

TESTS FOR MODEL SPECIFICATION IN THE PRESENCE OF ALTERNATIVE HYPOTHESES

Some Further Results*

James G. MacKINNON

Queen's University, Kingston, Ont. K7L 3N6, Canada

Halbert WHITE

University of California, San Diego, La Jolla, CA 92093, USA

Russell DAVIDSON

Queen's University, Kingston, Ont. K7L 3N6, Canada

In Davidson and MacKinnon (1981), two of the present authors proposed a novel and very simple procedure for testing the specification of a nonlinear regression model against the evidence provided by a non-nested alternative. In this paper we extend their results in several directions. First, we relax a number of the assumptions of the previous paper, we admit the possibility that the nonlinear regression functions may depend on lagged dependent variables, and we do not require that the error terms be normally distributed. Second, we show how the earlier procedure may straightforwardly be generalized to the case where the two non-nested models involve different transformations of the dependent variable. Finally, we propose a simple procedure for testing non-nested Linear regression models which have endogenous variables on the right-hand side, and have therefore been estimated by two-stage least squares.

1. Introduction

In recent years several procedures have been proposed for testing the specification of a nonlinear regression model against the evidence provided by a non-nested alternative hypothesis. The first such tests were due to Pesaran (1974) and Pesaran and Deaton (1978), and were explicitly based on the classic work of Cox (1961, 1962). More recently, Davidson and MacKinnon (1981) proposed much simpler procedures based on artificial regression models, and showed the resulting tests are asymptotically equivalent to Cox tests. Indeed, White (1982) has shown that if one implements the Cox test in a straightforward fashion, one of the procedures of Davidson and MacKinnon is obtained directly. For a survey of this material, see MacKinnon (1982).

*Davidson and MacKinnon gratefully acknowledge financial support for the research of this paper from the Social Sciences and Humanities Research Council of Canada, White's participation was supported by NSF grant JES81-07552, and MacKinnon and White thank Ian Domowitz for numerous helpful discussions.

In this paper we extend the results of Davidson and MacKinnon (1981), hereafter referred to as DM, in several directions. Most importantly, we relax the relatively restrictive assumptions of that paper. In particular, we allow the nonlinear regression functions to depend on lagged dependent variables, and we do not require that the error terms be normally distributed. In addition, we show how one of the procedures of DM may straightforwardly be generalized to the case where the two non-nested models involve different transformations of the dependent variable. Finally, we propose a simple procedure for testing non-nested linear regression models which have been estimated by two-stage least squares. This is also a straightforward generalization of one of the procedures suggested by DM.

2 The J test, the P test and an extension

DM considered the following situation. The hypothesis to be tested is a single-equation, possibly nonlinear regression model,

$$H_0: y_t = f_t(X_t, \beta) + \varepsilon_{0t}, \quad (1)$$

and the alternative model is

$$H_1: y_t = g_t(Z_t, \gamma) + \varepsilon_{1t}. \quad (2)$$

Here X_t and Z_t represent the t th observations on vectors of exogenous variables, where the index t runs from 1 to n , β and γ are respectively a k -vector and a vector of parameters to be estimated, and ε_{it} is assumed to be $NID(0, \sigma_i^2)$ if H_i actually generated the data. There are certain further technical assumptions, which are stated in DM (pp. 784-785). In section 3 below we will substantially weaken the assumptions just stated on H_0 and H_1 ; for the moment, however, we will retain those of DM.

All of the tests in DM are based on the artificial compound model

$$H_c: y_t = (1 - \alpha)f_t(X_t, \beta) + \alpha g_t(Z_t, \gamma) + \varepsilon_t. \quad (3)$$

By itself this model is not very useful, since α , β and γ will generally not be identifiable. DM therefore suggested that γ be replaced by $\hat{\gamma}$, its least squares estimate, and showed that the t -statistic on $\hat{\alpha}$ is asymptotically $N(0, 1)$ when H_0 is true. They called this test the J test, because α and β are estimated jointly.

The J test is clearly extremely easy to perform so long as H_0 is a linear regression model (see section 4 below). However, when H_0 is nonlinear, so is the J test regression. To avoid the computational problems this may cause, DM suggested that this regression be linearized about the point $(\beta = \hat{\beta}, \alpha = 0)$,

to yield

$$y_t - \hat{f}_t = \alpha(\hat{g}_t - \hat{f}_t) + \hat{F}_t b + \varepsilon_t, \quad (4)$$

where \hat{F}_t is row vector containing the derivatives of f_t with respect to β , evaluated at $\hat{\beta}$, and b is a vector of regression coefficients. It is easy to see that the t-statistic on $\hat{\alpha}$ from (4) is

$$(y - \hat{f})^T \hat{M}_0 (\hat{g} - \hat{f}) / \hat{\sigma} ((\hat{g} - \hat{f})^T \hat{M}_0 (\hat{g} - \hat{f}))^{1/2}, \quad (5)$$

where y , \hat{f} and \hat{g} are vectors whose t th components are y_t , \hat{f}_t and g_t , respectively, $\hat{\sigma}$ is the estimated standard error from (4), and

$$\hat{M}_0 = I - \hat{F}(\hat{F}^T \hat{F})^{-1} \hat{F}^T, \quad (6)$$

where \hat{F} is the matrix whose t th row is \hat{F}_t .

Under the assumptions of DM, it is fairly easy to show that under H_0 (5) tends in probability to

$$\varepsilon_0^T M_0 (g - f) / \sigma_0 ((g - f)^T M_0 (g - f))^{1/2}, \quad (7)$$

where quantities without hats are evaluated at β_0 , the true value of β , or at γ_0 , the plim of $\hat{\gamma}$ under H_0 . It is obvious that (7) is $N(0, 1)$. Because of the role played by the projection matrix M_0 in (5), DM called the test based on (4) the P test.

In many applied cases, alternative non-nested models will utilize different transformations of the dependent variable. For example, the regressand might be $\log y_t$, $\exp y_t$, y_t^2 or $W_t y_t$ where W_t is some exogenous variable. The P test as derived above cannot be applied to such cases. Let us therefore consider the situation where H_0 is still given by eq. (1), but the alternative model is now

$$H_1 : h_t(y_t) = g_t(Z_t, \gamma) + \varepsilon_{1t}, \quad (8)$$

where $h_t(\cdot)$ may be any monotonic, continuously differentiable function which does not depend on any unknown parameters. There is obviously no loss of generality in assuming that y_t itself appears on the left-hand side of (1), since y_t can always be redefined appropriately.

Now consider the artificial compound model

$$H_c : (1 - \alpha)(y_t - f_t(\beta)) + \alpha(h_t(y_t) - g_t(\gamma)) = \varepsilon_t, \quad (9)$$

where the dependence of f_t and g_t on X_t and Z_t has been suppressed for notational convenience. If γ were replaced by $\hat{\gamma}$ this model could presumably be estimated to yield some sort of J test, but a highly nonlinear maximum likelihood procedure would have to be used, since the loglikelihood function would contain a Jacobian term as well as a sum of squares term. Simply taking a Taylor series approximation around $(\beta = \hat{\beta}, \alpha = 0)$, as in the P test, will not yield a valid test, however. The derivative of the left-hand side of (9) with respect to a , evaluated at $(\hat{\beta}, 0)$, is

$$-y_t + \hat{f}_t + h_t(y_t) - \hat{g}_t, \quad (10)$$

so that a straightforward P test regression would have terms involving y_t on the right-hand side. But this problem can be avoided if we replace y_t by \hat{f}_t whenever it appears on the right-hand side, an idea which was utilized in a related context by Andrews (1971). Following this procedure, we obtain the artificial regression

$$y_t - \hat{f}_t = \alpha(\hat{g}_t - h_t(\hat{f}_t)) + \hat{F}_t b + \varepsilon_t, \quad (11)$$

which is a rather elementary generalization of the P test regression (4). We shall therefore refer to this procedure as the extended P test or P_E test.

The principal merit of the P, test is simplicity; it certainly cannot be expected to have any optimality properties. This is so for two reasons. First, the P_E test is not a Lagrange Multiplier test based on the compound model (9) with γ replaced by $\hat{\gamma}$, and can therefore be expected to have less power than LM, Wald or LR tests based thereon. Secondly, the artificial compound model (9) is not equivalent to an exponential combination of the likelihood functions corresponding to H_0 and H_1 , so that LM, Wald and LR tests based on H_c will not be asymptotically equivalent to Cox tests [see Atkinson (1970)]. Thus the P, test is two steps removed from a Cox test. However, our experience with the test suggests that it often has plenty of power in applied situations, so that its theoretical deficiencies may be of small consequence to applied workers who find its simplicity appealing.

3. Validity of the tests under weak conditions

In this section we prove that the P and P_E tests are valid under much weaker conditions than those imposed by DM. We make extensive use of the following martingale central limit theorem, for which we provide a proof in the appendix. This theorem is very useful for many econometric applications.

Theorem 1. Define $Q_{nt} \equiv (Q_{nt1}, \dots, Q_{ntp})$ and let $\{Q_{nt}\}$ and $\{\varepsilon_t\}$ be $1 \times p$ and 1×1 stochastic processes such that, for each $n \geq 1$,

$$(a.1) \quad E(Q_{nt}^T \varepsilon_t | Q_{nt-1}^T \varepsilon_{t-1}, \dots, Q_{n0}^T \varepsilon_0) = 0, \quad t = 1, \dots, n,$$

$$(a.2) \quad E(|\varepsilon_t^2 Q_{nti} Q_{ntj}|^{1+\delta}) \leq \Delta < \infty, \quad \text{for some } \delta > 0, \quad i, j = 1, \dots, p,$$

and all $t = 1, \dots, n$,

$$(a.3) \quad \text{var} \left(n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt}^T \varepsilon_t \right) = I_p,$$

$$(a.4) \quad m^{-1} \sum_{t=1}^m \varepsilon_t^2 Q_{nt}^T Q_{nt} - E(\varepsilon_t^2 Q_{nt}^T Q_{nt}) \xrightarrow{p} 0 \quad \text{as } n \geq m \rightarrow \infty.$$

Then

$$n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt}^T \varepsilon_t \overset{\Delta}{\sim} N(0, I_p).$$

We are now ready to state formally the assumptions we shall need. We are interested in hypotheses about

$$E(y_t | y_{t-1}, y_{t-2}, \dots; W_t, W_{t-1}, \dots), \quad (12)$$

where y_t denotes the t th observation on a dependent variable and W_t denotes the t th vector of observations on exogenous variables. Thus (12) is just the mean of y_t conditional on its own past and on the past and present of W_t ; we may write it more compactly as $E(y_t | \mathcal{F}_t)$, where \mathcal{F}_t is the σ -algebra generated by $(y_{t-1}, y_{t-2}, \dots; W_t, W_{t-1}, \dots)$. The hypothesis we wish to test is that

$$H_0: E(y_t | \mathcal{F}_t) = f_t(X_t, \beta), \quad (13)$$

where X_t denotes a vector of m_0 elements, selected from some of the y_{t-i} 's ($i \geq 1$) and some of the W_{t-i} 's ($i \geq 0$). An alternative hypothesis is that

$$H_1: E(h_t(y_t) | \mathcal{F}_t) = g_t(Z_t, \gamma), \quad (14)$$

where $h_t(y_t)$ is a known function of y_t and Z_t denotes another vector of m_1 elements including some of the y_{t-i} 's and some of the W_{t-i} 's. The assumptions we shall require are extensions of those introduced by Domowitz and White (1982):

$$(A1) \quad E(y_t | \mathcal{F}_t) = f_t(X_t, \beta_0),$$

$$E((y_t - f_t(X_t, \beta_0))^2 | X_t, Z_t) = \sigma_0^2.$$

We may denote $y_t - f_t(X_t, \beta_0)$ as ε_t , dropping the zero subscript for convenience.

(A2) (a) There exist domains $X \subset \mathbf{R}^{m_0}$ and $Z \subset \mathbf{R}^{m_1}$ such that, for each x in X or z in Z , $f_t(x, \beta)$ and $g_t(z, \gamma)$ are continuous functions of β or γ uniformly in t almost surely, and measurable functions of x or z for each β in B or γ in Γ ; B and Γ are compact subsets of \mathbf{R}^k or \mathbf{R}^l , where k and l are finite.

(b) $h_t(y)$ is monotonic and continuously differentiable in y .

(A3) (a) $\{(y_t - f_t(X_t, \beta))^2\}$ and $\{(h_t(y_t) - g_t(Z_t, \gamma))^2\}$ are dominated by uniformly $(r_1 + \delta)$ -integrable functions, $r_1 \geq 1, 0 < \delta \leq r_1$.

(b) $\{(h_t(f_t(X_t, \beta)) - g_t(Z_t, \gamma))^2\}$ is dominated by uniformly $(r_1 + \delta)$ -integrable functions, $r_1 \geq 1, 0 < \delta \leq r_1$.

(A4) Define

$$\bar{\sigma}_{0n}^2 = (1/n) \sum_{t=1}^n E(y_t - f_t(X_t, \beta))^2, \quad (15)$$

$$\bar{\sigma}_{1n}^2 = (1/n) \sum_{t=1}^n E(h_t(y_t) - g_t(Z_t, \gamma))^2. \quad (16)$$

Let β_0 and γ_n^* be identifiably unique minimizers of $\bar{\sigma}_{0n}^2$ and $\bar{\sigma}_{1n}^2$, respectively, such that β_0 is interior to B and γ_n^* is interior to Γ uniformly in n .

(A5) $f_t(x, \beta)$ and $g_t(z, \gamma)$ are twice continuously differentiable in β or γ uniformly in t a.s.

(A6) $\{(y_t - f_t(X_t, \beta)) \partial f_t(X_t, \beta) / \partial \beta_i\}^2$ and $\{(h_t(y_t) - g_t(Z_t, \gamma)) \partial g_t(Z_t, \gamma) / \partial \gamma_i\}^2$ are dominated by uniformly r_2 -integrable functions, $r_2 > 1, i = 1, \dots, k$ or $i = 1, \dots, l$.

(A7) Define

$$B_{a,n}^g = \text{var} \left[n^{-\frac{1}{2}} \sum_{t=a+1}^{a+n} (-2(h_t(y_t) - g_t(Z_t, \gamma))) \nabla_{\gamma} g_t \right] \quad (17)$$

where $\nabla_{\gamma} g_t$ is the $l \times 1$ vector with elements $\partial g_t(Z_t, \gamma) / \partial \gamma_i, i = 1, \dots, l$. Let there exist an $l \times l$ matrix B^g such that $\det B^g > 0$ and $\lambda^T B_{a,n}^g \lambda - \lambda^T B^g \lambda \rightarrow 0$ as $n \rightarrow \infty$ uniformly in a for all real non-zero $l \times 1$ vectors λ .

- (A8) (a) $\{(\partial f_t(X_t, \beta)/\partial \beta_i)(\partial f_t(X_t, \beta)/\partial \beta_j) - (y_t - f_t(X_t, \beta))\partial^2 f_t(X_t, \beta)/\partial \beta_i \partial \beta_j\}$ and $\{(\partial g_t(Z_t, \gamma)/\partial \gamma_i)(\partial g_t(Z_t, \gamma)/\partial \gamma_j) - (h_t(y_t) - g_t(Z_t, \gamma))\partial^2 g_t(Z_t, \gamma)/\partial \gamma_i \partial \gamma_j\}$ are dominated by uniformly $(r_1 + \delta)$ -integrable functions, $r_1 \geq 1$, $0 < \delta \leq r_1$, $i, j = 1, \dots, k(l)$.
- (b) $\{(\partial h_t(f_t(X_t, \beta))/\partial y)^2\}$ is dominated by uniformly $(r_1 + \delta)$ -integrable functions, $r_1 \geq 1$, $0 < \delta \leq r_1$.
- (c) $\{(\partial f_t(X_t, \beta)/\partial \beta_i)^2\}$ and $\{(\partial g_t(Z_t, \gamma)/\partial \gamma_i)^2\}$ are dominated by uniformly $(r_1 + \delta)$ -integrable functions, $r_1 \geq 1$, $0 < \delta \leq r_1$, $i = 1, \dots, k(l)$.
- (A9) The Hessian matrices $V^2 \bar{\sigma}_{0n}^2(\beta)$ and $V^2 \bar{\sigma}_{1n}^2(\gamma)$ have constant rank in some open neighborhood of β_0 or γ_n^* for all n sufficiently large, uniformly in n .
- (A10) There exists $\delta > 0$ and $A < \infty$ such that $E|\varepsilon_t^4|^{1+\delta} \leq A$ and $E|(h_t(y_t) - g_t(Z_t, \gamma_n^*))^4|^{1+\delta} \leq A$ for all t .
- (A11) The random vectors $\{y_t, X_t, Z_t\}$ are either (a) ϕ -mixing of size $r_2/(r_2 - 1)$ or (b) α -mixing of size $\max[r_1/(r_1 - 1), r_2/(r_2 - 1)]$, $r_1, r_2 > 1$, where the mixing coefficients are defined with respect to the σ -algebras generated by (y_t, X_t, Z_t) .

We remark that conditions A2(b), A3(b) and A8(b) are always satisfied when $h_t(y) = y$. Conditions A1-A4 and A11 ensure the consistency of $\hat{\beta}_n$ for β_0 and $\hat{\gamma}_n$ for γ_n^* . Conditions A1-A8(b), A9 and A11 ensure the asymptotic normality of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma_n^*)$. Conditions A8(c) and A10 are additional conditions needed to ensure the asymptotic normality of the P test statistic. For more details on the mixing conditions of A11 and the definitions of identifiable uniqueness and uniform $(r_1 + \delta)$ -integrability, see White and Domowitz (1981) and Domowitz and White (1982). For a discussion of the interpretation of these conditions in the context of misspecified nonlinear regression models, see Domowitz and White (1982).

We are now ready to prove that the P and P_E test statistics are asymptotically distributed as $N(0, 1)$ under H_0 . We begin by considering the simpler case of the P test.

Theorem 2. Under assumptions A1-A11,

$$(y - \hat{f})^T \hat{M}_0(\hat{g} - \hat{f}) / \sigma((\hat{g} - \hat{f})^T \hat{M}_0(\hat{g} - \hat{f}))^{1/2} \stackrel{A}{\sim} N(0, 1).$$

provided that for some $c > 0$ and all n sufficiently large

$$\begin{aligned} & \sigma_0^2 \mathbf{E}((g-f)^T(g-f)/n) - \mathbf{E}((g-f)^T F/n) \mathbf{E}(F^T F/n)^{-1} \mathbf{E}(F^T(g-f)/n) \\ & \equiv \bar{\omega}_n^2 \geq c. \end{aligned} \quad (18)$$

Condition (18) ensures that $n^{-\frac{1}{2}}$ times the numerator of the statistic does not have a degenerate distribution. This condition is an extremely important one, and it can be violated in certain unusual cases. In particular, if the H_0 and H_1 models are orthogonal, so that $\mathbf{E}(F^T G/n)$ is a zero matrix, condition (18) will not be satisfied. Thus we are explicitly ruling out this and certain other even more unusual cases; see DM, footnote 3. For further discussion, see Aguirre-Torres and Gallant (1982).

Proof. Consider first the numerator of the test statistic. The first-order conditions for nonlinear least squares estimation ensure that $(y - \hat{f})^T \hat{F} = 0$, so that

$$\begin{aligned} n^{-\frac{1}{2}}(y - \hat{f})^T \hat{M}_0(\hat{g} - \hat{f}) &= n^{-\frac{1}{2}}(y - \hat{f})^T(\hat{g} - \hat{f}) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^n (y_i - \hat{f}_i)(\hat{g}_i - \hat{f}_i). \end{aligned} \quad (19)$$

By the mean value theorem of Jennrich (1969, lemma 3), given A4 and A5, the right-hand side of expression (19) is equal to

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^n (y_i - f_i^0)(g_i^* - f_i^0) + n^{-\frac{1}{2}} \sum_{i=1}^n (f_i - \bar{g}_i) \nabla f_i \sqrt{n}(\hat{\beta}_n - \beta_0) \\ & - n^{-\frac{1}{2}} \sum_{i=1}^n (y_i - \bar{f}_i) \nabla f_i \sqrt{n}(\hat{\beta}_n - \beta_0) + n^{-\frac{1}{2}} \sum_{i=1}^n (y_i - \bar{f}_i) \nabla g_i \sqrt{n}(\hat{\gamma}_n - \gamma_n^*), \end{aligned} \quad (20)$$

where $f_i^0 = f_i(X_i, \beta_0)$, $g_i^* = g_i(Z_i, \gamma_n^*)$, $\bar{f}_i = f_i(X_i, \bar{\beta}_n)$, $\bar{g}_i = g_i(Z_i, \bar{\gamma}_n)$, and ∇f_i and ∇g_i are row vectors whose i th elements are respectively $\partial f_i(X_i, \beta) / \partial \beta_i$ and $\partial g_i(Z_i, \gamma) / \partial \gamma_i$ evaluated at $\bar{\beta}_n$ and $\bar{\gamma}_n$; the point $(\bar{\beta}_n, \bar{\gamma}_n)$ being somewhere on a line segment connecting $(\hat{\beta}_n, \hat{\gamma}_n)$ and (β_0, γ_n^*) . The expansion applies for a sequence tail-equivalent to $(\hat{\beta}_n, \hat{\gamma}_n)$, but we maintain the same notation for convenience.

Now we know that

$$(\hat{\beta}_n - \beta_0) \xrightarrow{a.s.} 0 \quad \text{and} \quad (\hat{\gamma}_n - \gamma_n^*) \xrightarrow{a.s.} 0,$$

given A1-A4 and A11, by Corollary 3.1 of Domowitz and White (1982). Assumptions A3, A8 and A11 ensure that Theorem 2.5 and then Theorem 2.3 of Domowitz and White (1982) apply, yielding the following results:

$$\left| n^{-1} \sum_{i=1}^n (\hat{f}_i - \bar{g}_i) \nabla \hat{f}_i - n^{-1} \sum_{i=1}^n E[(f_i^0 - g_i^*) \nabla f_i^0] \right| \xrightarrow{a.s.} 0, \quad (21)$$

$$\left| n^{-1} \sum_{i=1}^n (y_i - \hat{f}_i) \nabla \hat{f}_i - n^{-1} \sum_{i=1}^n E[(y_i - f_i^0) \nabla f_i^0] \right| \xrightarrow{a.s.} 0, \quad (22)$$

$$\left| n^{-1} \sum_{i=1}^n (y_i - \hat{f}_i) \nabla \bar{g}_i - n^{-1} \sum_{i=1}^n E[(y_i - f_i^0) \nabla g_i^*] \right| \xrightarrow{a.s.} 0. \quad (23)$$

The expectations in the second summations of (22) and (23) vanish by A1; hence the first summations tend to zero a.s. Given A1-A8(b), A9 and A11, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma_n^*)$ are $O_p(1)$ as a consequence of Corollary 3.3 of Domowitz and White (1982). Hence, by 2c.4(x.a) of Rao (1973), which treats products of random sequences, the last two terms of (20) vanish in probability.

Further, the leading, $O_p(1)$, term in $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is given by the fact that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) - [E(F^T F/n)]^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \nabla f_i^{0T} \varepsilon_i \xrightarrow{p} 0, \quad (24)$$

by use of A1-A8(b), A9, and A11. It then follows from the result 2c.4(xiv) of Rao (1973), which deals with continuous functions of random sequences, that the above results imply

$$\begin{aligned} & n^{-\frac{1}{2}}(y - \hat{f})^T (\hat{g} - \hat{f}) - n^{-\frac{1}{2}} \sum_{i=1}^n [(g_i^* - f_i^0) \\ & + E((f - g)^T F/n) E(F^T F/n)^{-1} \nabla f_i^0] \varepsilon_i \xrightarrow{p} 0. \end{aligned} \quad (25)$$

We now turn our attention to the denominator of the test statistic. Consider the quantity

$$\begin{aligned} \hat{\omega}_n^2 & \equiv (1/n) \hat{\sigma}_0^2 (\hat{g} - \hat{f})^T \hat{M}_0 (\hat{g} - \hat{f}) \\ & = \hat{\sigma}_0^2 [(\hat{g} - \hat{f})^T (\hat{g} - \hat{f})/n - ((\hat{g} - \hat{f})^T \hat{F}/n) (\hat{F}^T \hat{F}/n)^{-1} (\hat{F}^T (\hat{g} - \hat{f})/n)]. \end{aligned} \quad (26)$$

Given the domination conditions of A3 and A8, repeated application of Theorems 2.5 and 2.3 of Domowitz and White (1982), followed by

application of Lemma 3.2 of White (1980), implies that $(\hat{\omega}_n^2 - \bar{\omega}_n^2) \xrightarrow{p} 0$. Since by (18) $\bar{\omega}_n^2 \geq c$ for all n sufficiently large, it follows from 2c.4(xiv) of Rao (1973) that

$$\begin{aligned} & n^{-\frac{1}{2}}(y - \hat{f})^T(\hat{g} - \hat{f})/\hat{\omega}_n \\ & - n^{-\frac{1}{2}} \sum_{t=1}^n (1/\bar{\omega}_n)[(g_t^* - f_t^0) + E((f - g)^T F/n)E(F^T F/n)^{-1} \nabla f_t^0] \varepsilon_t \xrightarrow{p} 0. \end{aligned} \quad (27)$$

Moreover, it follows from 2c.4(x.d) of Rao (1973) that the two quantities in (27) have the same asymptotic distribution. This distribution will now be computed for the second term.

Define the scalar

$$Q_{nt} = (1/\bar{\omega}_n)[(g_t^* - f_t^0) + E((f - g)^T F/n)E(F^T F/n)^{-1} \nabla f_t^0], \quad (28)$$

and consider the distribution of

$$n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt} \varepsilon_t, \quad (29)$$

which is the second term of (27). By AI, $E(Q_{nt} \varepsilon_t | Q_{n,t-1} \varepsilon_{t-1}, \dots, Q_{n0} \varepsilon_0) = 0$ for $t = 1, \dots, n$ and each $n \geq 1$. Further, by the definition of $\bar{\omega}_n$, $\text{var}(n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt} \varepsilon_t) = 1$. Given A3, A8 and A10 it can be verified that $E(|Q_{nt}^2 \varepsilon_t^2|^{1+\delta}) \leq A < \infty$ for some $\delta > 0$. Assumptions A3, A8 and A11 allow the strong law of large numbers for mixing sequences of McLeish (1975, lemma 2.9) to be applied, ensuring that $m^{-1} \sum_{t=1}^m Q_{mt}^2 \varepsilon_t^2 - E(Q_{mt}^2 \varepsilon_t^2) \xrightarrow{p} 0$ as $n \geq m \rightarrow \infty$. Thus the conditions of Theorem 1 are satisfied, and we conclude that

$$n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt} \varepsilon_t \overset{A}{\sim} N(0, 1). \quad (30)$$

From (27) then,

$$n^{-\frac{1}{2}}(y - \hat{f})^T(\hat{g} - \hat{f})/\hat{\omega}_n \overset{A}{\sim} N(0, 1). \quad (31)$$

The quantity on the left of (31) is simply the P test statistic, (5), in a slightly different guise. Hence Theorem 2 is proved.

The extension of the above proof to the P_E test statistic is straightforward. We shall merely state the more general result.

Theorem 3. Under assumptions AI-All,

$$(y - \hat{f})^T \hat{M}_0(\hat{g} - \hat{h})/\hat{\sigma}_0((\hat{g} - \hat{h})^T \hat{M}_0(\hat{g} - \hat{h}))^{\frac{1}{2}} \overset{A}{\sim} N(0, 1),$$

provided that for some $c > 0$ and all n sufficiently large

$$\hat{\sigma}_0^2(\hat{g} - \hat{h})^T \hat{M}_0(\hat{g} - \hat{h}) - c > 0 \quad \text{a.s.}$$

Here \hat{h} denotes the vector whose t th component is $h_t(\hat{f}_t)$. The proof of Theorem 3 parallels that of Theorem 2, with $(\hat{g} - \hat{h})$ replacing $(\hat{g} - \hat{f})$ throughout; it is therefore omitted.

4. A test for models estimated by two-stage least squares

Up to this point we have assumed that the right-hand side variables X , and Z_t are contemporaneously uncorrelated with the errors ε_t of the true model. But the basic idea of the J test can be applied to situations where this assumption does not hold. In this section we describe how the test may be modified to handle this situation, and prove that the modified test is valid asymptotically. One reason this is worth doing is that recent papers by Ericsson (1982) and Godfrey (1982) have proposed adaptations of the Cox test to handle models estimated by 2SLS, and these modified Cox tests are inevitably much more cumbersome than a modified J test.

For simplicity, we restrict ourselves to the case of linear models. The two non-nested models may be written as

$$H_0: y = X\beta + \varepsilon_0, \quad (32)$$

and

$$H_1: y = Z\gamma + \varepsilon_1, \quad (33)$$

where the matrix notation we employ is standard. Some of the columns of the X and Z matrices may be correlated with ε_0 , the error terms of the true model. Thus OLS estimation is inappropriate. However, there is assumed to exist a matrix of instruments, W , with the usual properties, so that 2SLS estimation is feasible. In contrast to the notation of the previous section, W can now contain lagged dependent variables in addition to exogenous variables, here denoted by the matrix Q . Our assumptions about X , Z , W and ε_0 will be stated formally below.

We are explicitly assuming that both competing hypotheses specify the same matrix of instruments. This assumption is somewhat restrictive, but is, we believe, a good one, even though it is entirely possible to devise tests based on more general assumptions. Such tests would have the undesirable property that their results might depend on which instruments were associated with each hypothesis, rather than on the specifications of H_0 and H_1 themselves. Moreover, our assumption makes it impossible for the applied worker to treat the same variables as exogenous in one model and

endogenous in the other, an error which could easily cause non-nested tests to yield misleading results.

Two-stage least squares estimates of H_0 and H , may be obtained by OLS estimation of

$$y = P_W X \beta + e_0, \quad (34)$$

and

$$y = P_W Z \gamma + e_1, \quad (35)$$

where $P_W = W(W^T W)^{-1} W^T$. These 2SLS estimates are

$$\hat{\beta} = (X^T P_W X)^{-1} X^T P_W y, \quad (36)$$

and

$$\hat{\gamma} = (Z^T P_W Z)^{-1} Z^T P_W y. \quad (37)$$

In order to calculate the 2SLS J test statistic, we presumably wish to estimate the equation

$$y = X \beta + \alpha Z \hat{\gamma} + \varepsilon \quad (38)$$

by 2SLS [where now $\beta = (1 - \alpha)\hat{\beta}$]. This may be done by OLS estimation of

$$y = P_W X \beta + \alpha P_{ZW} y + e, \quad (39)$$

where $P_{ZW} = P_W Z (Z^T P_W Z)^{-1} Z^T P_W$. If we multiply both sides of (39) by $M_{XW} = I - P_W X (X^T P_W X)^{-1} X^T P_W$, we obtain

$$M_{XW} y = \alpha M_{XW} P_{ZW} y + e'. \quad (40)$$

By standard results, the estimate of α and of its standard error from (40) are identical to those from (39), except for degrees of freedom corrections. Thus we see that the t-statistic on $\hat{\alpha}$ is

$$y^T P_{ZW} M_{XW} y / (\hat{\sigma} (y^T P_{ZW} M_{XW} P_{ZW} y)^{\frac{1}{2}}), \quad (41)$$

where $\hat{\sigma}$ denotes the usual 2SLS estimated standard error from (39). We now wish to prove that, under appropriate conditions, the test statistic (41) is asymptotically $N(0, 1)$.

Our first assumption expresses H_0 more formally:

(B1) It is known that for finite β_0 and π_0

$$E(y_t | y_{t-1}, y_{t-2}, \dots; Q_t, Q_{t-1}, \dots) = W_t \pi_0 \beta_0,$$

where W_t is a row vector whose components are some of the y_{t-i} 's, $i \geq 1$, and some of the Q_{t-i} 's, $i \geq 0$. By definition,

$$\pi_0 \equiv \text{a.s.} \lim \left(n^{-1} \sum_{i=1}^n W_i^T W_i \right)^{-1} \left(n^{-1} \sum_{i=1}^n W_i^T X_i \right).$$

Defining e_t as $y_t - W_t \pi_0 \beta_0$, it is also known that

$$E(e_t^2 | W_t) = \sigma_0^2.$$

The next condition imposes the moment conditions used to ensure the consistency and asymptotic normality of the 2SLS estimators:

- (B2) (a) $\{Z_t^2\}$ and $\{X_t^2\}$ are uniformly $(r_1 + \delta)$ -integrable, $r_1 \geq 1$, $0 < \delta \leq r_1$;
 (b) $\{W_t^4\}$ and $\{e_t^4\}$ are uniformly $(r_1 + \delta)$ -integrable, $r_1 \geq 1$, $0 < \delta \leq r_1$.

The following condition ensures that both models are identified

(B3) There exists $c > 0$ such that, for all n sufficiently large,

$$\det E(W^T W/n) \geq c,$$

$$\det E(X^T W/n) E(W^T W/n)^{-1} E(W^T X/n) \geq c,$$

$$\det E(Z^T W/n) E(W^T W/n)^{-1} E(W^T Z/n) \geq c,$$

where W , X and Z are the matrices with rows W_t , X_t , and Z_t , respectively.

The final assumption imposes restrictions on the memory of the random sequences considered

- (B4) The random vectors $\{y_t, W_t, X_t, Z_t\}$ are either (a) ϕ -mixing of size $r_1/(2r_1 - 1)$ or (b) α -mixing of size $r_1/(r_1 - 1)$, $r_1 > 1$.

Theorem 4. Under assumptions B1-B4,

$$y^T P_{ZW} M_{XW} y / (\hat{\sigma}^2 (y^T P_{ZW} M_{XW} P_{ZW} y)^{\ddagger}) \stackrel{A}{\sim} N(0, 1),$$

provided that, for some $c > 0$ and all n sufficiently large,

$$\hat{\sigma}^2(y^T P_{ZW} M_{XW} P_{ZW} y/n) - c > 0 \quad \text{a.s.}$$

Proof. First consider the numerator, $y^T P_{ZW} M_{XW} y$. Provided that $W^T Z/n - E(W^T Z/n)$, $W^T W/n - E(W^T W/n)$, $W^T X/n - E(W^T X/n)$ and $W^T e/n$ all tend to zero in probability, which is ensured by **B1**, **B2** and **B4**; provided that **B3** holds; and provided that $n^{-\frac{1}{2}} \sum_{t=1}^n W_t^T e_t$ is $O_p(1)$, which is ensured by **B1** and **B2**; it follows that

$$\sqrt{n} y^T P_{ZW} M_{XW} y - n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt}^* e_t \xrightarrow{p} 0,$$

where

$$\begin{aligned} Q_{nt}^* &\equiv \beta_0^T E(X^T W/n) E(W^T W/n)^{-1} E(W^T Z/n) \\ &\quad [E(Z^T W/n) E(W^T W/n)^{-1} E(W^T Z/n)]^{-1} \\ &\quad E(Z^T W/n) E(W^T W/n)^{-1} \{I - E(W^T X/n) \\ &\quad [E(X^T W/n) E(W^T W/n)^{-1} E(W^T X/n)]^{-1} \\ &\quad E(X^T W/n) E(W^T W/n)^{-1}\} W_t^T. \end{aligned} \quad (42)$$

Let $\hat{\omega}_n^2 \equiv \text{var}(n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt}^* e_t)$. Then assumptions **B1**–**B4** ensure that $\hat{\omega}_n^2 = \hat{\sigma}^2(y^T P_{ZW} M_{XW} P_{ZW} y/n)$ is consistent for $\bar{\omega}_n^2$ by repeated application of Lemma 2.9 of McLeish (1975) and Lemma 3.2 of White (1980).

Since $\bar{\omega}_n^2 > c$ for all n sufficiently large by assumption, it follows from 2c.4(xiv) of Rao (1973) that

$$\sqrt{n} y^T P_{ZW} M_{XW} y / \hat{\omega}_n - n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt} e_t \xrightarrow{p} 0, \quad (43)$$

where $Q_{nt} \equiv Q_{nt}^* / \hat{\omega}_n$. It follows from 2c.4(x.d) of Rao (1973) that the two quantities in (43) have the same asymptotic distribution.

We now show that $n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt} e_t \xrightarrow{d} N(0, 1)$ by verifying that the conditions of Theorem 1 hold. By **B1**, $E(Q_{nt} e_t | Q_{n,t-1} e_{t-1}, \dots, Q_{n0} e_0) = 0$. Next, by the definitions of Q_{nt} and $\bar{\omega}_n^2$, $\text{var}(n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt} e_t) = 1$. Given **B2** it can be verified that $E(|Q_{nt}^2 e_t^2|^{1+\delta}) \leq \Delta < \infty$ for some $\delta > 0$. Finally,

$$m^{-1} \sum_{t=1}^m Q_{nt}^2 e_t^2 - E(Q_{nt}^2 e_t^2) \xrightarrow{p} 0,$$

by Lemma 2.9 of McLeish (1975), given assumptions **B1**–**B4**. We therefore conclude that both quantities in (43) are asymptotically $N(0, 1)$, which completes the proof.

5. Conclusion

We have shown that the model specification tests of DM are valid under quite general conditions. The P test may be applied to dynamic nonlinear regression models whose error terms are serially uncorrelated and homoskedastic, and obey weak moment conditions. We have proposed an extension of the P test, the P_E test, which may be used under similar conditions when the dependent variable is transformed differently in the H_0 and H_1 models. We have also shown that a procedure which is computationally identical to the J test, except that 2SLS replaces OLS throughout, may validly be used when the competing hypotheses are linear models with endogenous variables on the right-hand side. Thus these tests are widely applicable as well as computationally convenient.

Appendix

In this appendix we prove Theorem 1. We apply Scott's (1973) Theorem 2, a functional central limit theorem for triangular martingale arrays. Define

$$Z_t(n) \equiv \lambda^T Q_{nt}^T \varepsilon_t / (\sqrt{n(\lambda^T \lambda)^{\frac{1}{2}}}),$$

$$S_m(n) \equiv \sum_{t=1}^m Z_t(n),$$

$$s_m^2(n) \equiv E(S_m^2(n)),$$

where $A \neq 0$ is any real $p \times 1$ vector. We show that

$$S_n(n) = \lambda^T n^{-\frac{1}{2}} \sum_{t=1}^n Q_{nt}^T \varepsilon_t / (\lambda^T \lambda)^{\frac{1}{2}} \stackrel{A}{\sim} N(0, 1),$$

for any λ , so that the desired result follows from the Cramer-Wold device; see Rao (1973, 2c.4(xi)).

We begin by verifying that Scott's conditions (c) [Scott (1973, p. 130)] are satisfied. These conditions are

$$\sum_{t=1}^{m_n(\phi)} Z_t^2(n) \xrightarrow{P} \phi \quad \text{as } n \rightarrow \infty, \quad 0 < \phi \leq 1, \quad (1)$$

$$\sup_{t \leq n} Z_t^2(n) \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty, \quad (2)$$

where $m_n(\phi) = \max \{m \leq n : s_m^2(n) \leq \phi\}$; that is, the largest index such that $s_m^2(n)$ is equal to or less than ϕ .

First, we verify (2). For any $\eta > 0$,

$$\begin{aligned} P\left[\sup_{t \leq n} Z_t^2(n) > \eta\right] &= P\left[\bigcup_{t=1}^n \{Z_t^2(n) > \eta\}\right] \\ &\leq \sum_{t=1}^n P[Z_t^2(n) > \eta] \\ &= \sum_{t=1}^n E(|Z_t^2(n)|^{1+\delta})/\eta^{1+\delta} \end{aligned}$$

[see, for example, Tucker (1967, p. 39, theorem 2)]. From the definition of $Z_t^2(n)$, we have

$$\begin{aligned} E(|Z_t^2(n)|^{1+\delta}) &= E\left(\left|\sum_{i=1}^p \sum_{j=1}^p \lambda_i \lambda_j \varepsilon_i^2 Q_{ni} Q_{nj} / (n(\lambda^T \lambda))\right|^{1+\delta}\right) \\ &\leq \sum_{i=1}^p \sum_{j=1}^p \mu_{ij} |\lambda_i \lambda_j|^{1+\delta} E|\varepsilon_i^2 Q_{ni} Q_{nj}|^{1+\delta} / (n^{1+\delta} (\lambda^T \lambda)^{1+\delta}) \\ &\leq \Delta \sum_{i=1}^p \sum_{j=1}^p \mu_{ij} |\lambda_i \lambda_j|^{1+\delta} / (n^{1+\delta} (\lambda^T \lambda)^{1+\delta}), \end{aligned}$$

by repeated application of the c_r -inequality [Loève (1963, p. 155)] and condition (a.2) of the theorem, where the μ_{ij} are positive constants bounded above. Combining the above inequalities yields

$$P\left[\sup_{t \leq n} Z_t^2(n) > \eta\right] \leq \Delta'(\lambda) / n^\delta,$$

where

$$\Delta'(\lambda) \equiv \Delta \sum_{i=1}^p \sum_{j=1}^p \mu_{ij} |\lambda_i \lambda_j|^{1+\delta} / ((\lambda^T \lambda)^{1+\delta} \eta^{1+\delta}) < \infty,$$

and $\Delta'(\lambda)$ is independent of n . Hence (2) holds.

To verify (1) we successively establish that

$$m_n(\phi) \rightarrow \infty \quad \text{as } n \rightarrow \infty, \quad 0 < \phi \leq 1, \quad (3)$$

$$\sum_{t=1}^m Z_t^2(n) - s_m^2(n) \xrightarrow{p} 0 \quad \text{as } n \geq m \rightarrow \infty, \quad (4)$$

$$s_{m_n(\phi)}^2(n) \rightarrow \phi \quad \text{as } n \rightarrow \infty. \quad (5)$$

Combining (3)–(5), with m replaced by $m_n(\phi)$ in (4), we obtain (1).

From (a.1) it follows that

$$s_m^2(n) = (m/n)\lambda^T(1/m) \sum_{i=1}^m E(\varepsilon_i^2 Q_{mi}^T Q_{mi}) \lambda / (\lambda^T \lambda).$$

Further, (a.2) guarantees that the eigenvalues of $(1/m) \sum_{i=1}^m E(\varepsilon_i^2 Q_{mi}^T Q_{mi})$ are uniformly bounded above. Hence there exists $A^* < \infty$ such that

$$s_m^2(n) \leq A^* m/n$$

The function $\bar{s}_m^2(n) \equiv \min(A^* m/n, 1)$ never lies below $s_m^2(n)$, since $s_m^2(n)$ is non-decreasing in m and $s_m^2(n) = 1$ given (a.3). *Defining*

$$\bar{m}_n(\phi) \equiv \max\{m \leq n : \bar{s}_m^2(n) \leq \phi\},$$

we see that $m_n(\phi) \geq \bar{m}_n(\phi)$. For $0 < \phi < 1$, $\bar{m}_n(\phi) \geq (\phi n / A^*) - 1$, while for $\phi = 1$, $\bar{m}_n(\phi) = n$. Hence

$$\begin{aligned} m_n(\phi) &\geq (\phi n / A^*) - 1; & 0 < \phi < 1, \\ &= n, & \phi = 1, \end{aligned}$$

so that $m_n(\phi) \rightarrow \infty$ as $n \rightarrow \infty$, $0 < \phi \leq 1$, establishing (3).

Next, by the martingale property (a.1),

$$\sum_{i=1}^m Z_i^2(n) - s_m^2(n) = \sum_{i=1}^m [Z_i^2(n) - E(Z_i^2(n))]$$

so that

$$\sum_{i=1}^m Z_i^2(n) - s_m^2(n) = (m/n)\lambda^T \left[m^{-1} \sum_{i=1}^m \varepsilon_i^2 Q_{mi}^T Q_{mi} - E(\varepsilon_i^2 Q_{mi}^T Q_{mi}) \right] \lambda / (\lambda^T \lambda).$$

Since $m_{n,\phi} \leq 1$, it follows from (a.4) that (4) holds.

To see that (5) is also valid, note that $s_{m_n(\phi)+1}^2(n) > \phi$, since $m_n(\phi)$ is the largest index such that $s_m^2(n) \leq \phi$. Hence

$$E(Z_{m_n(\phi)+1}^2(n)) = s_{m_n(\phi)+1}^2(n) - s_{m_n(\phi)}^2(n) > \phi - s_{m_n(\phi)}^2(n).$$

But from (a.2) there exists $\Delta^0(\lambda)$ such that

$$E(Z_{m_n(\phi)+1}^2(n)) < \Delta^0(\lambda)/n,$$

so that

$$\Delta^0(\lambda)/n > \phi - s_{m_n(\phi)}^2(n) \geq 0,$$

which implies that $s_{m_n(\phi)}^2(n) \rightarrow \phi$ as $n \rightarrow \infty$. As noted above, (1) now follows from (3)–(5), with m replaced by $m_n(\phi)$ in (4). It then follows from Scott's (1973) Theorem 2 that $S_n(n) \overset{A}{\sim} N(0, 1)$ for any λ . Hence from 2c.4(xi) of Rao (1973), it follows that

$$n^{-\frac{1}{2}} \sum_{t=1}^n Q_{tn}^T \varepsilon_t \overset{A}{\sim} N(0, I_p). \quad \text{Q.E.D.}$$

References

- Aguirre-Torres, V. and A.R. Gallant, 1982, 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*, this issue.
- Andrews, D.F., 1971, A note on the selection of data transformations, *Biometrika* 58, 249–254.
- Atkinson, A.C., 1970, A method for discriminating between models, *Journal of the Royal Statistical Society B* 32, 323–353.
- Cox, D.R., 1961, Tests of separate families of hypotheses, in: *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol. 1 (University of California Press, Berkeley, CA) 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, *Journal of Econometrics* 20, 35–58.
- Eriasson, N.R., 1982, Asymptotic properties of statistics for testing non-nested hypotheses, *Review of Economic Studies*, forthcoming.
- Godfrey, L.G., 1982, Testing non-nested models after estimation by instrumental variables or least squares, *Econometrica*, forthcoming.
- Jennrich, R.I., 1969, Asymptotic properties of nonlinear least squares estimators, *Annals of Mathematical Statistics* 40, 633–643.
- Loève, M., 1963, *Probability theory* (Van Nostrand, New York).
- MacKinnon, J.G., 1982, Model specification tests against non-nested alternatives, *Econometric Reviews*, forthcoming.
- McLeish, D.L., 1975, A maximal inequality and dependent strong laws, *Annals of Probability* 3, 826–836.
- Pesaran, M.H., 1974, On the general problem of model selection, *Review of Economic Studies* 41, 153–171.
- Pesaran, M.H. and A.S. Deaton, 1978, Testing non-nested nonlinear regression models, *Econometrica* 46, 677–694.
- Rao, C.R., 1973, *Linear statistical inference and its applications* (Wiley, New York).
- Scott, D.J., 1973, Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach, *Advances in Applied Probability* 5, 119–137.
- Tucker, H.G., 1967, *A graduate course in probability* (Academic Press, New York).
- White, H., 1980, Nonlinear regression on cross-section data, *Econometrica* 48, 721–746.
- White, H., 1982, Regularity conditions for Cox's test of non-nested hypotheses, *Journal of Econometrics* 19, 301–318.
- White, H. and I. Domowitz, 1981, Nonlinear regression with dependent observations, Discussion paper no. 81-32 (Department of Economics, University of California, San Diego, CA).

MULTIPLE MODEL TESTING FOR NON-NESTED HETEROSKEDASTIC CENSORED REGRESSION MODELS

Marlene A. SMITH and G. S. MADDALA

University of Florida, Gainesville, FL 32611, USA

1. Introduction

Applied econometric research is often characterized by the search for a 'suitable' model describing an economic relationship. Misspecification is frequently determined by the outcome of estimated t-ratios, or the decision to include exogenous variables is based on minimum mean square error. In order to assure that the selected model satisfies the classical assumptions, residuals may be tested for heteroskedastic or autocorrelated behavior. Furthermore, some measure of goodness of fit is employed to compare competing models.

Perhaps the comments of Pesaran (1974, p. 154) best characterize the use of such techniques in economic analysis:

'There is no theoretical justification for expecting a correctly specified model to possess all the characteristics of the classical regression models. The assumptions underlying the classical regression models are made, not because they are optimal from the point of view of economic theory, but because they are extremely convenient for estimation and hypothesis testing purposes . . . Consequently, it seems more appropriate to treat the problem of choosing among alternative models as an hypothesis testing problem rather than as an arbitrary definition of what a "true model" should be.'

Several model selection criteria have recently been suggested which address these criticisms [see Sawyer (1980) for a survey]. Separate criteria demonstrate considerable flexibility in their ability to distinguish among models of different dimension, distributional specification, and functional form of the dependent and independent variables. As an example, an asymmetrical test was originally proposed by Cox (1961, 1962), and later applied to regression problems by Pesaran (1974). The Cox test of the null hypothesis is defined to be

$$T_f = L_f(\hat{\alpha}) - L_g(\hat{\beta}) - E_n[L_f(\hat{\alpha}) - L_g(\hat{\beta})],$$

where $L_f(\hat{\alpha})$ is the loglikelihood of the null model (H_0) evaluated at its maximum likelihood estimators ($\hat{\alpha}$), $L_g(\hat{\beta})$ is the loglikelihood of the alternative model (H_1) evaluated at its MLE ($\hat{\beta}$), and E_x refers to the asymptotic expectation under the null model when α is evaluated at its MLE. As an interpretation, the test compares the loglikelihood ratio of the competing models to their expected values when constrained to the theoretical values under the null hypothesis. The joint test of T_f and

$$T_g = L_g(\hat{\beta}) - L_f(\hat{\alpha}) - E_{\hat{\beta}}[L_g(\hat{\beta}) - L_f(\hat{\alpha})]$$

yields nine possible outcomes, including the rejection or acceptance of both models. That is, the Cox test, like most separate criteria, may not give an absolute ranking to competing models (as would, for example, an \bar{R}^2 or likelihood ratio test). Pesaran and Deaton (1978) extend the results to non-linear, multivariate regression models.

Many of the separate criteria applied to non-nested models, such as the tests suggested by Cox are constrained to binary comparisons of competing models. A natural evolution in the econometric literature has been the development of multiple model selection criteria. This is the emphasis of the work by Davidson and MacKinnon (1981) and Sawyer (1980). Specifically, Sawyer develops the multiple model equivalent of the Cox test. Furthermore, simulation results conducted there indicate that joint tests of all models under consideration are more powerful than pairwise comparisons when the competing models are sufficiently disparate. Thus, in the choice among several separate hypotheses, one would expect that the probability of selecting the 'true' model is enhanced by the use of a multiple model criterion.

This paper investigates the applicability of multiple model selection in censored regression models characterized by a heteroskedastic disturbance structure. Smith (1982) derived the Cox test for the Tobit model with spherical error terms. However, the Cox test is intractable for the heteroskedastic limited variable model. Therefore, we rely on the linear embedding procedures suggested by Davidson and MacKinnon. Section 2 discusses the heteroskedastic Tobit model. Section 3 contains a description of the artificially-embedded procedures for the non-linear regression model. In section 4 we consider an empirical study of the demand for demand deposits, that motivated the current discussion. We present there the results of the J test for model selection. The final section gives an interpretation of the results as well as a discussion of the problems arising from multiple model testing.

2. The heteroskedastic Tobit model

The statistical properties of a censored heteroskedastic regression model

are discussed in Fische, Maddala and Trost (1979). Specifically, it is shown that ignoring heteroskedasticity in the Tobit model will yield inconsistent estimates of the unknown parameters.

Fische et al. provide an estimator for the Tobit model based on the procedure of Rutemiller and Bowers (1968). Consider the model

$$\begin{aligned} Y_i &= X_i \beta_0 + U_{0i} \quad \text{if } \text{RHS} > 0, \\ &= 0 \quad \text{otherwise,} \end{aligned} \quad (1)$$

where it is assumed that

$$U_{0i} \sim \text{IN}(0, \sigma_{0i}^2), \quad \sigma_{0i}^2 = (\alpha_0 + \alpha_1 \bar{X}_i)^2.$$

Here, \bar{X}_i is some subset of the X_i variables in (1). Within this specification of the variance, an appropriate test for $\alpha_1 = 0$ is used to detect the presence of heteroskedasticity. The estimation of β_0 , α_0 , and α_1 requires the use of a non-linear maximization technique. More specifically, the loglikelihood function of (1) can be written as

$$\begin{aligned} L_0 &= \sum_2 \log(1 - F_i) - (N_1/2) \log(2\pi) \\ &\quad - \frac{1}{2} \sum_1 \log(\alpha_0 + \alpha_1 \bar{X}_i)^2 - \frac{1}{2} \sum_1 (\alpha_0 + \alpha_1 \bar{X}_i)^{-2} (Y_i - X_i \beta_0)^2, \end{aligned} \quad (2)$$

where

$$F_i = \int_{-\infty}^{X_i \beta_0 (\alpha_0 + \alpha_1 \bar{X}_i)^{-1}} (1/\sqrt{2\pi}) \exp[-t^2/2] dt.$$

\sum_1 and \sum_2 refer to summation over observations for which $Y_i > 0$ and $Y_i = 0$, respectively. The likelihood equations are

$$\frac{\partial L_0}{\partial \beta_0} = -\sum_2 h_{0i} X_i' + \sum_1 \frac{1}{\sigma_{0i}^2} (Y_i - X_i \beta_0) X_i' = 0,$$

$$\frac{\partial L_0}{\partial \alpha_0} = \sum_2 \frac{h_{0i}(X_i \beta_0)}{\sigma_{0i}} + \sum_1 \frac{(Y_i - X_i \beta_0)^2 - \sigma_{0i}^2}{\sigma_{0i}^3} = 0,$$

$$\frac{\partial L_0}{\partial \alpha_1} = \sum_2 \frac{h_{0i}(X_i \beta_0) \bar{X}_i'}{\sigma_{0i}} + \sum_1 \left[\frac{(Y_i - X_i \beta_0)^2 - \sigma_{0i}^2}{\sigma_{0i}^3} \right] \bar{X}_i' = 0,$$

where

$$h_{0i} = f_{0i}/(1 - F_{0i}),$$

$$f_{0i} = \frac{1}{\sigma_{0i}\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_{0i}^2}(X_i\beta_0)^2\right].$$

These equations are solved iteratively using the Berndt et al. (1974) method to give the ML estimates of the different parameters.

3. Artificial embedding procedures for model selection

The artificial embedding procedures involve combining the non-nested models under consideration into a comprehensive model by introducing an artificial embedding parameter. Estimation and hypothesis testing of the embedding parameter serve as the selection criterion. Within this general framework, there are obviously many different ways to construct the model. The emphasis of many of the artificial embedding procedures has been the construction of comprehensive models which permit straightforward identification of the distribution and statistical properties of the embedding parameter.

Davidson and MacKinnon (1981) design a series of linear embedding procedures which may be applied to linear, or non-linear, non-nested models. They have the advantage of being computationally simple, and are easily extended to multiple model testing. Consider the situation of two non-nested, non-linear models,

$$H_0: Y_i | X_i, Z_i \sim N(f_i(X_i, \beta_0), \sigma_0^2) \quad \text{for some } \beta_0 \text{ in } B, \quad (3)$$

$$H_1: Y_i | X_i, Z_i \sim N(g_i(Z_i, \gamma_0), \sigma_1^2) \quad \text{for some } \gamma_0 \text{ in } \Gamma, \quad (4)$$

where Y_i is the vector of explanatory variables, X_i and Z_i are non-stochastic vectors of explanatory variables under H_0 and H_1 , respectively, and β_0 and γ_0 are unknown parameters,

Davidson and MacKinnon suggest the comprehensive model

$$Y_i = (1 - \lambda)f_i + \lambda g_i + U_i, \quad (5)$$

Here λ serves as the artificial embedding parameter, and the arguments of $f_i(X_i, \beta_0)$ and $g_i(Z_i, \gamma_0)$ have been suppressed for brevity. The test for $\lambda = 0$ is the test for the truth of H_0 . However, λ is not identified, and no inferences can be drawn from the estimation of (5). Three procedures are suggested the J test, the P test and the C test. Since we will be using only the J test in the subsequent work, we will omit the discussion of the P test and the C test.

The J test requires the substitution of β_1 by its maximum likelihood estimator ($\hat{\beta}_1$). The comprehensive model then becomes

$$Y_i = (1 - \lambda)f_i + \lambda\hat{g}_i + U_i^*, \quad (6)$$

where $\hat{g}_i = g_i(Z_i, \hat{\gamma}_0)$. $\hat{\gamma}_0$ will be independent of U_i^* as the sample size increases. Thus, it is theoretically correct to perform an asymptotic t test on the estimated value of λ . The J test is computationally simple to perform with conventional software packages when H_0 is linear. The steps required are:

- (i) compute the maximum likelihood estimator of $\hat{\beta}_1$ in H_1 ,
- (ii) substitute the predicted values, \hat{g}_i , into (5) and obtain estimates of λ ,
- (iii) test for the truth of H_0 using an asymptotic t test or likelihood ratio test.

It is tempting to test $\lambda = 1$ in the comprehensive model as an indication of the truth of H_1 . However, Davidson and MacKinnon show that this test lacks power, since the t statistics generated from (6) are valid only when H_0 is true. Thus, it is suggested that the roles of H_0 and H_1 be reversed. For example, the J test for H_1 should be computed from

$$Y_i = (1 - \lambda)g_i + \lambda\hat{f}_i + U_i^*. \quad (7)$$

When neither H_0 nor H_1 is true, it will be possible to reject both hypotheses.

Finally, the linear embedding procedures easily generalize to simultaneous multiple model selection. For m alternative models, construct

$$Y_i = \left(1 - \sum_{j=1}^m \lambda_j\right) f_i + \sum_{j=1}^m \lambda_j \hat{g}_{ij} + U_i \quad (8)$$

This is the multiple model generalization of the J test, where the joint test of $\lambda_j = 0$ is appropriate. The likelihood ratio test is employed which will be asymptotically chi-squared distributed with m degrees of freedom.

The computational simplicity of the artificial embedding technique becomes a major advantage over the **Cox** test in the presence of models characterized by complicated likelihood functions. As an illustration, the following section presents an empirical application of the multiple model J test to the heteroskedastic Tobit model. We investigate the following m non-nested models:

$$\begin{aligned} H_j: Y_i &= X_{ij}\beta_j + U_{ji} & \text{if } \text{RHS} > 0, & \quad i = 1, 2, \dots, N, \\ &= 0 & \text{otherwise,} & \quad j = 1, 2, \dots, m, \end{aligned} \quad (9)$$

where

$$u_{ji} \sim \text{IN}(0, \sigma_{ji}^2), \quad \sigma_{ji}^2 = (\alpha_{j0} + \alpha_{j1} Z_i)^2. \quad (10)$$

Note that we assume that the variables that affect the variance are the same in all the models though the explanatory variables X_{ji} in (9) differ. The comprehensive model for testing the k th model against all others can be written as

$$Y_i = \begin{cases} \left(1 - \sum_{\substack{j=1 \\ j \neq k}}^m\right) X_{ki} \beta_k + \sum_{\substack{j=1 \\ j \neq k}}^m \lambda_j X_{ji} \hat{\beta}_j + u_{ki}^* & \text{if RHS} > 0, \\ = 0 & \text{otherwise.} \end{cases} \quad (11)$$

Because of the assumptions in (10), u_{ki}^* will be $\text{IN}(0, \sigma_{ki}^{*2})$ where σ_{ki}^* is a linear function of Z_i .

Here $\hat{\beta}_j$ refer to the heteroskedastic Tobit estimates obtained from (9) individually and, as before, λ_j refer to the embedding parameters. We will use

$$2 \ln [L(\hat{\lambda}) - L(0)] \quad (12)$$

as the test statistic. Here $L(0)$ and $L(\hat{\lambda})$ are the maximised loglikelihood values of (9) and (11), respectively. This statistic will have a chi-square distribution with m degrees of freedom.¹

4. The demand for demand deposits: An illustration

We will illustrate the application of the J test to the heteroskedastic Tobit model with a problem in the estimation of demand for money. Previous studies based on time series estimation of the role of wealth (in a portfolio sense) and income (in a transactions sense) in the money demand schedule yield conflicting results. This may be due to the high collinearity of the aggregate measures of wealth and income available for time series estimation. In order to investigate this proposition, a cross-sectional data set is used in a heteroskedastic Tobit model. The data used are the Projector and Weiss (1966) sample of household data² collected as of December 31, 1962. The sample purposely contains a disproportionately large percentage of wealthy

¹An anonymous referee pointed out that if σ_{ji}^2 in (10) is defined as $\sigma_{ji}^2 = (\alpha_{j0} + \alpha_{j1} Z_{ji})^2$ then the variance of u_{ki}^* in the comprehensive model (11) would be $(\gamma_0 + \sum_{j=1}^m \gamma_j Z_{ji})^2$. For the purpose of this study, however, the specification of the variance we have used is appropriate. The residual variance in all models is assumed to be a function of income and wealth.

²We are grateful to our colleagues William Bomberger and David Denslow for providing us with these data.

households. Thus, all observations in which wealth exceeds \$1 million are excluded from this study. Demand deposits (*DEM*) are used as the dependent variable (cash holdings are unavailable), where approximately 25% of the households reported non-positive demand deposits. Income represents pretax income during 1962. The total sample size is 1884.

Whereas economic theory gives strong arguments for the inclusion of wealth and income, little is known about the role played by personal and occupational characteristics in the demand for money. Furthermore, there is incomplete theoretical evidence about the appropriate functional form. For this reason we will use the J test for the selection of a 'best' model. Five models were chosen in various combinations for this purpose. Consider

$$H_j: DEM_i = \alpha_0 + \alpha_1 WEALTH_i + \alpha_2 INCOME_i + \alpha_3 Z_{ji} + U_{ji} \\ = 0, \quad (13)$$

for $j = 1, \dots, 5$ and $i = 1, \dots, N$. Here, DEM_i denotes the demand deposits held by the i th household (similarly for $INCOME$, and $WEALTH$), and Z_j refers to the remaining exogenous variables chosen for the j th model. It is further assumed the $U_{ji} \sim IN(0, \sigma_{ji}^2)$, where

$$\sigma_{ji} = \beta_0 + \beta_{1j} WEALTH_i + \beta_{2j} INCOME_i, \quad (14)$$

The exogenous variables in Z_j are listed below for each model:

Model	Z_i
H_1	<i>BINHNT</i> ,
H_2	$W_i^2, Y_i^2, (W \cdot Y)_i$
H_3	<i>DOCCI, DOCC2_i, DOCC3_i, DOCC4_i</i>
H_4	<i>SINGLE,, BLACK,, SIZE,, AGE,, AGE_i²</i>
H_5	<i>AGE,, CITY_i, ED, FEMALE,</i>

W_i^2, Y_i^2 and $(W \cdot Y)_i$ are polynomial expressions for wealth and income.

Furthermore,

$$BIHNT_i = 1 \quad \text{if inheritances were a substantial portion of assets,} \\ = 0 \quad \text{if little or no assets were inherited;} \\ DOCCI_i = 1 \quad \text{if occupation of head of household, is self-} \\ \text{employment,} \\ = 0 \quad \text{otherwise;}$$

- $DOCC2_i$ = 1 if head of household, is retired,
 = 0 otherwise;
 $DOCC3_i$ = 1 if head of household, did not work,
 = 0 otherwise;
 $DOCC4_i$ = 1 if occupation of head of household, is farm worker,
 = 0 otherwise;
 $SINGLE_i$ = 1 if head of household, is not married,
 = 0 otherwise;
 $BLACK_i$ = 1 if head of household, is black,
 = 0 otherwise;
 $SIZE_i$ = number of members in household;
 AGE_i = age of head of household;
 AGE_i^2 = AGE_i squared
 $CITY_i$ = 3 if size of place, exceeds 1,000,000,
 = 2 if size of place, includes 250,000 to 1,000,000,
 = 1 if size of place, is less than 250,000,
 = 0 if household, is outside an urban area;
 ED_i = years of education of head of household;
 $FEMALE_i$ = 1 if head of household, is female,
 = 0 otherwise.

Table 1
Heteroskedastic Tobit estimates.^a

	H_1	H_2	H_3	H_4	H_5
<i>CONSTANT</i>	-125.010 (39.2390)	-48.924 (39.8278)	-114.892 (40.0977)	-154.135 (39.4066)	-17.588 (39.7652)
<i>WEALTH</i>	0.017 (0.0023)	0.017 (0.0025)	0.016 (0.0023)	0.015 (0.0023)	0.014 (0.0023)
<i>INCOME</i>	0.031 (0.0059)	0.031 (0.0059)	0.037 (0.0063)	0.045 (0.0058)	0.026 (0.0062)
a					
<i>CONSTANT</i>	245.097 (6.6078)	242256 (6.3490)	248.808 (6.6800)	250.560 (6.8564)	243.175 (6.7145)
<i>WEALTH</i>	0.030 (0.0004)	0.030 (0.0004)	0.030 (0.0004)	0.030 (0.0004)	0.030 (0.0004)
<i>INCOME</i>	0.016 (0.0008)	0.015 (0.0008)	0.015 (0.0009)	0.013 (0.0011)	0.013 (0.0012)
$L(0)$	-13423.82	-13415.97	-13418.81	-13399.15	-13383.75
$L(\hat{\lambda})$	-13371.13	-13369.61	-13364.48	-13374.80	-13343.02

^a $N = 1884$; standard errors in parentheses.

Table 1 (continued)
Personal characteristics.¹

	H_1	H_2	H_3	H_4	H_5
<i>BINHT</i>	0.148 (0.1115)				
W^2		-1.995 (15.417)			
Y^2		69354 (7.065)			
$W \cdot Y$		-27.161 (37.994)			
<i>DOCC1</i>			0.300 (0.058)		
<i>DOCC2</i>			0.1064 (0.200)		
<i>DOCC3</i>			0.105 (0.262)		
<i>DOCC4</i>			0.460 (0.172)		
<i>SINGLE</i>				-0.044 (0.086)	
<i>BLACK</i>				-1.845 (1.460)	
<i>SIZE</i>				-0.033 (0.016)	
<i>AGE</i>				-0.381 (0.125)	0.126 (0.020)
<i>AGE</i> ²				0.247 (0.060)	
<i>CITY</i>					-0.006 (0.024)
<i>ED</i>					0.0467 (0.061)
<i>FEMALE</i>					-0.193 (0.141)

¹ $N = 1884$; standard errors in parentheses

A brief justification of the decision to construct these models is necessary. H_1 suggests that wealth, income, and inheritance affect demand deposits, where no reference is made to personal or occupational characteristics. H_2 investigates non-linearities in wealth and income. The remaining models are constructed so as to isolate occupational and personal characteristics. Thus, H_3 hypothesizes that occupational characteristics are of primary importance;

H_5 , personal characteristics. H_5 is constructed to include educational and personal characteristics.

Table I presents the estimation results. For each model, the regression coefficients from (13) are given, where the numbers in parentheses are estimated standard errors. $L(0)$ represents the maximized likelihood value of (13), and similarly $L(\hat{\lambda})$ corresponds to the maximum likelihood of the embedded model shown in (11).³

The results show that wealth and income are both significant, and the hypothesis of homoskedasticity must be rejected. Furthermore, most of the personal and occupational characteristics are insignificant. The results of the J test, however, are somewhat surprising. When each model is tested against the remaining four simultaneously, twice the difference of $L(\hat{\lambda})$ and $L(0)$ will be asymptotically distributed as chi-squared if the model being tested is true. Assuming a significance level of 5% [$\chi^2_{0.05}(4)=14.86$], all models under consideration must be rejected!

Let us first look at the results of each individual model. There is convincing evidence of heteroskedasticity in the residuals. Further, both the income and wealth coefficients are not only significant but also of plausible magnitudes. In addition all variables that are included in the Z_j set that are significant have the correct signs. Thus, anyone estimating only one of these models would find the model satisfactory.

However, when it comes to the problem of choosing the 'best' model, it is a different story. The criteria of model selection used here have the ability to reject or accept all models under consideration. In applied analysis, the researcher would hope to isolate an optimal model for predictive and inferential purposes. The tests we have performed have rejected all models. This result can be interpreted in several ways. One interpretation is that none of the models considered can be regarded as being consistent with the way in which the data were generated. This, however, is not a satisfactory interpretation in view of the results mentioned earlier. A more reasonable interpretation is that none of the models can be chosen in preference to the others because each of them has some important omitted variables. In fact, the results seem to suggest that a model in which the Z_j set includes Y' , $DOCCI$, $DOCC4$, AGE, AGE' and ED may be appropriate. There are strong arguments for the inclusion of these variables (though we arrived at this conclusion after testing each of the five models against the others and looking at the results), Y^2 captures the non-linearity of the income effects. $DOCCI$ and $DOCC4$ are included on the grounds that self-employed persons and farm workers need to hold higher cash balances. AGE and AGE' are included on the grounds that persons in their middle age hold higher cash

³The heteroskedastic Tobit algorithm by Fische and Trost [which uses the Berndt et al. (1974) method] was used in the computations

balances than those at lower age (lower wealth effect) and higher age (lower transaction effect).⁴

In summary, the multiple model selection criteria are useful in choosing between different non-nested non-linear models. The paper illustrates the use of heteroskedastic Tobit models with an important application to the problem of measuring income and wealth effects in the demand for money and also how to interpret the results when the multiple model selection criteria reject all models.

⁴An alternative explanation for the fact that all models have been rejected is that they are not really very different from each other. Sawyer and Davidson and MacKinnon point out that this is possible when the models under consideration are too similar.

References

- Berndt, E.R., B.H. Hall, R.E. Hall and J.A. Hausman, 1974, Estimation and inference in non-linear structural models, *Annals of Economic and Social Measurement* 3, 653-665.
- Bomberger, W.A. and D.A. Denslow, 1980, Income, wealth, and the demand for demand deposits: Some cross-sectional results, Discussion paper (University of Florida, Gainesville, FL).
- Cox, D.R., 1961, Tests of separate families of hypotheses, in: *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol. 1 (University of California Press, Berkeley, CA) 105-123.
- Cox, D.R., 1962, Further results on test 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.
- Fishe, R.P.H., G.S. Maddala and R.P. Trost, 1979, Estimation of a heteroskedastic Tobit model. Discussion paper (University of Florida, Gainesville, FL).
- Pesaran, M.H., 1974, On the general problem of model selection, *Review of Economic Studies* 41, 153-171.
- Pesaran, M.H. and A.S. Deaton, 1978, Testing non-nested non-linear regression models, *Econometrica* 46, 677-694.
- Projector, D. and G. Weiss, 1966, Survey of financial characteristics of consumers (Board of Governors of the Federal Reserve System, Washington, DC).
- Rutemiller, H.C. and D.A. Bowers, 1968, Estimation in a heteroskedastic regression model *Journal of the American Statistical Association* 63, 522-557.
- Sawyer, K.R., 1980, The theory of econometric model selection, Doctoral dissertation (Australian National University, Canberra).
- Smith, M.A., 1982, Model selection criteria in the limited dependent variable model, Unpublished doctoral dissertation (University of Florida, Gainesville, FL).

n
e
of
.e
ie
ie
ie
st
ll
e.
ar
is
ie
le
re
se
:
ie
id
ls.
at
ie
ry
ie
ae
st,
-2
ng
is
id
ts.
ns
.re
sh