ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH

QUEEN'S UNIVERSITY

Kingston, Ontario, Canada K7L 3N6

MODEL SPECIFICATION TESTS BASED ON ARTIFICIAL LINEAR REGRESSIONS

DISCUSSION PAPER NO. 390

Russell Davidson
and
James G. MacKinnon

Department of Economics
Queen's University

May, 1980

Abstract

In this paper we develop an extremely general procedure for performing a wide variety of model specification tests by running artificial linear regressions and then using conventional significance tests. In particular, this procedure allows us to develop non-nested hypothesis tests for any set of models which attempt to explain the same dependent variable(s), even when the error specifications of the various models are not the same. For example, it is straightforward to test linear regression models against loglinear ones. These procedures are illustrated by an empirical application, in which we estimate and test several competing models of personal savings behavior in Canada.

Introduction

In this paper we develop an extremely general procedure for performing a wide variety of tests for model misspecification by running artificial linear regressions and then using asymptotic t-tests or likelihood ratio tests. Our basic result can be used to test for, among other things, both heteroskedasticity and serial correlation in the context of very general nonlinear regression models. More important, it can be used to develop non-nested hypothesis tests for any two (or more) models which purport to explain the same dependent variable (or set of dependent variables), even when the error specifications of the various models are different. For example, it is straightforward using this procedure to test a linear regression model against a loglinear one.

In Section 1 we derive the principal theoretical results of the paper. In Section 2 we use them to develop the non-nested hypothesis testing procedure alluded to above. In Section 3 we apply them to some other problems in testing model specification. Finally, in section 4, we apply some of the techniques we have proposed, especially the ones for non-nested hypothesis testing, to an empirical example.

1. The Principal Results

Consider the model

$$f_t(y_t, \bar{y}_t, \theta) = \varepsilon_t, \qquad \varepsilon_t \sim N(0,1) \tag{1}$$

where y_t is the t^{th} observation on a dependent variable, $\bar{y}_t$ is a vector of past values of y, θ is a vector of r parameters to be estimated, and $f_t(\cdot)$ is a function which may, and usually will, depend on a vector of exogenous variables X_t. It is assumed that f_t and its derivative with respect to y_t, f_t', are twice continuously differentiable with respect to θ. There are assumed to be $n > r$ observations.

The loglikelihood function for this model is:

$$L = -(n/2) \log 2\pi - (1/2) f^t f_t + \iota^t k_t \qquad (2)$$

where ι^t denotes the t^{th} component of an n-vector of ones and

$$k_t = \log |f_t'(y_t, \bar{y}_t, \theta)| \ . \qquad (3)$$

The notation used in (2) perhaps requires some introduction. We use the Einstein summation convention that repeated indices are summed over, when one is a subscript and the other is a superscript. Thus $f^t f_t$ means $\Sigma_{t=1}^{n} f_t^2$, and both f_t and f^t denote $f_t(y_t, \bar{y}_t, \theta)$. The notation "|...|" in (3) denotes absolute value.

The loglikelihood function (2) is particularly simple for two reasons. First, we have assumed that the errors have a variance of unity (because the parameter(s) determining the actual variance have been subsumed in θ). Secondly, the assumption that f_t depends only on current and past values of y means that the Jacobian matrix of the transformation from the y's to the ε's is lower triangular, so that only its diagonal elements appear in the likelihood function. Nevertheless, as we shall see, (1) is general enough to include a wide variety of econometric models as special cases.

Each of the likelihood equations for a maximum of (2) with respect to θ may be written as

$$L_\alpha = - f^t f_{t,\alpha} + \iota^t k_{t,\alpha} = 0, \tag{4}$$

where the subscript α denotes differentiation with respect to the α^{th} component of θ. Each element of the Hessian may be written as

$$L_{\alpha\beta} = - f^t_{,\beta} \; f_{t,\alpha} - f^t \; f_{t,\alpha\beta} + \iota^t \; k_{t,\alpha\beta} \tag{5}$$

where the subscript $\alpha\beta$ denotes differentiation with respect to both the α^{th} and β^{th} components of θ. From (3) one easily calculates that

$$k_{t,\alpha\beta} = (f'_{t,\alpha\beta} \,/\, f'_t) - (f'_{t,\alpha} \; f'_{t,\beta})/(f'^2_t). \tag{6}$$

The first term in (6) is simply the derivative of $f_{t,\alpha\beta}$ with respect to ε_t (since the derivative of y_t with respect to ε_t is $1/f'_t$). On the other hand, each of the terms in the middle sum in (5) is $- f_t \; f_{t,\alpha\beta}$. Thus (5) may be rewritten as

$$L_{\alpha\beta} = - f^t_{,\beta} \; f_{t,\alpha} - k^t_{,\beta} \; k_{t,\alpha} + \iota^t \; [(f'_{t,\alpha\beta} \,/\, f'_t) - f_t \; f_{t,\alpha\beta}]. \tag{7}$$

But the expectation of the last term in (7) is zero. It is an easily derived property of the standard normal distribution that, when x is standard normal and g is any continuously differentiable function, $E(xg(x)) = E(g'(x))$ if the expectations exist. In the last term in (7) f_t plays the role of x, $f_{t,\alpha\beta}$ plays the role of g(x) and, as indicated above, $(f'_{t,\alpha\beta}/f'_t)$ is then equal to g'(x). Thus we conclude that

$$-E(L_{\alpha\beta}) = E(f^t_{,\beta} \; f_{t,\alpha} + k^t_{,\beta} \; k_{t,\alpha}). \tag{8}$$

Now consider the linear regression:

$$\begin{bmatrix} \hat{f}_t \\ \iota_t \end{bmatrix} = \begin{bmatrix} -\hat{f}_{t,\alpha} \\ \hat{k}_{t,\alpha} \end{bmatrix} b^{\alpha} + \text{errors.} \qquad (9)$$

This regression has 2n observations and r regressors. For the first n observations the regressand is f_t and each of the regressors is minus the derivative of f_t with respect to one of the components of θ, both evaluated at the maximum likelihood estimates $\hat{\theta}$. For the last n observations, the regressand is unity and each of the regressors is the derivative of k_t with respect to the same component of θ, again evaluated at $\hat{\theta}$. It follows immediately from the likelihood equations (4) that the OLS estimates of b from (9) will be zero identically. Moreover, a typical element of the X^TX matrix from that regression will be

$$X^TX_{\alpha\beta} = \hat{f}^t_{,\beta}\hat{f}_{t,\alpha} + \hat{k}^t_{,\beta}\;\hat{k}_{t,\alpha} \qquad (10)$$

which is a consistent estimate of (8). Also, the ML estimate of the variance from (9) will be

$$(\hat{f}_t\hat{f}^t + n)/2n \qquad (11)$$

which clearly has a plim of unity. Thus the estimated covariance matrix of b from the linear regression (9) will provide a consistent estimate of the covariance matrix of $\hat{\theta}$, which by the standard result is asymptotically given by the inverse of (8).

Now suppose that we evaluate the regressand and regressors in (9) not at the ML estimates $\hat{\theta}$ but at $\tilde{\theta}$, which maximizes the likelihood function (2) subject to R distinct restrictions. If those restrictions are valid, that is, if they are satisfied by the true parameters θ of (1), then it can be seen (details in Appendix 1) that, asymptotically, $(\tilde{\theta}-\theta) + b$, where b denotes the OLS estimates from (9), is normally distributed with mean vector zero and covariance matrix given by the inverse of (8) and thus consistently estimated by the covariance matrix from the linear regression. It follows that restrictions on θ may be straightforwardly tested by asymptotic t-tests (if R = 1 and the restriction is of the form $\theta^{\beta} = \bar{\theta}^{\beta}$ for some β) or by Wald or likelihood-ratio tests (if $R \geq 1$), applied to regression (9). Thus regression (9) provides a way to test any set of restrictions on (1). One need only estimate (1) subject to the restrictions and then estimate (9).

Our approach is clearly similar in spirit to the "Lagrange Multiplier" principle of Aitchison and Silvey [1,2]. Such LM tests have recently been applied to econometrics by Breusch and Pagan [3,4], Godfrey [11,12] and Engle [10], among others. However, the actual test statistics we suggest will not be LM test statistics. Instead, we are following the general approach of Durbin's [9] "alternative procedure". Indeed, regression (9) can be regarded as a particular implementation of that procedure. We believe that it is usually simpler and more natural to apply Wald or LR tests to artificial regressions than it is to work out LM test statistics explicitly. In several of the papers mentioned above, (see especially [10] and [12]) the LM test procedure turns out to involve an artificial regression which is a special case of (9), and a test statistic which is asymptotically equivalent to the LR or Wald statistics based on (9).

Econometricians do not often encounter regression models in which the error variance is known to be unity *a priori*. Therefore consider the model

$$g_t(y_t,\phi) = u_t, \qquad u_t \sim N(0,\sigma^2) \tag{12}$$

where σ^2 is to be estimated. This model can easily be put into the form of (1), by defining

$$f_t(y_t,\theta) = f_t(y_t,\phi,\sigma) = (1/\sigma)g_t(y_t,\phi). \tag{13}$$

Then regression (9) becomes

$$\begin{bmatrix} \hat{g}_t/\hat{\sigma} \\ \iota_t \end{bmatrix} = \begin{bmatrix} -\hat{g}_{t,\gamma}/\hat{\sigma} \\ \hat{g}'_{t,\gamma}/\hat{g}'_t \end{bmatrix} b^{\gamma} + \begin{bmatrix} \hat{g}_t/\hat{\sigma}^2 \\ -\iota_t/\hat{\sigma} \end{bmatrix} b^{\sigma} + \text{errors} \tag{14}$$

This is the basic result for univariate regression models.

Observe that, if $g'_{t,\gamma}$ is zero for all elements γ of ϕ, we may replace (14) by

$$\hat{g}_t = -\hat{g}_{t,\gamma}\, b^{\gamma} + \hat{u}_t \tag{15}$$

It is possible to do this because when $\hat{g}'_{t,\gamma}$ is zero, the last regressor in (14) is orthogonal to all of the other regressors as a consequence of the likelihood equations (4). Since that last column (like all the others) is also orthogonal to the regressand, we may drop it without affecting the properties of (14). All the remaining regressors are zero for the last n observations, so we may drop those observations. Finally, multiplying the regressand and all regressors by $\hat{\sigma}$ yields (15).

If (15) is valid, we may write (12) as

$$\ell_t(y_t) = h_t(\phi) + u_t \tag{16}$$

for some suitable functions h_t and ℓ_t, so that (15) becomes

$$\ell_t(y_t) - \hat{h}_t = \hat{h}_{t,\gamma}\, b^{\gamma} + \hat{u}_t. \tag{17}$$

This is familiar for the case in which $\ell_t(y_t) = y_t$. It is reasonably well known that regression (17) then has the properties we have claimed for (9), and hence for (14); see in particular Durbin [9] and Davidson and MacKinnon [6]. Using those properties of (17) one may easily develop a variety of useful procedures. For example, all of the tests for serial correlation in linear regression models proposed by Godfrey [11,12] may be derived from (17), in the context of nonlinear regression models.[1] Regression (17) may also be used to calculate a consistent estimate of the covariance matrix for linear regression models with serial correlation and lagged dependent variables; conveniently so, since most regression packages provide an inconsistent estimate of this matrix (see Cooper [5]).

Regression (9) may also be used to test multivariate regression models. Consider the model

$$g_{it}(y_{it},\phi) = u_{it}, \qquad (u_{1t} \ldots u_{mt})^{T} \sim N(0,\Sigma) \tag{18}$$

where the dependent variables are indexed by i = 1 to m and the observations are indexed by t = 1 to n. We assume that g_{it} may depend additionally on exogenous variables and on lagged values of all of the endogenous variables, but that it does not depend on y_{jt} for $j > i$. The covariance matrix Σ is to be estimated. Special cases of (18) include systems of non-simultaneous equations, such as consumer demand systems, and linear simultaneous equation systems written in terms of the restricted reduced form. Nonlinear simultaneous equation systems are not special cases of (18), however, because Σ would have to depend on t if the errors originally adhered to the structural rather than reduced-form equations.

We now rewrite (18) in the form of (1). First, we define an upper-triangular m x m matrix P by the equation

$$P\,P^T = \Sigma^{-1} \tag{19}$$

and take note of the standard result that if $U \sim N(0,\Sigma)$ where U is a row vector of length m, then

$$U\,P \sim N(0,I). \tag{20}$$

Hence we may make the definition

$$f_{it}(y_{it},\theta) = P^j_i\; g_{jt}(y_{jt},\phi), \tag{21}$$

where $P^j_i = 0$ for $j > i$. If we then replace the double subscript "it" by the single subscript $s = n(i-1) + t$, it is evident that the model will have the form of (1). It will be more convenient to retain the double subscript, however. Note that the it^{th} Jacobian term in the loglikelihood function will be

$$k_{it} = \log \left| P_{ii}\; g'_{it}(y_{it},\phi)\right| \tag{22}$$

where g'_{it} denotes the derivative of g_{it} with respect to y_{it}. In order to write down regression (9) we need the first derivatives of f_{it} and k_{it}.

These are:

$$\frac{\partial f_{it}}{\partial \phi_\gamma} = P^j_{\ i}\, g_{jt,\gamma}$$

$$\frac{\partial f_{it}}{\partial P^j_{\ \ell}} = g_{jt} \quad \text{if } \ell = i$$
$$= 0 \text{ otherwise} \tag{23}$$

$$\frac{\partial k_{it}}{\partial \phi_\gamma} = \frac{g'_{it,\gamma}}{g'_{it}}$$

$$\frac{\partial k_{it}}{\partial P^j_{\ \ell}} = 1/P_{ii} \quad \text{if } i = j = \ell$$
$$= 0 \text{ otherwise.}$$

Substituting these derivatives, evaluated at the ML estimates $\hat{P}^j_{\ i}$ and $\hat{\phi}$, into regression (9) yields the result we want. This regression, which has 2mn observations and ℓ + (m+1)m/2 regressors, is rather cumbersome to write out, so we merely describe it here. The regressand is $\hat{P}^j_{\ i}\,\hat{g}_{jt}$ for the first mn observations, and unity for the last mn. The regressor corresponding to the γ^{th} component of ϕ is $-\hat{P}^j_{\ i}\,\hat{g}_{jt,\gamma}$ for the first mn observations, and $\hat{g}'_{it,\gamma}/\hat{g}'_{it}$ for the last mn. The regressor corresponding to P_{ii} (a diagonal element of P) is $-\hat{g}_{it}$ in the it^{th} place for all t, $1/\hat{P}_{ii}$ in the $(it + mn)^{th}$ place, and zero everywhere else. Finally, the regressor corresponding to P_{ij} (an off-diagonal element of P) is $-\hat{g}_{jt}$ in the it^{th} place, and zero everywhere else.

2. Non-nested Hypothesis Testing

In [6] and [8] we developed techniques for non-nested hypothesis testing in the context of univariate and multivariate models respectively.

Like earlier authors (notably Pesaran and Deaton [13]) we restricted our attention to models with the same error specification. In the univariate case, we proposed that in order to test

$$H_0: \; y_t = f_t(\beta) + u_{0t}, \qquad u_{0t} \sim N(0, \sigma_0^2)$$

against

$$H_1: \; y_t = g_t(\gamma) + u_{1t} \qquad u_{1t} \sim N(0, \sigma_1^2)$$

one can estimate the possibly nonlinear regression:

$$y_t = (1-\alpha)f_t(\beta) + \alpha\hat{g}_t + u_t \tag{24}$$

(where $\hat{g}_t \equiv g_t(\hat{\gamma})$, and $\hat{\gamma}$ denotes the ML estimate from H_1) and test whether $\alpha = 0$. Since $\hat{g}_t$ is asymptotically non-stochastic, this procedure, which we called the J-test because one estimates α and β jointly, is asymptotically valid. As an alternative to the J-test, we suggested the P-test, which requires that one estimate:

$$y_t - \hat{f}_t = \hat{F}_t b + \alpha(\hat{g}_t - \hat{f}_t) + u_t, \tag{25}$$

where $\hat{F}_t$ is a vector of derivatives of $f_t(\beta)$ evaluated at the ML estimates $\hat{\beta}$. Thus (25) is a linear approximation to (24) about the point $(\hat{\beta},0)$. This P-test procedure is simply an application of regression (17); its validity follows immediately from the validity of the J-test and from the results in Section 1.

We now wish to consider a somewhat more general case. The model to be tested is

$$H_0: \; f_t(y_t, \bar{y}_t, \beta) = u_{0t}, \qquad u_{0t} \sim N(0, \sigma_0^2) \tag{26}$$

and the alternative model is

$$H_1: \; g_t(y_t, \bar{y}_t, \gamma) = u_{1t}, \qquad u_{1t} \sim N(0, \sigma_1^2). \tag{27}$$

The notation here is the same as the notation used in Section 1. To perform the J-test in this case one would have to estimate

$$(1-\alpha) \; f_t \; (\beta) + \alpha g_t(\hat{\gamma}) = u_t \tag{28}$$

by nonlinear methods. This procedure will clearly be valid if under H_0 $\hat{\gamma}$ tends asymptotically to some probability limit γ^0, since in that case $\hat{\gamma}$ will be asymptotically non-stochastic. Now $\hat{\gamma}$ is what White [14] calls a quasi-maximum likelihood estimator or QMLE; that is, the estimator one obtains by applying maximum likelihood to a model which is false. For the case where the random variables y are independently and identically distributed, White proves under weak regularity conditions that a QMLE does indeed converge to some probability limit. We conjecture that this result carries over to the independently and not identically distributed case; a proof would be beyond the scope of this paper. For now, we merely assume that g_t and f_t satisfy whatever conditions are necessary for $\hat{\gamma}$ to be asymptotically non-stochastic.

Since the likelihood function for (28) will include the Jacobian term

$$\iota^t \log \mid (1-\alpha) \; f_t'(\beta) + \alpha g_t'(\hat{\gamma}) \mid ,$$

performing the J-test will typically be a nontrivial undertaking beyond the capability of most regression packages. As an alternative to the J-test, we therefore suggest what we shall call the L-test, which uses (14) to construct a linear regression to test for $\alpha = 0$. Straightforward application of (14) yields a regression which is unnecessarily complicated, but which can be simplified to:

$$\begin{bmatrix} \hat{f}_t \\ \hat{\sigma}_0 \iota_t \end{bmatrix} = \begin{bmatrix} -\hat{f}_{t,\beta} \\ \hat{\sigma}_0 \hat{f}'_{t,\beta}/\hat{f}'_t \end{bmatrix} b^{\beta} + \begin{bmatrix} \hat{f}_t \\ -\hat{\sigma}_0 \iota_t \end{bmatrix} b^{\sigma} + \begin{bmatrix} -\hat{g}_t \\ \hat{\sigma}_0 \hat{g}'_t/\hat{f}'_t \end{bmatrix} b^{\alpha} + \text{errors} \qquad (29)$$

In deriving (29) we have made use of the facts that adding a constant multiple of any regressor to any other regressor, or multiplying any regressor by a constant, have no effect on the fit of the regression. In this case, because all the other regressors are orthogonal to the regressand by construction, the t-test for $\alpha = 0$ is also not affected by these operations.[2]

As one very simple and concrete example of (29), suppose that the two models are, in matrix notation,

$$H_0: \; y = X\beta + u_0$$

$$H_1: \; y^* = Z\gamma + u_1 \qquad (30)$$

where y^* is a vector of the natural logarithms of y. Then regression (29), again in matrix notation, is simply:

$$\begin{bmatrix} y - X\hat{\beta} \\ \hat{\sigma}_0 \iota \end{bmatrix} = \begin{bmatrix} X \\ 0 \end{bmatrix} b + \begin{bmatrix} y - X\hat{\beta} \\ -\hat{\sigma}_0 \iota \end{bmatrix} b^{\sigma} + \begin{bmatrix} Z\hat{\gamma} - y^* \\ \hat{\sigma}_0 y^{-1} \end{bmatrix} b^{\alpha} + \text{errors} \qquad (31)$$

where y^{-1} means the vector whose t^{th} element is $1/y_t$. Thus it is absolutely straightforward to test a linear model against a loglinear one; so, of course, is it to test a loglinear model against a linear one.

There is no reason at all why a hypothesis should be tested against only one alternative hypothesis at a time. When there are several alternatives, regression (29) will have several regressors like the last one, each corresponding to a different alternative. In this case, one appropriate test is a likelihood ratio test. The value of the loglikelihood function under the null can easily be computed by regressing the dependent variable from (29) on any one or more of the regressors (except for the last one), since all are orthogonal to the regressand by construction.

At this point it is perhaps worthwhile to inject a word of caution. Although it is very easy to run regression (29) with a good modern regression package, there is always the possibility that one or more of the regressors may be constructed wrongly. For example, (29) does <u>not</u> include a constant term, and some programs may automatically include one unless told not to. If this happens, it is quite likely that the t-statistic on $\hat{\alpha}$ may be extremely large, and entirely invalid. To guard against this, we would suggest that one run (29) <u>without</u> the column corresponding to α first, to make sure that all the b's are in fact zero, and that one then take great care to ensure that the last column is constructed correctly.

In [8] we provided two different generalizations of the P-test to the case of multivariate models. The simpler of these we called the P_0-test. Suppose that the model to be tested is

$$H_0: \quad y_{it} = f_{it}(\beta) + u^0_{it}, \qquad (u^0_{1t} \;.... \; u^0_{mt})^T \sim N(0,\Sigma_0)$$

and that we wish to test it against

$$H_1: \quad y_{it} = g_{it}(\gamma) + u^1_{it}, \qquad (u^1_{1t} \; .... \; u^1_{mt})^T \sim N(0, \Sigma_1).$$

To perform the P_0-test, one merely has to compute the regression

$$P^j{}_i(y_{jt} - \hat{f}_{jt}) = P^j{}_i \; \hat{f}_{jt,\beta} b^\beta + \alpha P^j{}_i(\hat{g}_{jt} - \hat{f}_{jt}) + \varepsilon_{it} \qquad (32)$$

where P_{ji} is the ji^{th} element of an upper triangular matrix P such that $P\,P^T = \hat{\Sigma}_0^{-1}$. This is easily seen to be a straightforward generalization of the univariate P-test, regression (25).

We now consider the more general case in which regressions like (32) are no longer valid. The model to be tested is

$$H_0: \; f_{it}(y_{it},\beta) = u^0_{it}; \qquad (u^0_{it} \; .... \; u^0_{mt}) \sim N(0,\Sigma_0) \qquad (33)$$

and the alternative model is

$$H_1: \; g_{it}(y_{it},\gamma) = u^1_{it}, \qquad (u^1_{it} \; .... \; u^1_{mt}) \sim N(0,\Sigma_1). \qquad (34)$$

These models are assumed to have the same characteristics as the multivariate models discussed in section 1. One way to perform a J-test in this case would be to estimate

$$(1 - \alpha) f_{it}(\beta) + \alpha\hat{g}_{it} = u_{it} \qquad (35)$$

by highly nonlinear methods. As in the univariate case, such a procedure will be valid as long as $\hat{\gamma}$ tends to some probability limit γ_0, and we assume that this is the case.

The Jacobian term in the loglikelihood function for (35) is

$$\iota^{it} \log | P_{ii} [(1-\alpha) f'_{it}(\beta) + \alpha \hat{g}'_{it}]| . \qquad (36)$$

It is now straightforward to derive an L-test as an application of the multivariate regression described in Section 1. The L-test regression will be identical to that regression, with g replaced by f everywhere, and with one additional regressor. That regressor, which corresponds to α, will have $P^j_i(\hat{f}_{jt} - \hat{g}_{jt})$ in the first mn observations and $(\hat{g}'_{it}/\hat{f}'_{it} - 1)$ in the last mn observations. It is evident that, in the case where $\hat{f}'_{it,\beta} = 0$ for all elements of β, and $\hat{g}'_{it} \equiv \hat{f}'_{it}$, this L-test regression may validly be replaced by the P_0-test regression (32). Thus we call this L-test the L_0-test.

One word of warning is in order. If $\hat{f}_{it}$ and $\hat{g}_{it}$ are very different in magnitude, truncation error in the calculation of $\hat{g}'_{it} - \hat{f}'_{it}$ and $\hat{f}_{it} - \hat{g}_{it}$ may be severe, and may cause almost perfect collinearity. It is therefore important to scale f and g similarly. This is less of a problem in the univariate case, because it was there possible to simplify the regression so that terms like $\hat{f}_t - \hat{g}_t$ did not appear.

3. Some Nested Hypothesis Tests

The results of Section 1 can be used to derive a large number of tests for model specification which do not rely on the presence of non-nested alternative models. Here we illustrate a few of these.

First of all, it is evident that all of Godfrey's [11,12] tests for AR, MA and ARMA errors in the context of linear regression models, and of course Durbin's [9] "alternative" procedure for testing for AR(1) errors, are applications of regression (17). Similar tests can be derived from regression (14) for models which have a non-zero Jacobian term in the loglikelihood function. We illustrate this for the case where the null hypothesis has no serial correlation, and the alternative hypothesis is that the errors follow an AR(1) process. We write the null hypothesis as

$$H_0:\ f_t(y_t,\bar{y}_t,\beta) = u_t, \qquad u_t \sim N(0,\sigma^2). \tag{37}$$

The alternative hypothesis is that

$$u_t = \rho u_{t-1} + e_t, \qquad e_t \sim N(0, \sigma_e^2).$$

This implies that the general model can be written as

$$H_1:\ (1-\rho^2)^{\frac{1}{2}} f_1\ (y_1,\bar{y}_1,\beta) = e_1$$

$$f_t(y_t,\bar{y}_t,\beta) - \rho f_{t-1}(y_{t-1},\bar{y}_{t-1},\beta) = e_t, \quad t=2 \ldots n \tag{38}$$

Unlike most authors, we do not drop the initial observation at this stage, and we do explicitly assume that the error process is stable. As we shall see, these assumptions are inconsequential.

We now calculate the derivatives of (38) and of the log of its Jacobian with respect to β and ρ, and evaluate them at $\beta = \hat{\beta}$, $\rho = 0$ and $\sigma = \hat{\sigma}$. The testing regression can then be derived immediately from (14). After we multiply all variables by $\hat{\sigma}$ or $\hat{\sigma}^2$ to simplify things slightly, we obtain the following regression:

$$\begin{bmatrix} \hat{f}_t \\ \hat{\sigma}\iota_t \end{bmatrix} = \begin{bmatrix} -\hat{f}_{t,\beta} \\ \hat{\sigma}\ \hat{f}'_{t,\beta}/\hat{f}'_t \end{bmatrix} b^{\beta} + \begin{bmatrix} \hat{f}_t \\ -\hat{\sigma}\iota_t \end{bmatrix} b^{\sigma} + \begin{bmatrix} \hat{f}_{t-1} \\ 0 \end{bmatrix} b^{\rho} + \text{errors} \tag{39}$$

where $\hat{f}_0$ is defined to be zero. It is interesting to observe that, although we did not explicitly ignore the first observation, it turns out to play no role at all, because for observation 1 the derivative of (38) with respect to ρ, evaluated at $\rho = 0$, is zero.

The results of Section 1 are particularly useful for deriving tests for heteroskedasticity. Suppose that the null hypothesis is

$$H_0: \; y_t = f_t(\beta) + u_t, \qquad u_t \sim N(0,\sigma^2) \tag{40}$$

and the alternative is

$$H_1: \; y_t = f_t(\beta) + u_t, \qquad u_t \sim N(0,\sigma_t^2)$$

$$\sigma_t = h(Z_t,\alpha) \tag{41}$$

where α is a vector of parameters which can be partitioned into α_0 (a scalar) and α_1 to α_m. Under the null hypothesis, $\alpha_1 = \ldots \alpha_m = 0$, so that $\sigma_t = h(\alpha_0) = \sigma$. This formulation is similar to that of Breusch and Pagan [3], but they assume that $h(Z_t,\alpha)$ can be written as $h(\alpha_0 + Z_{ti}\alpha^i)$, which is much more restrictive.

In this case the testing regression turns out to be

$$\begin{bmatrix} y_t - \hat{f}_t \\ \\ \hat{\sigma}\iota_t \end{bmatrix} = \begin{bmatrix} \hat{f}_{t,\beta} \\ \\ 0 \end{bmatrix} b^{\beta} + \begin{bmatrix} \hat{h}_{t,\alpha}(y_t - \hat{f}_t) \\ \\ -\hat{\sigma}\,\hat{h}_{t,\alpha} \end{bmatrix} a^{\alpha} + \text{errors.} \tag{42}$$

One then uses a likelihood ratio or Wald test for the hypothesis that a^1 through a^m (but of course not a^0) are zero.

Another alternative hypothesis which it might be interesting to investigate is that

$$\sigma_t = \sigma + \alpha f_t(\beta) \tag{43}$$

since this includes both homoskedastic errors and errors whose variance is proportional to the square of the dependent variable as special cases. The testing regression for this alternative is

$$\begin{bmatrix} y_t - \hat{f}_t \\ \hat{\sigma}\iota_t \end{bmatrix} = \begin{bmatrix} \hat{f}_{t,\beta} \\ 0 \end{bmatrix} b^{\beta} + \begin{bmatrix} y_t - \hat{f}_t \\ -\hat{\sigma}\iota_t \end{bmatrix} b^{\sigma} + \begin{bmatrix} \hat{f}_t(y_t - \hat{f}_t) \\ -\hat{\sigma}\hat{f}_t \end{bmatrix} a^{\alpha} + \text{errors} \tag{44}$$

These procedures merely illustrate a few of the situations in which (14) can fruitfully be applied. The reader should have no trouble developing new tests based on (14) for a variety of other problems.

The results of Section 1 may also be used to derive a battery of tests for multivariate models. However, it is difficult to develop interesting and useful tests in this context without specifying a lot more about the structure of the model than we have done so far, and it seems best to derive such tests in the context of practical applications.

An Application

In this section we provide an illustration of how some of the procedures developed above, in particular the L-test for non-nested models with different error specifications, may fruitfully be used in applied econometrics. The objective is to estimate a single-equation model of aggregate personal savings in Canada using quarterly time series data, not seasonally adjusted, for 1954 to 1978. The data used are described in Appendix 2. One of the models we estimate has previously been estimated by us in [7], using American as well as Canadian data, but in that paper it was tested only against models with the same error specification.

We initially estimate a rather naive model, which asserts that the level of real savings, S_t, should depend linearly on real personal disposable income, Y_t, the rate of inflation, π_t, and lagged savings, S_{t-1}. We obtain

Model 1

$$S_t = -1352 + .2721\, Y_t + 2840\, \pi_t + .2174\, S_{t-1} + \text{others}$$
$$\quad\ (329) \quad (.0503) \qquad (3197) \qquad (.0932)$$

$\log L = -640.862 \quad \bar{R}^2 = .9664 \quad AR1 = 0.33 \quad AR4 = 0.06$

Standard errors are in parentheses. The numbers reported as AR1 and AR4 are test statistics for AR(1) and simple AR(4) errors, obtained by re-estimating the model with residuals lagged one or four periods as additional regressors and taking the t-statistics on those additional regressors. These tests are elementary applications of (17) above; see also Durbin [9]. The notation "+ others" indicates that Model 1 includes additional regressors, the coefficients of which are not reported. In fact, there are nine seasonal dummy variables (three straight dummies, three dummies multiplied by an annually-increasing linear time trend and three dummies multiplied by an annually-increasing quadratic trend, each set of dummies being constrained to sum to zero over the year), and two trend terms (linear and quadratic, increasing quarterly). The seasonal dummies were necessary to eliminate evidence of seasonal variation in the residuals,[3] and the trends were included to pick up the effects of long-term changes in demographic structure and tax incentives to save (which have been important in Canada, but are difficult to model explicitly). All of the models we consider include these eleven additional explanatory variables.

On the surface, Model 1 may seem quite satisfactory. The adjusted R^2 is high, there is no evidence of serial correlation, all coefficients have the right signs, and only the coefficient on π_t is insignificantly different from zero (which would not be disturbing to many investigators because the role of inflation in savings functions is not yet well established). Nevertheless, further investigation will reveal that Model 1 is thoroughly false.

First of all, it seems a little unreasonable that errors which adhere to the level of real savings should have a constant variance. There are at least two plausible alternatives to the assumption that $\sigma_t = \sigma$. One is that $\sigma_t = \sigma + \alpha Y_t$, and the other is that $\sigma_t = \sigma + \alpha f_t$, where f_t denotes the fitted value of S_t from Model 1. The procedures developed in Section 3 above may be used to test against either of these alternatives; the test statistics are 3.00 and 3.85 respectively. The Lagrange Multiplier test of Breusch and Pagan [3] may also be used to test against the former alternative; the square root of their χ^2 test statistic (which should be N(0,1) under the null hypothesis) is 3.09. Hence we may certainly conclude that the error specification of Model 1 is wrong.

One alternative to Model 1 is an equally naive model in which the dependent variable is the savings rate, $s_t = S_t/Y_t$, so that Y_t does not appear on the right hand side at all. We obtain:

Model 2

$$s_t = \underset{(.00901)}{.03213} + \underset{(.2373)}{.7409}\,\pi_t + \underset{(.0924)}{.3246}\,s_{t-1} + \text{others}$$

$\log L = -635.990$ $\bar{R}^2 = .9286$ AR1 = -2.66 AR4 = 1.16

According to the likelihood function, Model 2 fits rather better than Model 1 (the likelihood for Model 2 is of course computed for S rather than s as the dependent variable; this involves subtracting $\Sigma \log Y_t$ = 938.800 from the ordinary loglikelihood value printed by the regression package). However, there is now evidence of first-order serial correlation, so that the error specification of Model 2 is also apparently wrong.

Models 1 and 2 may be tested against each other by means of the L-test. The test statistics are 4.56 and 2.60 when Models 1 and 2 are tested in turn. Thus both models are clearly false, with Model 1 being rather more soundly rejected, as one would expect on the basis of its lower likelihood.

A somewhat different approach would be to make consumption rather than savings the dependent variable. Since it is just as implausible that homoskedastic errors should adhere to the level of consumption as to the level of savings, it seems natural to specify the model in logarithms. A very traditional consumption function, which ignores inflation effects entirely, is:

<u>Model 3</u>

$$\log C_t = \underset{(.4834)}{1.5843} + \underset{(.0582)}{.6729} \log Y_t + \underset{(.0670)}{.1378} \log C_{t-1} + \text{others}$$

$$\log L = -634.526 \quad \bar{R}^2 = .9987 \quad AR1 = 2.11 \quad AR4 = 0.27$$

This model fits slightly better than Models 1 or 2, according to the likelihood function (which once again is computed in terms of S rather than C; in this case the Jacobian term is $-\Sigma \log C_t = -932.340$). However, it appears to suffer from first-order serial correlation.

Model 3 may be tested against Models 1 and 2 by means of the L-test; the test statistics are 2.11 and 3.20 respectively, so that Model 3 is surely false. When the tests are reversed, Models 1 and 2 are in turn rejected by Model 3; the test statistics are 4.46 and 3.56. Because the Jacobian of the transformation from S_t or s_t to $\log C_t = \log (Y_t - S_t)$ is negative, the column corresponding to α in the test regression (29) was multiplied by minus one for all of these L-tests. Otherwise, the test statistics would have been negative, giving the (probably) erroneous impression for all of these tests that the truth lies away from the direction of the alternative model, rather than towards it.

We now estimate a model which may not be false. In [7] we argued that, in the absence of inflation, a reasonable model of savings behavior is:

$$S_t = b_0 Y_t + b_1 (S_{t-1} - Y_{t-1}) + u_t \qquad (45)$$

where the standard deviation of u_t is proportional to Y_t. This model emerges if one postulates that savings are proportional to income in the long run, that the flow of consumption is subject to a first-order partial adjustment process, and that errors adhere to the savings rate. In times of inflation, however, measured income and measured savings overstate the true amounts perceived by consumers because of the loss of real value of financial assets due to inflation. Thus real income and savings should be adjusted downwards to take account of this. We estimate the perceived loss on financial assets due to inflation as a weighted average on current and past inflation rates times an estimate of the real value of financial assets, and call this variable Z; for details, see Appendix 2, and also [7]. Since not all financial assets lose value in terms of inflation (e.g., many common stocks), and since not everyone may correctly perceive how inflation affects the income from these assets, we allow the amount by which savings and income are reduced to depend on a parameter α which is expected to lie between zero and one. Thus (45) becomes

$$S_t = b_0 Y_t + (1 - b_0)\alpha Z_t + b_1(S_{t-1} - Y_{t-1}) + u_t \qquad (46)$$

Dividing all regressors by Y_t to eliminate heteroskedasticity, we obtain

<u>Model 4</u>

$$s_t = \underset{(.0461)}{.2243} + \underset{(.0585)}{.3893}\, Z_t/Y_t + \underset{(.0481)}{.1884}\, (S_{t-1} - Y_{t-1})/Y_t + \text{others}$$

log L = -619.231 $\bar{R}^2$ = .9489   AR1 = -0.84   AR4 = 0.82

Model 4 evidently fits much better than any of the first three models. It is thus inevitable that when they are tested against it, all are rejected; the test statistics for Models 1, 2 and 3 are 6.73, 5.84 and 5.77 respectively (the first and last of these are L-tests, the second, a P-test, since the error specifications of Models 2 and 4 are the same). On the other hand, none of the three earlier models rejects Model 4; the test statistics are 1.43, -0.32 and 0.32 respectively. Model 4 also displays no evidence of serial correlation or of any obvious form of heteroskedasticity. For example, when the hypothesis that $\sigma_t = \sigma$ is tested against the alternative that $\sigma_t = \sigma + \alpha f_t$, where f_t denotes the fitted value of s_t from Model 4, the test statistic is -0.35.

Of course, Model 4 is not the only model which works well. Alternatively, one could start from Model 2 and modify it to take account of the overmeasurement of income due to inflation. That would mean replacing log

Y_t by log $(Y_t - \alpha Z_t)$. Nonlinear estimation can be avoided by noting that a first-order Taylor series approximation to log $(Y_t - \alpha Z_t)$ around $\alpha = 0$ is log $Y_t - \alpha Z_t/Y_t$, which is surely a valid approximation since Z_t is much smaller than Y_t. We thus obtain:

Model 5

$$\log C_t = \underset{(.5055)}{.1235} + \underset{(.0567)}{.8052} \log Y_t - \underset{(.0765)}{.4023}\, Z_t/Y_t + \underset{(.0590)}{.1762} \log C_{t-1} + \text{others}$$

$$\log L = -620.439 \quad \bar{R}^2 = .9990 \quad AR1 = -0.53 \quad AR4 = 0.38$$

When Model 5 is tested against Model 4 by an L-test the test statistic is 1.82, and when Model 4 is tested against Model 5 it is -0.39. Thus neither of these models can reasonably be rejected, although there would seem to be slightly more evidence in favour of Model 4.

So far we have only used the L-test to test models against a single alternative at a time. Let us now see what happens when we test the two best models, numbers 4 and 5, against the other three models of the set {1, 2, 4, 5}; Model 3 is excluded because it is just a special case of Model 5. The test statistics for Models 4 and 5 respectively are 7.39 and 10.02; these are simply likelihood ratio test statistics, which should be asymptotically distributed as Chi-squared with three degrees of freedom under the null hypothesis. Since the .05 critical value is 7.81, we conclude that Model 4 cannot be rejected, but that Model 5 should be. These results illustrate the possibility, which we have stressed in [6], that non-nested hypothesis tests against several different alternatives at once may prove more powerful than several tests against the same alternatives singly.

Conclusion

In this paper we have proposed a very general way of linearizing nonlinear regressions about given specifications of the parameters. This allows us to develop a wide variety of test procedures, all based on linear regressions and standard test statistics produced by conventional regression packages. These procedures yield easily implemented tests of model specification in a great many cases of econometric interest. In particular, it is possible to test non-nested hypotheses against each other even when the error specifications are different, as in the case of linear versus logarithmic models. The empirical application of Section 4 shows that these non-nested tests have substantial ability to reject false hypotheses, and that they can yield a good deal of information with comparatively little effort. These tests should prove extremely useful in many areas of applied econometrics.

Appendix 1

We have to establish the result that when (1) is estimated subject to R distinct restrictions to yield parameter estimates $\tilde{\theta}$, then asymptotically $(\tilde{\theta} - \theta) + b$ is normally distributed with mean vector 0 and covariance matrix $[-E(L_{\alpha\beta})]^{-1}$. (b denotes the OLS estimates from (9)). First, we notice that we may reparametrise the model so that the restrictions take the form $\theta^a = 0$ for a = 1, ... R, and for convenience let the true values of the other r-R parameters also be zero. We shall use Latin indices to denote the R restricted parameters and Greek indices for the others, so that the summation convention will yield restricted sums when these are needed.

The estimates $\tilde{\theta}$ are defined by the r-R equations:

$$L_\alpha (0, \tilde{\theta}) = 0,$$

so that, asymptotically,

$$\tilde{\theta}^\alpha \sim - H^{\alpha\beta} L_\beta \qquad \text{(A1)}$$

where the (r-R) x (r-R) matrix $H^{\alpha\beta}$ is the inverse of the block $L_{\alpha\beta}$ of the Hessian of L. Both $H^{\alpha\beta}$ and L_β are evaluated here at (0,0). Now the OLS estimates from (9) can, in view of the properties of (9) established in the text, be written asymptotically as

$$b^a \sim - L^{ac} L_c(0,\tilde{\theta})$$

$$\sim - L^{ac} L_c - \tilde{\theta}^\beta L_{\beta c} L^{ac}$$

and $$b^\alpha \sim - L^{\alpha c} L_c (0,\tilde{\theta})$$

$$\sim - L^{\alpha c} L_c - \tilde{\theta}^\beta L_{c\beta} L^{\alpha c} \qquad \text{(A2)}$$

Here Taylor expansions have been used, and $L^{\alpha\beta}$, $L^{\alpha a}$, L^{ac} denote elements of the inverse of the full Hessian of L. Use of (A1) and (A2) gives

$$(0^a, \tilde{\theta}^\alpha) + (b^a, b^\alpha)$$

$$\sim(- L^{ac} (L_c - H^{\beta\gamma} L_\gamma L_{\beta c}), - H^{\alpha\beta} L_\beta - L^{\alpha c} (L_c - H^{\beta\gamma} L_\gamma L_{c\beta})).$$

It is easy to establish the following two results:

$$H^{\beta\gamma} L_{\beta c} L^{ca} = - L^{\gamma a}$$

$$H^{\beta\gamma} L_{\beta c} L^{c\alpha} = H^{\gamma\alpha} - L^{\gamma\alpha},$$

whence

$$(0^a, \tilde{\theta}^\alpha) + (b^a, b^\alpha)$$

$$\sim - (L^{ac} L_c + L^{a\alpha} L_\alpha, \ L^{\alpha c} L_c + L^{\alpha\beta} L_\beta),$$

or, expressed in conventional matrix and vector notation,

$$(\tilde{\theta}-\theta) + b \sim - (D^2L)^{-1} (DL) \qquad \text{(A3)}$$

where D^2L is the Hessian and DL the gradient of L. Our assertion follows immediately from (A3) by standard arguments.

Appendix 2

In this Appendix we describe the data used in this study. All data were not seasonally adjusted, and were taken from the CANSIM database, as of November 1979.

CN = personal consumption expenditure in current dollars, CANSIM # D 40043.

CR = personal consumption expenditure in constant 1971 dollars, CANSIM # D 40562.

P = CN/CR.

YN = personal disposable income in current dollars, CANSIM # D 40057.

Y = YN/P.

SN = personal savings excluding change in farm inventories, CANSIM # D 40055.

S = SN/P.

IN = interest, dividends and miscellaneous investment income, CANSIM # D 40036.

I = IN/P.

r = quarterly averages of the McLeod, Young, Weir 40 bond yield average, CANSIM # B 14031 (monthly), divided by 400.

$\pi = \log P_t - \log P_{t-1}$.

$\pi^* = .2\pi_t + .3\pi_{t-1} + .3\pi_{t-2} + .2\pi_{t-3}$.

$Z = \pi^* I/r$.

TREND = 1 in 1950-1, increasing by 1 each quarter.

Sample Period: 1954-1 to 1978-4.

Footnotes:

1. The regressions Godfrey proposes are precisely those one would derive from (17), where h_t is the regression model transformed to eliminate serial correlation (and ignoring the initial observations), which is then evaluated at the ML estimates conditional on serial correlation being absent. The test statistic Godfrey suggests is n times the R^2 from this regression. Since the R^2 would be zero if the restrictions were precisely true, one can easily show that nR^2 is asymptotically equivalent to the usual likelihood ratio test statistic.

2. Regression (28) is not the only plausible way to formulate a J-test in this case. Another alternative would be to estimate

$$(1-\alpha)\ f_t\ (\beta)/\sigma_0 + \alpha g_t(\hat{\gamma})/\hat{\sigma}_1 = \varepsilon_t.$$

However, it turns out that this formulation also leads to (29) as the L-test regression.

3. An alternative approach to modelling seasonality would be to assume that the error terms follow an AR or ARMA process in powers of four. Several experiments with this approach produced models which fit rather badly. For example, when the six trending seasonal dummies in Model 4 below were dropped, and a simple AR(4) error process was introduced instead (using pre-sample data to obtain lagged errors for the first four observations), the value of the loglikelihood function fell by approximately 20 points. This was so despite the fact that the estimates looked quite satisfactory (the estimate of the AR(4) parameter was 0.49 with a t-statistic of 5.82), and that there was no evidence at all of further fourth-order serial correlation. In view of these unsatisfactory

results, including a large number of seasonal dummy variables would seem to be a preferable approach to modelling seasonality, at least for this problem.

References

[1] Aitchison, J., and S.D. Silvey: "Maximum-likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29 (1958), 813-828.

[2] __________: "Maximum-likelihood Estimation and Associated Tests of Significance," *Journal of the Royal Statistical Society*, Series B, 22 (1960), 154-171.

[3] Breusch, T.S., and A.R. Pagan: "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47 (1979), 1287-1294.

[4] __________: "The Lagrange Multiplier Test and its Applications to Model Specification," *Review of Economic Studies*, 47 (1980).

[5] Cooper, J.P.: "Asymptotic Covariance Matrix of Procedures for Linear Regression in the Presence of First-Order Autoregressive Disturbances," *Econometrica*, 40 (1972), 305-310.

[6] Davidson, R., and J.G. MacKinnon: "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, forthcoming.

[7] __________: "Inflation and the Savings Rate," Queen's University Institute for Economic Research Discussion Paper No. 351, 1979.

[8] __________: "Relations Among Some Non-nested Hypothesis Tests," Queen's University Institute for Economic Research Discussion Paper No. 369, 1980.

[9] Durbin, J.: "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38 (1970), 410-421.

[10] Engle, R.F.: "A General Approach to the Construction of Model Diagnostics Based Upon the Lagrange Multiplier Principle," Warwick Economic Research Papers No. 156, 1979.

[11] Godfrey, L.G.: "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables," *Econometrica,* 46 (1978), 1293-1301.

[12] __________: "Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables," *Econometrica,* 46 (1978), 1303-1310.

[13] Pesaran, M.H., and A.S. Deaton: "Testing Non-Nested Nonlinear Regression Models," *Econometrica,* 46 (1978), 677-694.

[14] White, H.: "Maximum Likelihood Estimation of Misspecified Models: I," Department of Economics, University of Rochester, Discussion Paper No. 80-1, 1980.