots, (y_n, x_n)$ are i.i.d. random vectors in $\boldsymbol{R} \times \boldsymbol{R}^k$, where the y_j satisfy $\mathrm{E}y_j^2 < \infty$.

The i.i.d. assumption is, of course, rather restrictive, though not more restrictive than the 'constant in repeated samples' assumption made by Pesaran and Deaton (1978) and Davidson and MacKinnon (1981). The main reason for this i.i.d. assumption is that it avoids complications in identifying the null hypothesis. For example, if we retain the independence assumption but allow for non-stationarity, then the null hypothesis (5) becomes a sequence $(H_{0,j})$ of null hypotheses,

$$H_{0,j}\colon P[\mathrm{E}(y_j \mid x_j) = f(x_j, \theta_0)] = 1 \quad \text{for some} \quad \theta_0 \in \Theta,$$

and for testing each of these hypotheses we have only one observation x_j available. However, if the distribution functions F_j, say, of the (y_j, x_j) satisfy

$$\lim \frac{1}{n} \sum_{j=1}^{n} F_j(y, x) = F(y, x),$$

for any continuity point of F, where F is a distribution function, and if we denote $\bar{g}$ the conditional expectation function implied by F, then we can test the null hypothesis

$$H_0\colon \bar{g}(x) = f(x, \theta_0) \quad \text{a.s.,} \quad \text{w.r.t. } F, \quad \text{for some} \quad \theta_0 \in \Theta,$$

against the alternative hypothesis that this null is false. In the case of independent data this generalisation seems not too hard, using the results of Bierens (1981a, ch. 3) and White (1980). Moreover, in view of the results of Bierens (1981a, ch 5; 1982) and Domowitz and White (1982) it seems not impossible to extend our tests to the case of non-stationary dependent data, but this extension would be far from trivial and would require a discussion which would substantially increase the length of this paper. We have already done some preliminary work in this direction. See Bierens (1981c).

The condition $\mathrm{E}y_j^2<\infty$ is stronger than (1) but needed in order to assure that the errors u_j, defined by

$$u_j=y_j-\mathrm{E}(y_j\,|\,x_j)=y_j-g(x_j),$$

have finite variance. Thus Assumption 1 implies

$$\mathrm{E}(u_j\,|\,x_j)=0 \quad \text{a.s.}, \qquad \mathrm{E}u_j^2=\sigma^2<\infty. \tag{38}$$

However, Assumption 1 does not imply that $\mathrm{E}(u_j^2\,|\,x_j)=\sigma^2$ a.s., and therefore we assume in addition that:

Assumption 2. u_j and x_j are mutually independent.

In section 8 we show what happens if we drop this assumption.

We shall limit our attention to specific $f(x,\theta)$ satisfying the following conditions:

Assumption 3. The function $f(x,\theta)$ and its first and second partial derivatives $(\partial/\partial\theta_i)f(x,\theta)$ and $(\partial/\partial\theta_{i_1})(\partial/\partial\theta_{i_2})f(x,\theta)$, $i,i_1,i_2=1,2,\ldots,m$, are for each $x\in\boldsymbol{R}^k$ continuous real functions on $\Theta\subset\boldsymbol{R}^m$ and for each $\theta\in\Theta$ Borel measureable real functions on $\boldsymbol{R}^k$. Moreover, the set Θ is convex and compact.

Assumption 4. Let

$$\mathrm{E}\sup_{\theta\in\Theta}|f(x_j,\theta)|^2<\infty,$$

$$\mathrm{E}\sup_{\theta\in\Theta}|(\partial/\partial\theta_i)f(x_j,\theta)|^2<\infty,$$

$$\mathrm{E}\sup_{\theta\in\Theta}|(\partial/\partial\theta_{i_1})(\partial/\partial\theta_{i_2})f(x_j,\theta)|<\infty, \qquad i,i_1,i_2=1,2,\ldots,m.$$

In order to estimate the true parameter θ_0 consistently by least-squares estimation if H_0 is true it should be unique in the sense that it is the only point in Θ such that

$$0=\mathrm{E}\{g(x_j)-f(x_j,\theta_0)\}^2=\inf_{\theta\in\Theta}\mathrm{E}\{g(x_j)-f(x_j,\theta)\}^2. \tag{39}$$

But even if H_0 is false we can define a point θ_0 in Θ by the right-hand-side equality in (39), and if such a point is unique it can be estimated consistently by least squares. Therefore we assume:

Assumption 5. There exists a unique point $\theta_0\in\Theta$ such that

$$\mathrm{E}\{g(x_j)-f(x_j,\theta_0)\}^2=\inf_{\theta\in\Theta}\mathrm{E}\{g(x_j)-f(x_j,\theta_0)\}^2.$$

Now the Assumptions 1 through 4, together with Theorem 2 of Jennrich (1969), imply that

$$\sup_{\theta\in\Theta}|\hat{Q}(\theta)-Q(\theta)|\to 0 \quad \text{a.s.}, \tag{40}$$

where

$$\hat{Q}(\theta)=\frac{1}{n}\sum_{j=1}^{n}\{y_j-f(x_j,\theta)\}^2, \tag{41}$$

and

$$Q(\theta)=\sigma^2+\mathrm{E}\{g(x_j)-f(x_j,\theta)\}^2. \tag{42}$$

Assumption 3 implies that $Q(\theta)$ is a continuous function on Θ and Assumption 5 implies that $Q(\theta)$ takes a unique infimum on Θ at θ_0. Comparing these results with the conditions of Lemma 3.1.3 of Bierens (1981a) we then see that the least-squares estimator $\hat{\theta}$ is strongly consistent, i.e.:

Theorem 3. Let the Assumptions 1 through 4 be satisfied. Then $\hat{\theta}\to\theta_0$ a.s. under H_0 as well as under H_1.

Note that this result is essentially Corollary 2.2. of White (1981).

Next suppose:

Assumption 6. The matrix $A=\mathrm{E}\{(\partial/\partial\theta')f(x_j,\theta_0)\}\{(\partial/\partial\theta)f(x_j,\theta_0)\}$ is positive definite.

Assumption 7. If H_0 is true then θ_0 is an interior point of Θ.

From the argument in Section 3.1.3 of Bierens (1981a) it then follows that:

Theorem 4. *If H_0 is true and if the Assumptions 1 through 7 hold, then*

$$\operatorname*{plim}_{n\to\infty}\left\|\sqrt{n}(\hat{\theta}-\theta_0)-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}u_jA^{-1}(\partial/\partial\theta')f(x_j,\theta_0)\right\|=0,$$

and consequently by the central limit theorem,

$$\sqrt{n}(\hat{\theta}-\theta_0)\to \mathrm{N}(0,\sigma^2A^{-1})\quad \textit{in distr.}$$

Finally, in the case of test 2 we shall also need the following assumption:

Assumption 8. The matrix Ω defined by (26) is non-singular.

6. Asymptotic theory of test 1

In this section we show that (18) and (19) are true if the Assumptions 1 through 7 hold. Let us first introduce some notation:

$$\hat{\xi}(t)=\frac{1}{n}\sum_{j=1}^{n}(y_j-f(x_j,\hat{\theta}))\,\mathrm{e}^{\mathrm{i}t'z_j},\tag{43}$$

$$\hat{\xi}^{(2)}(t)=\frac{1}{n}\sum_{j=1}^{n}\{u_j+(\theta_0-\hat{\theta})'(\partial/\partial\theta')f(x_j,\theta_0)\}\,\mathrm{e}^{\mathrm{i}t'z_j},\tag{44}$$

$$\hat{\xi}^{(3)}(t)=\frac{1}{n}\sum_{j=1}^{n}u_j\{\mathrm{e}^{\mathrm{i}t'z_j}-(\partial/\partial\theta)f(x_j,\theta_0)A^{-1}\hat{b}_0(t)\},\tag{45}$$

$$\hat{\xi}^{*}(t)=\frac{1}{n}\sum_{j=1}^{n}u_j\{\mathrm{e}^{\mathrm{i}t'z_j}-(\partial/\partial\theta)f(x_j,\theta_0)A^{-1}b_0(t)\}=\frac{1}{n}\sum_{j=1}^{n}u_j\rho_j(t),\tag{46}$$

say, where z_j is defined by (12),

$$\rho_j(t)=\mathrm{e}^{\mathrm{i}t'z_j}-(\partial/\partial\theta)f(x_j,\theta_0)A^{-1}b_0(t),\tag{47}$$

$$\hat{b}_0(t)=\frac{1}{n}\sum_{j=1}^{n}(\partial/\partial\theta')f(x_j,\theta_0)\,\mathrm{e}^{\mathrm{i}t'z_j},\tag{48}$$

$$b_0(t)=\mathrm{E}\hat{b}_0(t)=\mathrm{E}(\partial/\partial\theta')f(x_j,\theta_0)\,\mathrm{e}^{\mathrm{i}t'z_j}.\tag{49}$$

Note that from (10) and (43),

$$\hat{\eta}=\int_{N_0}|\hat{\xi}(t)|^2\,\mathrm{d}t. \tag{50}$$

Now put

$$\hat{\eta}^*=\int_{N_0}|\hat{\xi}^*(t)|^2\,\mathrm{d}t. \tag{51}$$

We then have:

Theorem 5. If H_0 is true and if the Assumptions 1 through 7 hold, then

$$\operatorname*{plim}_{n\to\infty}|n\hat{\eta}-n\hat{\eta}^*|=0.$$

Proof. Using the mean value theorem we can write

$$\begin{aligned}\operatorname{Re}\hat{\xi}(t)&=\frac{1}{n}\sum_{j=1}^{n}[u_j+f(x_j,\theta_0)-f(x_j,\hat{\theta})]\cos(t'z_j)\\&=\frac{1}{n}\sum_{j=1}^{n}[u_j-(\hat{\theta}-\theta_0)'(\partial/\partial\theta')f(x_j,\tilde{\theta}_1(t))]\cos(t'z_j),\end{aligned} \tag{52}$$

and similarly we have

$$\operatorname{Im}\hat{\xi}(t)=\frac{1}{n}\sum_{j=1}^{n}[u_j-(\hat{\theta}-\theta_0)'(\partial/\partial\theta')f(x_j,\tilde{\theta}_2(t))]\sin(t'z_j), \tag{53}$$

where $\tilde{\theta}_1(t)$ and $\tilde{\theta}_2(t)$ are mean values satisfying

$$\|\tilde{\theta}_l(t)-\theta_0\|\leqq\|\hat{\theta}-\theta_0\|\ \text{a.s.,}\quad\text{for all}\quad t\in\boldsymbol{R}^k,\quad l=1,2. \tag{54}$$

Using Theorem 2 of Jennrich (1969) it is not hard to show

$$\operatorname*{plim}_{n\to\infty}\sup_{t\in N_0}|\sqrt{n}\hat{\xi}(t)-\sqrt{n}\hat{\xi}^{(2)}(t)|=0, \tag{55}$$

and using Theorem 4 we see that

$$\operatorname*{plim}_{n\to\infty}\sup_{t\in N_0}|\sqrt{n}\hat{\xi}^{(2)}(t)-\sqrt{n}\hat{\xi}^{(3)}(t)|=0. \tag{56}$$

Again using Theorem 2 of Jennrich (1969) we may conclude that

$$\operatorname*{plim}_{n\to\infty} \sup_{t\in N_0} |\hat{b}_0(t)-b_0(t)|=0, \tag{57}$$

and consequently that

$$\operatorname*{plim}_{n\to\infty} \sup_{t\in N_0} |\sqrt{n}\hat{\xi}^{(3)}(t)-\sqrt{n}\hat{\xi}^*(t)|=0. \tag{58}$$

Combining (55), (56) and (58) we obtain

$$\operatorname*{plim}_{n\to\infty} \sup_{t\in N_0} |\sqrt{n}\hat{\xi}(t)-\sqrt{n}\hat{\xi}^*(t)|=0, \tag{59}$$

and applying Lemma 3.3.1 of Bierens (1981a) we conclude from (59) that also

$$\operatorname*{plim}_{n\to\infty} \sup_{t\in N_0} \left| |\sqrt{n}\hat{\xi}(t)|^2-|\sqrt{n}\hat{\xi}^*(t)|^2\right|=0. \tag{60}$$

This proves the theorem. Q.E.D.

Theorem 5 implies that, under $H_0, n\hat{\eta}$ and $n\hat{\eta}^*$ have the same limiting distribution. But what is the limiting distribution of $n\hat{\eta}^*$? If we substitute (46) in (51) we get

$$n\hat{\eta}^*=\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} u_{j_1}u_{j_2}\int_{N_0} \rho_{j_1}(t)\overline{\rho_{j_2}(t)}\,\mathrm{d}t, \tag{61}$$

where $\bar{\rho}_j$ is the complex conjugate of ρ_j. If we could write

$$\int_{N_0} \rho_{j_1}(t)\overline{\rho_{j_2}(t)}\,\mathrm{d}t$$

as a product of i.i.d. random variables r_{j_1} and r_{j_2}, say, then $n\hat{\eta}^*$ would converge in distribution to χ_1^2 times $\sigma^2\mathrm{E}r_j^2$, but it appears impossible to split up the integral involved in this way. So the limiting distribution of $n\hat{\eta}^*$ is probably of an unknown type.

A possible way out of this problem is to compute the expectation of $n\hat{\eta}^*$ and to apply Chebishev's inequality. This expectation is

$$\mu=\mathrm{E}\,n\hat{\eta}^*=\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\,u_j^2\,\mathrm{E}\int_{N_0}\rho_j(t)\bar{\rho}_j(t)\,\mathrm{d}t=\sigma^2\mathrm{E}\int_{N_0}\rho_j(t)\bar{\rho}_j(t)\,\mathrm{d}t, \tag{62}$$

which is obviously independent of the sample size n. We then have by Chebishev's inequality

$$P\left[n\hat{\eta}^* > \frac{1}{\alpha}\mu\right] \leqq \mathrm{E}n\hat{\eta}^* \Big/ \frac{1}{\alpha}\mu = \alpha, \tag{63}$$

and consequently by Theorem 5,

$$\limsup_{n\to\infty} P\left[n\hat{\eta} > \frac{1}{\alpha}\mu\right] \leqq \alpha. \tag{64}$$

So if we can find a consistent estimator $\hat{\mu}$, say, of μ then

$$\limsup_{n\to\infty} P\left[n\hat{\eta} > \frac{1}{\alpha}\hat{\mu}\right] \leqq \alpha. \tag{65}$$

We can construct such an estimator $\hat{\mu}$ as follows. Put

$$\hat{b}(t) = \frac{1}{n}\sum_{j=1}^{n} (\partial/\partial\theta') f(x_j, \hat{\theta})\, \mathrm{e}^{\mathrm{i}t'z_j}, \tag{66}$$

and

$$\hat{\rho}_j(t) = \mathrm{e}^{\mathrm{i}t'z_j} - (\partial/\partial\theta) f(x_j, \hat{\theta}) \hat{A}^{-1} \hat{b}(t), \tag{67}$$

where $\hat{A}$ is defined by (16). Then

$$\sup_{t\in N_0} \left| \frac{1}{n}\sum_{j=1}^{n} \hat{\rho}_j(t)\overline{\hat{\rho}_j(t)} - \frac{1}{n}\sum_{j=1}^{n} \rho_j(t)\overline{\rho_j(t)} \right|$$

$$\leqq \sup_{t\in N_0} \left|\hat{b}(t)'\hat{A}^{-1}\hat{b}(t) - b_0(t)'\hat{A}^{-1}b_0(t)\right| \to 0 \quad \text{a.s.}, \tag{68}$$

because by Theorem 2 of Jennrich (1969)

$$\hat{A} \to A \quad \text{a.s.}, \tag{69}$$

where A is defined as in Assumption 6, and

$$\sup_{t\in N_0} |\hat{b}(t) - b_0(t)| \to 0 \quad \text{a.s.} \tag{70}$$

Moreover,

$$\sup_{t\in N_0} \left| \frac{1}{n}\sum_{j=1}^{n} \rho_j(t)\overline{\rho_j(t)} - \mathrm{E}\rho_j(t)\overline{\rho_j(t)} \right| \to 0 \quad \text{a.s.} \tag{71}$$

So if we combine (68) and (71) we then may conclude

$$\int_{N_0} \frac{1}{n} \sum_{j=1}^{n} \hat{\rho}_t(t)\overline{\hat{\rho}_j(t)}\,\mathrm{d}t \rightarrow \int_{N_0} \mathrm{E}\{\rho_j(t)\overline{\rho_j(t)}\}\,\mathrm{d}t = \mathrm{E}\int_{N_0} \rho_j(t)\overline{\rho_j(t)}\,\mathrm{d}t \text{ a.s.} \tag{72}$$

Furthermore, it is not hard to prove that

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n} [y_j - f(x_j, \hat{\theta})]^2 \rightarrow \sigma^2 \text{ a.s.,} \quad \text{if } H_0 \text{ is true.} \tag{73}$$

Thus the statistic $\hat{\mu}$, defined by

$$\hat{\mu} = \hat{\sigma}^2 \frac{1}{n}\sum_{j=1}^{n} \int_{N_0} \hat{\rho}_j(t)\overline{\hat{\rho}_j(t)}\,\mathrm{d}t, \tag{74}$$

is a strongly consistent estimator of μ. We leave it to the reader to verify that this estimator $\hat{\mu}$ can be written as (15). This proves the following result:

Theorem 6. If H_0 is true and the Assumptions 1 through 7 hold, then

$$\limsup_{n\rightarrow\infty} P\left(n\hat{\eta} > \frac{1}{\alpha}\hat{\mu}\right) \leqq \alpha \quad \textit{for every} \quad \alpha \in (0,1).$$

Now suppose that H_0 is false but that the Assumptions 1 through 6 still hold. Using Theorem 3 and Theorem 2 of Jennrich (1969) it is not hard to show

$$\hat{\eta} \rightarrow \eta_0 \quad \text{a.s.,} \tag{75}$$

where

$$\eta_0 = \int_{N_0} |\mathrm{E}(g(x_j) - f(x_j, \theta_0))\,\mathrm{e}^{it'z_j}|^2\,\mathrm{d}t. \tag{76}$$

But Corollary 3 implies that $\eta_0 > 0$, hence

$$n\hat{\eta} \rightarrow \infty \quad \text{a.s.} \tag{77}$$

Moreover, it is not hard to show that now

$$\hat{\sigma}^2 \rightarrow \sigma_0^2 = \sigma^2 + \mathrm{E}\{g(x_j) - f(x_j, \theta_0)\}^2 \quad \text{a.s.,} \tag{78}$$

and that (72) carries over. Thus we have

$$\hat{\mu} \to \mu_0 \quad \text{a.s.,} \tag{79}$$

where

$$\mu_0 = \sigma_0^2 \int_{N_0} \mathrm{E}\rho_j(t)\overline{\rho_j(t)}\,\mathrm{d}t < \infty. \tag{80}$$

This proves:

Theorem 7. If H_1 is true and the Assumptions 1 through 6 hold, then

$$\lim_{n\to\infty} P\left(n\hat{\eta} > \frac{1}{\alpha}\hat{\mu}\right) = 1 \quad \textit{for every} \quad \alpha \in (0,1).$$

7. Asymptotic theory of test 2

It is easily verified from (20) and (21) that if H_0 is true and if the Assumptions 1 through 7 are satisfied, then

$$\operatorname*{plim}_{n\to\infty} \sup_{t\in N_0} \left|\hat{R}_l(t\,|\,\beta_n) - \{-\tfrac{1}{2}t't\}\right| = 0, \qquad l = 1, 2, \tag{81}$$

for $\operatorname{plim}\beta_n\hat{\eta} = 0$. Applying Lemma 3.1.8 of Bierens (1981a) it then follows from (81) and (23) that

$$\operatorname*{plim}_{n\to\infty} \hat{t}_l = 0, \qquad l = 1, 2. \tag{82}$$

On the other hand, if H_0 is false then by Theorem 2 of Jennrich (1969), Theorem 3, condition (22) and the fact that $\hat{\eta} \to \eta_0 > 0$ a.s., we have

$$\sup_{t\in N_0} \left|\frac{1}{\beta_n}\hat{R}_1(t\,|\,\beta_n) - s(t)\right| \to 0 \quad \text{a.s.,} \tag{83}$$

$$\sup_{t\in N_0} \left|\frac{1}{\beta_n}\hat{R}_2(t\,|\,\beta_n) + s(t)\right| \to 0 \quad \text{a.s.,} \tag{84}$$

where

$$\begin{aligned} s(t) &= \eta_0 \mathrm{E}(g(x_j) - f(x_j, \theta_0))(\cos(t'z_j) + \sin(t'z_j)) \\ &= \eta_0 \operatorname{Re}\{\mathrm{E}(g(x_j) - f(x_j, \theta_0))\mathrm{e}^{\mathrm{i}t'z_j}\} \\ &\quad + \eta_0 \operatorname{Im}\{\mathrm{E}(g(x_j) - f(x_j, \theta_0))\,\mathrm{e}^{\mathrm{i}t'z_j}\}. \end{aligned} \tag{85}$$

Now (83), (84) and (23) imply that

$$s(\hat{t}_1) \to \sup_{t \in N_0} s(t) \quad \text{a.s.,} \tag{86}$$

$$s(\hat{t}_2) \to \inf_{t \in N_0} s(t) \quad \text{a.s.} \tag{87}$$

But Corollary 3 implies that $s(t_*) \neq 0$ for some $t_* \in N_0$. Hence

$$\{s(\hat{t}_1) - s(\hat{t}_2)\} \to \sup_{t \in N_0} s(t) - \inf_{t \in N_0} s(t) > 0 \quad \text{a.s.,} \tag{88}$$

and consequently by the uniform continuity of $s(t)$ on N_0,

$$\|\hat{t}_1 - \hat{t}_2\| > \delta \text{ a.s.,} \quad \text{for some} \quad \delta > 0 \quad \text{as} \quad n \to \infty. \tag{89}$$

This proves (25).

Now let us return to the case that H_0 is true. We shall prove (24) by showing first that

$$\frac{\sqrt{n}}{\beta_n \hat{\eta}} (\hat{t}_1 - \hat{t}_2) \quad \text{and} \quad 2 \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{u}_j z_j$$

converge in distribution to the same limiting distribution, and second that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{u}_j z_j \to \mathrm{N}(0, \sigma^2 \Omega) \quad \text{in distr.}$$

Thus consider the following Taylor expansion of $\sqrt{n}(\partial/\partial t)\hat{R}_1(\hat{t}_1 | \beta_n)$ around $t = 0$:

$$\begin{aligned} \sqrt{n}(\partial/\partial t)\hat{R}_1(\hat{t}_1 \mid \beta_n) &= \sqrt{n}(\partial/\partial t)\hat{R}_1(0 \mid \beta_n) + \sqrt{n}\hat{t}_1'(\partial/\partial t)(\partial/\partial t')\hat{R}_1(\hat{t}_1^* \mid \beta_n) \\ &= \beta_n \hat{\eta} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{u}_j z_j \qquad (90) \\ &\quad - \sqrt{n}\hat{t}_1' \left\{ I + \beta_n \hat{\eta} \frac{1}{n} \sum_{j=1}^{n} \hat{u}_j z_j z_j' (\cos(\hat{t}_1^{*\prime} z_j) + \sin(\hat{t}_1^{*\prime} z_j)) \right\}, \end{aligned}$$

where $\hat{t}_1^*$ is a mean value satisfying

$$\|\hat{t}_1^*\| \leqq \|\hat{t}_1\|. \tag{91}$$

Since plim $\hat{t}_1 = 0$ under H_0 it follows that the probability that $\hat{t}_1$ is an interior point of N_0 converges to 1, and consequently that

$$\lim_{n\to\infty} P[(\partial/\partial t')\hat{R}_1(\hat{t}_1 \mid \beta_n) = 0] = 1. \tag{92}$$

In its turn (92) implies

$$\lim_{n\to\infty} P\left[\frac{\sqrt{n}}{\beta_n \hat{\eta}}(\partial/\partial t')\hat{R}_1(\hat{t}_1 \mid \beta_n) = 0\right] = 1. \tag{93}$$

Combining (90) and (93) we then get

$$\lim_{n\to\infty} P\left[\frac{\sqrt{n}t_1}{\beta_n \hat{\eta}} - (I + \hat{M}_1)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \hat{u}_j z_j = 0\right] = 1, \tag{94}$$

where

$$\hat{M}_1 = \beta_n \hat{\eta}\frac{1}{n}\sum_{j=1}^{n} \hat{u}_j z_j z_j'(\cos(\hat{t}_1^{*\prime} z_j) + \sin(\hat{t}_1^{*\prime} z_j)). \tag{95}$$

In a similar way we get

$$\lim_{n\to\infty} P\left[\frac{\sqrt{n}\hat{t}_2}{\beta_n \hat{\eta}} + (I - \hat{M}_2)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \hat{u}_j z_j = 0\right] = 1, \tag{96}$$

where

$$\hat{M}_2 = \beta_n \hat{\eta}\frac{1}{n}\sum_{j=1}^{n} \hat{u}_j z_j z_j'(\cos(\hat{t}_2^{*\prime} z_j) + \sin(\hat{t}_2^{*\prime} z_j)), \tag{97}$$

with $\hat{t}_2^*$ some vector in N_0 satisfying

$$\|\hat{t}_2^*\| \leqq \|\hat{t}_2\|. \tag{98}$$

Realizing that under H_0

$$\operatorname*{plim}_{n\to\infty} \hat{M}_1 = O, \qquad \operatorname*{plim}_{n\to\infty} \hat{M}_2 = O, \tag{99}$$

we now conclude from (94) and (95) that

$$\operatorname*{plim}_{n\to\infty}\left\|\frac{\sqrt{n}(\hat{t}_1-\hat{t}_2)}{\beta_n\hat{\eta}}-2\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{u}_jz_j\right\|=0, \tag{100}$$

provided

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{u}_jz_j \quad \text{converges in distr.}$$

For proving the latter, we observe that by the mean value theorem

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{u}_jz_j&=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}u_jz_j+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(f(x_j,\theta_0)-f(x_j,\hat{\theta}))z_j\\&=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}u_jz_j-\frac{1}{n}\sum_{j=1}^{n}z_j(\partial/\partial\theta)f(x_j,\hat{\theta}_*)\sqrt{n}(\hat{\theta}-\theta_0),\end{aligned} \tag{101}$$

where $\hat{\theta}_*$ is a mean value satisfying $\|\hat{\theta}_*-\theta_0\|\leqq\|\hat{\theta}-\theta_0\|$. Thus $\hat{\theta}_*$ is also a consistent estimator of θ_0, and hence

$$\frac{1}{n}\sum_{j=1}^{n}z_j(\partial/\partial\theta)f(x_j,\hat{\theta}_*)\to C \quad \text{a.s.}, \tag{102}$$

where C is defined by (28). Now from (101), (102) and Theorem 4 we obtain

$$\operatorname{plim}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{u}_jz_j-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}u_j\{z_j-CA^{-1}(\partial/\partial\theta')f(x_j,\theta_0)\}\right|=0. \tag{103}$$

Since from the central limit theorem

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}u_j\{z_j-CA^{-1}(\partial/\partial\theta')f(x_j,\theta_0)\}\to \mathrm{N}(0,\sigma^2\Omega) \quad \text{in distr.}, \tag{104}$$

where Ω is defined by (26), we conclude that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{u}_jz_j\to \mathrm{N}(0,\sigma^2\Omega) \quad \text{in distr.} \tag{105}$$

Combining this result with (100) we obtain the desired result (24). Since Theorem 3 in this paper and Theorem 2 of Jennrich (1969) imply that (33) and even the stronger result $\hat{\Omega} \to \Omega$ a.s. hold under H_0 as well as under H_1, it is now easy to complete the proof of Theorem 8 below.

Theorem 8. Let the Assumptions 1 through 8 be satisfied and let $\hat{T}$ be defined by (35), where the sequence (β_n) satisfies condition (22). Then $\hat{T} \to \chi^2(k)$ in distr. if H_0 is true, and $\hat{T} \to \infty$ a.s. if H_1 is true. In the latter case the rate of a.s. convergence to infinity is n/β_n^2.

8. The consequence of dropping Assumption 2

If we drop Assumption 2, we then no longer have that $\mathrm{E}(u_j^2 \mid x_j) = \mathrm{E}u_j^2 = \sigma^2$ a.s., but only that

$$\mathrm{E}(u_j^2 \mid x_j) = \sigma_*^2(x_j) \quad \text{a.s.,} \tag{106}$$

where σ_*^2 is a Borel measurable real function on $\boldsymbol{R}^k$. However, Assumption 2 is not essential for the validity of Theorem 3 and Theorem 4, as is not hard to show. Thus:

Theorem 9. If the Assumptions 1 and 3 through 5 are satisfied then $\hat{\theta} \to \theta_0$ a.s. under H_0 as well as under H_1. If in addition the Assumptions 6 and 7 are satisfied and if H_0 is true, then

$$\operatorname*{plim}_{n\to\infty} \left\| \sqrt{n}(\hat{\theta} - \theta_0) - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} u_j A^{-1} (\partial/\partial\theta') f(x_j, \theta_0) \right\| = 0.$$

Moreover, if we augment Assumption 4 with:

Assumption 9. $\mathrm{E}\sigma^2(x_j) \sup_{\theta \in \Theta} |(\partial/\partial\theta_i) f(x_j, \theta)|^2 < \infty$,

and if we put

$$A_2 = \mathrm{E}\sigma_*^2(x_j)\{(\partial/\partial\theta') f(x_j, \theta_0)\}\{(\partial/\partial\theta) f(x_j, \theta_0)\}, \tag{107}$$

then similar to White (1981):

Theorem 10. If H_0 is true and if the Assumptions 1, 3 through 7 and 9 are satisfied then

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathrm{N}(0, A^{-1} A_2 A^{-1}) \quad \textit{in distr.}$$

Furthermore, observe that under the conditions of Theorem 10 the matrix

$$\hat{A}_2 = \frac{1}{n}\sum_{j=1}^{n} \hat{u}_j^2 \{(\partial/\partial\theta')f(x_j,\hat{\theta})\}\{(\partial/\partial\theta)f(x_j,\hat{\theta})\} \tag{108}$$

is a strongly consistent estimator of A_2,

$$\hat{A}_2 \to A_2 \text{ a.s.,} \quad \text{if } H_0 \text{ is true,} \tag{109}$$

whereas

$$\hat{A}_2 \to A_2 + \mathrm{E}\{g(x_j) - f(x_j,\theta_0)\}^2\{(\partial/\partial\theta')f(x_j,\theta_0)\}$$
$$\times\{(\partial/\partial\theta)f(x_j,\theta_0)\} \text{ a.s.,} \quad \text{if } H_1 \text{ is true.} \tag{110}$$

Now if we change $\hat{\mu}$ to

$$\hat{\mu} = 2^k\left\{\hat{\sigma}^2\prod_{l=1}^{k}\varepsilon_l - 2\,\mathrm{tr}(\hat{A}^{-1}\hat{B}_2) + \mathrm{tr}(\hat{A}^{-1}\hat{A}_2\hat{A}^{-1}\hat{B})\right\}, \tag{111}$$

where $\hat{A}$, $\hat{A}_2$ and $\hat{B}$ are defined by (16), (108) and (17), respectively, and

$$\hat{B}_2 = \frac{1}{n^2}\sum_{j_2=1}^{n}\sum_{j_2=1}^{n} u_{j_1}^2\{(\partial/\partial\theta')f(x_{j_1},\hat{\theta})\}\{(\partial/\partial\theta)f(x_{j_2},\hat{\theta})\}$$
$$\times\prod_{l=1}^{k}\frac{\sin(\varepsilon_l(z_{l,j_1} - z_{l,j_2}))}{z_{l,j_1} - z_{l,j_2}}, \tag{112}$$

then it is not too hard to verify that the Theorems 6 and 7 carry over.

Moreover, if we put

$$D_2 = \mathrm{E}(y_j - f(x_j,\theta_0))^2 z_j z_j', \tag{113}$$

$$C_2 = \mathrm{E}(y_j - f(x_j,\theta_0))^2 z_j(\partial/\partial\theta)f(x_j,\theta_0), \tag{114}$$

$$\Omega_2 = D_2 - CA^{-1}C_2' - C_2A^{-1}C' + CA^{-1}A_2A^{-1}C', \tag{115}$$

$$\hat{D}_2 = \frac{1}{n}\sum_{j=1}^{n}\hat{u}_j^2 z_j z_j', \tag{116}$$

$$\hat{C}_2 = \frac{1}{n}\sum_{j=1}^{n}\hat{u}_j^2 z_j(\partial/\partial\theta)f(x_j,\hat{\theta}), \tag{117}$$

$$\hat{\Omega}_2 = \hat{D}_2 - \hat{C}\hat{A}^{-1}\hat{C}_2' - \hat{C}_2\hat{A}^{-1}\hat{C}' + \hat{C}\hat{A}^{-1}\hat{A}_2\hat{A}^{-1}\hat{C}', \tag{118}$$

$$\hat{T} = (n/4\beta_n^2\hat{\eta}^2)(\hat{t}_1 - \hat{t}_2)'\hat{\Omega}_2^{-1}(\hat{t}_1 - \hat{t}_2), \tag{119}$$

and if we assume instead of Assumption 8 that:

Assumption 10. The matrix Ω_2 is non-singular,

then Theorem 8 carries over.

9. Some numerical applications

In this section we will demonstrate the performance of our tests by some numerical examples. The first set of examples concerns the same model as is used in Bierens (1981b) as a counter example of the consistency of the P-test of Davidson and MacKinnon (1981), i.e.,

Model 1:

$$y_j = x_{1,j} + x_{2,j} + x_{1,j}x_{2,j} + u_j, \qquad j = 1, 2, \ldots, n, \tag{120}$$

where the $x_{1,j}$'s, the $x_{2,j}$'s and the u_j's are independent normally distributed with zero mean and variances

$$\operatorname{var}(x_{1,j}) = \operatorname{var}(x_{2,j}) = 1, \qquad \operatorname{var}(u_j) = \sigma_u^2, \tag{121}$$

respectively. The second set of examples concerns

Model 2:

$$y_j = x_{1,j} + x_{2,j} + u_j, \qquad j = 1, 2, \ldots, n, \tag{122}$$

where the $x_{1,j}$'s, $x_{2,j}$'s and u_j's are as before. In both cases we test the null hypothesis:

$$H_0\colon y_j = \theta_1 x_{1,j} + \theta_2 x_{2,j} + \theta_3 + u_j, \quad \mathrm{E}(u_j \mid x_{1,j}, x_{2,j}) = 0 \text{ a.s.} \tag{123}$$

Clearly in the case of model 1 this hypothesis is false, and in the case of model 2 the hypothesis is true. Note that for model 1 White's version of Hausman's test [of the hypothesis (123)] is not consistent when the non-efficient estimator is the weighted least-squares estimator with weights

$$w(x_j) = x_{1,j}^2 + x_{2,j}^2,$$

for if $\sigma_u^2 = 1$, then the test statistic $\hat{m}$, say, involved satisfies

$$\hat{m} \to \frac{13}{2}\chi^2(k) \quad \text{in distr.,}$$

although the null is false.

We have generated 10 data sets $\{(y_1, x_{1,1}, x_{2,1}), \ldots, (y_{50}, x_{1,50}, x_{2,50})\}$ (thus $n=50$) according to model 1, where $\sigma_u = 0.2$, and we have done the same according to model 2, but now with $\sigma_u = \sqrt{1+(0.2)^2} = 1.0198$. The reason for choosing σ_u^2 in the latter way is that it then equals the variance of $x_{1,j}x_{2,j} + u_j$ in the former case, so that both models are comparable with respect to the least-squares fit. Moreover, in both cases we have chosen

$$N_0 = [-\varepsilon, \varepsilon] \times [-\varepsilon, \varepsilon] \quad \text{(compare Corollary 3)}, \tag{124}$$

$$\beta_n = 0.1 n^{0.1} \quad \text{(compare (22))}, \tag{125}$$

and

$$z_j = \Phi\begin{pmatrix} x_{1,j} \\ x_{2,j} \end{pmatrix} = \begin{pmatrix} \operatorname{atan}((x_{1,j} - \bar{x}_1)/s_1) \\ \operatorname{atan}((x_{2,j} - \bar{x}_2)/s_2) \end{pmatrix}, \tag{126}$$

[compare (7)], where $\bar{x}_i$ is the sample mean and s_i^2 is the sample variance of the $x_{i,j}$'s $(i=1,2)$. The reason for standardizing the $x_{i,j}$'s in (126) will be discussed later on.

For computational convenience we have applied our tests in the form as described in the sections 3 and 4. Moreover, in order to see how sensitive the results are for the choice of ε we have applied each test twice, namely for $\varepsilon = 1$ and $\varepsilon = 5$. The results are shown in tables 1 and 2.

In table 1 we see that the test statistics of both tests are rather sensitive for the choice of ε. The best results of test 1 are obtained for $\varepsilon = 1$ and the best results of test 2 are obtained for $\varepsilon = 5$. If we test the (false) null at the 90% confidence level then test 1 with $\varepsilon = 1$ rejects the null in all the cases and with

Table 1

Test results in the case of model 1.

Test 1: $\hat{\mu}/n\hat{\eta}$		Test 2: $\hat{T}$	
$\varepsilon=1$	$\varepsilon=5$	$\varepsilon=1$	$\varepsilon=5$
0.0541	0.3525	10.5797	20.4945
0.0492	0.2757	6.4709	18.0454
0.0611	0.4151	0.4353	0.4915
0.0523	0.2983	2.9768	8.5483
0.0522	0.2721	1.9082	2.8222
0.0507	0.2787	3.7641	4.3709
0.0509	0.3198	10.9897	22.3332
0.0506	0.2857	1.3695	1.9423
0.0526	0.2112	5.4794	11.5123
0.0569	0.3299	8.0171	12.1894

Table 2

Test results in the case of model 2.

Test 1: $\hat{\mu}/n\hat{\eta}$		Test 2: $\hat{T}$	
$\varepsilon=1$	$\varepsilon=5$	$\varepsilon=1$	$\varepsilon=5$
2.2595	1.6975	0.2046	0.2047
1.1244	0.8763	1.4869	1.4917
2.1964	1.0666	0.8812	0.8825
1.8216	0.8586	1.1774	1.1867
1.3288	1.0666	0.5223	0.5223
0.8583	0.8072	4.3740	4.3858
5.3457	1.2226	0.9707	0.9709
3.4192	1.2758	0.4052	0.4053
0.7901	0.8144	4.7971	4.8176
1.2477	0.7984	1.1161	1.1172

$\varepsilon=5$ in none of the cases, whereas test 2 with $\varepsilon=1$ rejects the null in 5 out of 10 cases and with $\varepsilon=5$ in 6 out of 10 cases.

The results in table 2 show that if the null is true then the test statistics are less sensitive to the choice of ε. Again proceeding to test at the 90% confidence level we see that test 1 accepts the (true) null in all the cases and test 2 accepts the null in 9 out of 10 cases.

We have also done some experiments with varying β_n, and it turns out that test 2 is not only sensitive to the choice of ε but also to the choice of β_n, especially if the null is false. However, in view of Theorem 8 this is not surprising.

Further research will be needed to establish optimal decision rules for choosing the various test parameters.

Finally, we will pay attention to standardizing the explanatory variables in (126). In the cases under review this would not be necessary, for plim $\bar{x}_i=0$ and plim $s_i=1$, but in general it avoids having the argument of the function $\operatorname{atan}(\cdot)$ become so large in absolute value that z_j is nearly constant. The reason for standardization is to preserve sufficient variation of the z_j's. This is necessary because otherwise we would have

$$\frac{1}{n}\sum_{j=1}^{n}\hat{u}_j \mathrm{e}^{\mathrm{i}t'z_j} \sim \frac{1}{n}\sum_{j=1}^{n}\hat{u}_j \frac{1}{n}\sum_{j=1}^{n}\mathrm{e}^{\mathrm{i}t'z_j}=0,$$

if the specification $f(x,\theta)$ contains a constant term, for in that case $\sum_{j=1}^{n}\hat{u}_j=0$. Obviously this situation would destroy all the power of the tests.

The asymptotic theory in the sections 6, 7 and 8 does not account for z_j's of the type (126). However, it is not too hard to verify that after some minor modifications the results in these sections carry over.

Appendix: Proofs of the Theorems 1 and 2

Proof of part I of Theorem 1

From Chung (1974, theorem 9.1.2) it follows that there exists a Borel measurable real function r, say, on $\boldsymbol{R}^k$ such that

$$\mathrm{E}(v \mid z) = r(z) \text{ a.s.} \tag{A.1}$$

Put

$$r_1(\cdot) = \max\{r(\cdot), 0\}, \qquad r_2(\cdot) = \max\{-r(\cdot), 0\}. \tag{A.2}$$

Then obviously r_1 and r_2 are non-negative Borel measurable real functions on $\boldsymbol{R}^k$ satisfying

$$r = r_1 - r_2. \tag{A.3}$$

Now assume for the moment

$$c_1 = \mathrm{E}r_1(z) > 0, \qquad c_2 = \mathrm{E}r_2(z) > 0. \tag{A.4}$$

Then we can define probability measures ν_1 and ν_2 on the Euclidean Borel field $\mathscr{B}$ as follows:

$$\nu_j(B) = \int_B r_j(x)\,\mathrm{d}\nu(x)/c_j, \qquad j = 1, 2, \tag{A.5}$$

where ν is the probability measure generated by the random vector z and B is an arbitrary Borel set in $\mathscr{B}$. Then we may write

$$\begin{aligned} \mathrm{E}v\,\mathrm{e}^{\mathrm{i}t'z} = \mathrm{E}r(z) &= \int r(x)\,\mathrm{e}^{\mathrm{i}t'x}\,\mathrm{d}\nu(x) \\ &= \int r_1(x)\,\mathrm{e}^{\mathrm{i}t'x}\,\mathrm{d}\nu(x) - \int r_2(x)\,\mathrm{e}^{\mathrm{i}t'x}\,\mathrm{d}\nu(x) \\ &= c_1 \int \mathrm{e}^{\mathrm{i}t'x}\,\mathrm{d}\nu_1(x) - c_2 \int \mathrm{e}^{\mathrm{i}t'x}\,\mathrm{d}\nu_2(x) \\ &= c_1\eta_1(t) - c_2\eta_2(t), \end{aligned} \tag{A.6}$$

say, where

$$\eta_j(t) = \int \mathrm{e}^{\mathrm{i}t'x}\,\mathrm{d}\nu_j(x)/c_j, \qquad j = 1, 2, \tag{A.7}$$

are the characteristic functions of the probability measures ν_j $(j = 1, 2)$, respectively.

If $\mathrm{E}v\,\mathrm{e}^{it'z} \equiv 0$ for every $t \in \boldsymbol{R}^k$ then it follows from (A.6) that

$$c_1\eta_1(t) - c_2\eta_2(t) = 0 \quad \text{for every} \quad t \in \boldsymbol{R}^k. \tag{A.8}$$

Hence, substituting $t=0$, we get

$$c_1\eta_1(0) - c_2\eta_2(0) = c_1 - c_2 = 0, \tag{A.9}$$

so that from (A.4), (A.8) and (A.9)

$$\eta_1(t) = \eta_2(t) \quad \text{for every} \quad t \in \boldsymbol{R}^k. \tag{A.10}$$

But (A.10) implies that the probability measures ν_1 and ν_2 are equal, i.e.,

$$\nu_1(B) = \nu_2(B) \quad \text{for every Borel set } B. \tag{A.11}$$

From (A.5), (A.9) and (A.11) we now obtain

$$\int_B r_1(x)\,\mathrm{d}\nu(x) = \int_B r_2(x)\,\mathrm{d}\nu(x) \quad \text{for every Borel set } B, \tag{A.12}$$

and consequently

$$\int_B r(x)\,\mathrm{d}\nu(x) = 0 \quad \text{for every Borel set } B. \tag{A.13}$$

But

$$B_1 = \{x \in \boldsymbol{R}^k : r(x) > 0\} \tag{A.14}$$

is a Borel set, and thus

$$0 = \int_{B_1} r(x)\,\mathrm{d}\nu(x), \tag{A.15}$$

which is only possible if B_1 is a null set with respect to ν. Similarly we conclude that the Borel set

$$B_2 = \{x \in \boldsymbol{R}^k : r(x) < 0\} \tag{A.16}$$

is a null set with respect to ν, and hence

$$B_1 \cup B_2 = \{x \in \boldsymbol{R}^k : r(x) \neq 0\} \tag{A.17}$$

is a null set with respect to v. This means that $r(z)=0$ a.s. This proves that if (A.4) holds and if $\mathrm{E}v\,\mathrm{e}^{\mathrm{i}t'z}=0$ for all $t\in\boldsymbol{R}^k$, then $\mathrm{E}(v\,|\,z)=0$ a.s. However, if (A.4) does not hold, then our conclusion still holds as is not hard to prove. This completes the 'only if' part of part I of Theorem 1. The 'if' part is trivial. Q.E.D.

Proof of part II of Theorem 1

Since z is now bounded we may write

$$\mathrm{E}v\,\mathrm{e}^{\mathrm{i}t'z}=\mathrm{E}v\sum_{j=0}^{\infty}\frac{i^j}{j!}(t'z)^j=\sum_{j=0}^{\infty}\frac{i^j}{j!}\,\mathrm{E}v(t'z)^j. \tag{A.18}$$

So if $\mathrm{E}v\,\mathrm{e}^{\mathrm{i}t'_*z}\neq 0$ for some $t_*\in\boldsymbol{R}^k$, then there exists a non-negative integer j_* such that

$$\mathrm{E}v(t'_*z)^{j*}\neq 0. \tag{A.19}$$

Assuming that j_* is minimal, we therefore have

$$\lim_{\lambda\downarrow 0}(\partial/\partial\lambda)^{j*}\mathrm{E}v\,\mathrm{e}^{\mathrm{i}\lambda t'_*z}=i^{j*}\mathrm{E}v(t'_*z)^{j*}\neq 0, \tag{A.20}$$

which implies that $\mathrm{E}v\,\mathrm{e}^{\mathrm{i}\lambda t'_*z}\neq 0$ for an arbitrarily small $\lambda>0$, say λ_*. Putting $t_0=\lambda_* t_*$, this proves part II of Theorem 1. Q.E.D.

Proof of Theorem 2

We can write $\mathrm{E}v(t'_*z)^{j*}$ as a sum of terms proportional to $\mathrm{E}v\prod_{l=1}^{k}z_l^{m_l}$, where $\sum_{l=1}^{k}m_l=j_*$. Thus Theorem 2 follows from (A.19). Q.E.D.

References

Aguirre-Torres, V. and A.R. Gallant, 1983, The null and non-null asymptotic distribution of the Cox test for multivariate nonlinear regression: Alternatives and a new distribution-free Cox test, Journal of Econometrics 21, no. 1, forthcoming.

Atkinson, A.C., 1969, A test for discriminating between models, Biometrica 56, 337–347.

Atkinson, A.C., 1970, A method for discriminating between models, Journal of the Royal Statistical Society B 32, 323–353.

Bierens, H.J., 1981a, Robust methods and asymptotic theory in nonlinear econometrics, in: Lecture notes in economics and mathematical systems, Vol. 192 (Springer-Verlag, Heidelberg).

Bierens, H.J., 1981b, A test for model specification in the absence of alternative hypotheses, Discussion paper no. 81–154 (Center for Economic Research, University of Minnesota, Minneapolis, MN).

Bierens, H.J., 1981c, A test for model specification of nonlinear time series regressions, Discussion paper no. 81–156 (Center for Economic Research, University of Minnesota, Minneapolis, MN).

Bierens, H.J., 1982, A uniform weak law of large numbers under ϕ-mixing with application to nonlinear least squares estimation, Statistica Neerlandica 36, 81–86.

Chung, K.L., 1974, A course in probability theory (Academic Press, New York).

Cox, D.R., 1961, Test for separate families of hypotheses, Proceedings of the 4th Berkeley Symposium 1, 105–123.

Cox, D.R., 1962, Further results on tests of separate families of hypotheses, Journal of the Royal Statistical Society B 24, 406–424.

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

Domowitz, I. and H. White, 1982, Misspecified models with dependent observations, Annals of Applied Econometrics, this issue.

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

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

Pereira, B. de B., 1977, A note on the consistency and on the finite sample comparisons of some tests of separate families of hypotheses, Biometrica 64, 109–113.

Pereira, B. de B., 1978, Tests and efficiencies of separate regression models, Biometrica 65, 319–327.

Pesaran, M.H. and A.S. Deaton, 1978, Testing non-nested nonlinear regression models, Econometrica 46, 677–694.

Quandt, R.E., 1974, A comparison of methods for testing nonnested hypotheses, Review of Economics and Statistics 56, 92–99.

Ramsey, J.B., 1969, Tests for specification errors in classical linear least-squares regression analysis, Journal of the Royal Statistical Society B 31, 350–371.

Ramsey, J.B., 1970, Models, specification error, and inference: A discussion of some problems in econometric methodology, Bulletin of the Oxford Institute of Economics and Statistics 32, 301–318.

White, H., 1980, Nonlinear regression on cross-section data, Econometrica 48, 721–746.

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