Econometrica, Vol. 46, No. 3 (May, 1978)

TESTING NON-NESTED NONLINEAR REGRESSION MODELS[1]

BY M. H. PESARAN AND A. S. DEATON

In Pesaran [**9**], the test developed by Cox for comparing separate families of hypotheses was applied to the choice between two non-nested linear single-equation econometric models. In this paper, the analysis is extended to cover multivariate nonlinear models whenever full information maximum likelihood estimation is possible. This allows formal comparisons not only of competing explanatory variables but also of alternative functional forms. The largest part of the paper derives the results and shows that they are recognizable as generalizations of the single-equation case. It is also shown that the calculation of the test statistic involves very little computation beyond that necessary to estimate the models in the first place. The paper concludes with a practical application of the test to the analysis of the U.S. consumption function and it is demonstrated that formal tests can give quite different results to conventional informal selection procedures. Indeed, in the case examined, five alternative hypotheses, some of which appear to perform quite satisfactorily, can all be rejected using the test.

1. INTRODUCTION

THE NEED FOR STATISTICAL PROCEDURES for testing separate families of hypotheses has become more acute with the increased use of econometric techniques in practice. The usual F tests can only be applied to test nested hypotheses, i.e. those which are members of the same family. However, in practice, one is frequently faced with the problem of testing non-nested hypotheses.

In an earlier article, Pesaran [**9**] applied the test developed by Cox [**3**, **4**], for separate families of hypotheses to single-equation linear regression models both with autocorrelated and nonautocorrelated disturbances. In that paper, the question was confined to the selection of appropriate explanators for a given dependent variable. However, in much applied work, the investigator is required not merely to select variables but simultaneously to find an appropriate functional form. This problem can be especially acute since in many areas of research, economic theory can guide us in the choice of variables, but helps very little in the choice of functional form. As computing capacity has increased, and nonlinear estimation has become routine, the use of linearity has become more a matter of choice than of necessity; the criteria for such a choice are thus of considerable practical importance. In this paper, we extend the earlier analysis to cover these problems by deriving the comparable statistics without assuming linearity of the models. This allows formal comparisons of different explanatory variables, of different functional forms, and of the interactions between the two. We also extend the results to cover competing systems of nonlinear equations whenever full-information maximum-likelihood estimation is possible. This allows the test to be applied to non-nested simultaneous equation models as well

[1] An earlier version of this paper was presented to the European meeting of the Econometric Society, Helsinki, September, 1976.

as to systems of regression equations such as are frequently encountered in demand analysis or in investment and employment studies. Finally, we present some illustrative calculations of how the test can be used in practical situations involving nonlinear models.

We believe that, in the final analysis, the usefulness and importance of this class of tests can only be established by practical experience. However, there has in recent years been a continuing debate on the appropriate methodology for non-nested testing; see, for example, the papers by Atkinson [**1**], Quandt [**10**], and Gaver and Geisel [**6**]. Consequently, it is desirable that we make our position clear at the outset. This can best be done by a clear statement of what we believe to be involved in the use of the Cox test and the grounds for our belief that alternative procedures are unsatisfactory by comparison.

We are faced with a body of data and a set of alternative hypotheses. Since the latter are non-nested by assumption, we cannot rank them by level of generality as can be done when the models are nested. There is thus no maintained hypothesis; each model is on an equal footing with every other model. To follow Cox's procedure we take the alternatives one at a time, assuming each one in turn to be true and inferring from the behavior of the alternatives against the data whether or not our temporarily maintained or working hypothesis can or cannot explain what we then observe. We thus make pairwise tests of each pair of hypotheses and we ask the question, is the performance of H_j against the data consistent with the truth of H_i? By making such tests, we are using the hypotheses in the same way that one usually uses the data; for example, the formulation of a previously unconsidered hypothesis can lead to new inferences about existing models just as would the discovery of new data. This highlights a basic feature of empirical methodology, that hypotheses are responsible for organizing data in order to yield meaningful information and that, without such organization, observations are meaningless, if not impossible. We thus consider that not only are procedures such as the Cox test necessary to make comparisons between hypotheses, but that the ability to make meaningful inferences about the truth of any single hypothesis demands the presence of at least one non-nested alternative. In econometrics, we never have a maintained hypothesis which we believe with certainty; we must always use the models we possess to organize the evidence in different ways and to ask whether the patterns which result are consistent with the views we currently hold.

It is important that notions of the absolute fit or performance of individual models play no part in the analysis. Indeed, it should be clear from the previous discussion that, apart from the nested case, we regard such indicators as meaningless. In considering whether an alternative hypothesis, together with the data, contains sufficient information to reject the currently maintained hypothesis, the question of whether that alternative "fits" well or badly, even if meaningful, is certainly irrelevant. An hypothesis, which one would not wish to consider seriously in its own right, can be a perfectly effective tool for disproving an alternative, even if that alternative may in some respects seem much more promising. It is thus important that tests between non-nested hypotheses or

models should encompass the possibility of rejecting both, as does the Cox procedure. This is notably not the case for tests which compare relative fits, for example comparisons of R^2 statistics or likelihoods. Nor is it true of Bayesian procedures which, in the absence of discriminatory prior information, reduce to comparisons of likelihoods. Since the Bayesian approach assumes that the models under consideration exhaust all the possibilities, we are led to select that model which, when prior and sample evidence are combined, does least violence to the facts. This is quite reasonable as the solution of a statistical decision problem but we do not find it convincing as a general basis for statistical inference in applied economics. And even if one accepts this framework, there appear to be quite serious problems in the formulation of satisfactory priors in a number of important cases; see the discussion in Gaver and Geisel [**6**, 69–72].

The other possible approach to non-nested hypothesis testing consists of embedding the alternatives in a general combined model against which the original alternatives may be compared using standard techniques. Variants of this methodology have been suggested by Atkinson [**1**] and, more recently, by Quandt [**10**]. The original comments by Pesaran [**9**, p. 155] still seem pertinent: there is a degree of arbitrariness in the way a comprehensive model is constructed; the redefinition of the problem poses a quite different question from the original one and will only yield answers to the original problem in special cases; and on practical grounds, collinearity between variables will often prevent satisfactory estimation of the general model at all. This last point is likely to be even more serious in discriminating among functional forms containing the same variables, and when a large number of non-nested alternatives is being considered, a comprehensive model is likely to be so general as to be useless in practice. In our view, there are very few economic hypotheses, if any, that we really are prepared to maintain, so that it is a great advantage of the Cox procedure that we do not have to do so. The construction of artificial composite hypotheses to which we are forced to become committed, without possibility of test, only avoids the problems of inference with which we are concerned.

Many statisticians would hold that statistical inference can only reject hypotheses *in favor of a well-defined alternative*; see for example the persuasive arguments put forward by Hacking [**7**]. From this point of view, one should look with suspicion upon any test which allows even the possibility of rejecting all hypotheses under consideration. While it is clear that in many practical situations it is necessary to select a particular hypothesis as a basis for further action, such a choice does not necessarily imply a belief that the chosen model is correct. In natural sciences, at least historically, theories have tended to be rejected only in the face of strong evidence in favor of a particular alternative. In economics, where firmly established models are much less frequent, this has been less true. To take an example, the resurgence of monetarism over the last decade has convinced many economists that the naive Keynesian models are no longer tenable. But few have been wholeheartedly converted to monetarism; many would assert that the insights gained from the Keynesian view of macroeconomics are still sufficient to cast severe doubt on many of the monetarist

positions. Consequently, many economists find themselves believing neither Keynesian nor monetarism; each contains enough to invalidate the other. We believe that in economics, at least, there is much to commend a statistical test which gives formal recognition to the possibility of ignorance of this kind.

2. THE DERIVATION OF THE STATISTIC

The discussion in this section is carried out in terms of the most general case, when the competing hypotheses are each systems of nonlinear regression equations. In order to make the exposition more transparent, we shall give the single-equation results separately; this also enables a fairly direct comparison to be made with the single-equation linear results derived in Pesaran [**9**].

The two competing models, H_0 and H_1, will be written in the form

$$(1) \qquad H_0: \qquad y_{ti} = f_i(\theta_0; x_t) + u_{ti0},$$

$$(2) \qquad H_1: \qquad y_{ti} = g_i(\theta_1; z_t) + u_{ti1}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$.

In each case we have n equations (index i) defined over T observations (index t). y_{ti} is thus the tth observation on the ith dependent variable. The functions $f_i(\cdot)$ and $g_i(\cdot)$ are continuous and second order differentiable with respect to all their arguments; $f_i(\cdot)$ is not nested within $g_i()$ nor *vice versa*. θ_0 and θ_1 are vectors of parameters of length k_0 and k_1 respectively. Note that these vectors are not indexed by i. There are many important cases where individual parameters appear in several different equations; note that our notation does *not* imply that every parameter appears in every equation. x_t and z_t are vectors of predetermined variables. The expressions u_{ti0} and u_{ti1} are random disturbances, for each t independently and identically multinormally distributed with means zero and covariance matrices Ω_0 and Ω_1 respectively. A wide class of models can be written in this form; for example, we can regard (1) and (2) as the reduced forms of systems of simultaneous linear equations where the θ parameters are the *structural* rather than reduced-form coefficients.

The systems (1) and (2) may be written in vector form as

$$(3) \qquad H_0: \qquad y = f(\theta_0; X) + u_0,$$

$$(4) \qquad H_1: \qquad y = g(\theta_1; Z) + u_1,$$

where y is the $nT \times 1$ vector of observations on all the n dependent variables, $f(\cdot)$ and $g(\cdot)$ are the corresponding $nT \times 1$ vectors of predictions, u_0 and u_1 of errors, and X and Z are matrices of predetermined variables. We denote the complete parameter sets of each model by the vectors α_0 and α_1 so that

$$\alpha_0' = \{\theta_0', \mathrm{st}(\Omega_0)'\} \quad \text{and} \quad \alpha_1' = \{\theta_1', \mathrm{st}(\Omega_1)'\}$$

where st (Ω_0) and st (Ω_1) are the vectors formed by stacking the matrices Ω_0 and Ω_1 by columns. Denoting the log likelihood functions of H_0 and H_1 by $L_0(\alpha_0)$ and $L_1(\alpha_1)$, respectively, and by $\hat{L}_{10}$ the log of the maximum likelihood ratio, then

if we maintain H_0 against H_1, the Cox statistic is given by

$$T_0 = \hat{L}_{10} - T\left[\operatorname*{plim}_{T\to\infty}{}_0 \left(\frac{\hat{L}_{10}}{T}\right)\right]_{\alpha_0=\hat{\alpha}_0} \tag{5}$$

where plim_0 denotes the probability limit when H_0 is true, $\hat{L}_{10} = L_0(\hat{\alpha}_0) - L_1(\hat{\alpha}_1)$, and $\hat{\alpha}_0$ and $\hat{\alpha}_1$ denote the maximum likelihood estimators of α_0 and α_1 under H_0 and H_1, respectively. Given that H_0 is true, Cox [**3**] shows that T_0 is asymptotically normally distributed with mean zero and variance $V(T_0)$. Defining $L_{10} = L_0(\alpha_0) - L_1(\alpha_{10})$, where $\alpha_{10} = \operatorname{plim}_0 \hat{\alpha}_1$, then

$$V_0(T_0) = V_0(L_{10}) - \frac{1}{T}\eta' Q^{-1}\eta, \tag{6}$$

where Q is the asymptotic information matrix of H_0, i.e.

$$Q = -\operatorname{plim}_0 \frac{1}{T}\frac{\partial^2 L_0}{\partial\alpha_0 \partial\alpha_0'} \tag{7}$$

and

$$\eta = T\frac{\partial[\operatorname{plim}_0(\hat{L}_{10}/T)]}{\partial\alpha_0}. \tag{8}$$

Our main task is to derive expressions for (5) and (6) with H_0 and H_1 given by (3) and (4). In order to do so, we shall first make the following assumptions:

(i) u_0 and u_1 are distributed as multivariate normal with mean zero and covariance matrices $\Omega_0 \otimes I$ and $\Omega_1 \otimes I$, respectively, where $\otimes$ denotes the Kronecker product and I is a $T \times T$ identity matrix. Ω_0 and Ω_1 are assumed to be nonsingular.

(ii) The number of observations, T, is at least as large as the number of equations, n, so that the maximum likelihood estimators of Ω_0 and Ω_1 are also nonsingular. (In practice T may have to be very much larger than n.)

(iii) Either the x_t and z_t variables are nonstochastic, or we require that the functions $f(\theta_0; X)$ and $g(\theta_1; Z)$, as well as their derivatives, are distributed independently of the disturbances u_0 and u_1.

(iv) We require that the following limits exist and be finite:

$$\operatorname*{plim}_{T\to\infty}{}_0 \left[\frac{1}{T}\left(\frac{\partial f(\theta_0)}{\partial\theta_0}\right)'(\Omega_0^{-1}\otimes I)\left(\frac{\partial f(\theta_0)}{\partial\theta_0}\right)\right] = \Sigma_{00},$$

$$\operatorname*{plim}_{T\to\infty} \left[\frac{1}{T}\left(\frac{\partial g(\theta_1)}{\partial\theta_1}\right)'(\Omega_1^{-1}\otimes I)\left(\frac{\partial g(\theta_1)}{\partial\theta_1}\right)\right] = \Sigma_{11}.$$

The matrices Σ_{00} and Σ_{11} are nonsingular, i.e., θ_0 is asymptotically identified under H_0 as is θ_1 under H_1.

(v) H_0 and H_1 are both non-nested and non-orthogonal.

(vi) The regularity conditions on the likelihood functions are satisfied so that we can write

$$\operatorname{plim}_0 \hat{\alpha}_0 = \alpha_0 \quad \text{and} \quad \operatorname{plim}_1 \hat{\alpha}_1 = \alpha_1.$$

We shall adopt the following notation. Variables are subscripted by numerical and literal subscripts; the former relate to the hypotheses H_0 or H_1, the latter to vector or matrix elements. Thus ω_{ij0} is the i, jth element of the matrix Ω_0; the literal subscripts always precede the numerical ones. The numerical subscript 10 refers to an asymptotic expectation of a parameter of H_1 given that H_0 is true. Superscripts denote elements of matrix inverses, e.g., ω_0^{ij} is the i, jth element of the inverse of Ω_0. Superimposed hats denote maximum likelihood estimates. We shall also use $y_{\cdot i}$ to denote the vector of T observations on the ith dependent variable; similarly for $f_{\cdot i}(\theta_0)$ and $g_{\cdot i}(\theta_0)$.

Given this notation and our assumptions, the log-likelihood functions for H_0 and H_1 may be written

$$L_0(\alpha_0) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log|\Omega_0| - \tfrac{1}{2}\{y - f(\theta_0)\}'(\Omega_0^{-1}\otimes I)\{y - f(\theta_0)\}, \tag{9}$$

$$L_1(\alpha_1) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log|\Omega_1| - \tfrac{1}{2}\{y - g(\theta_1)\}'(\Omega_1^{-1}\otimes I)\{y - g(\theta_1)\}, \tag{10}$$

where, for brevity, we have suppressed the X and Z arguments of f and g. As is well known, the maximum likelihood estimates of Ω_0 and Ω_1, $\hat{\Omega}_0$ and $\hat{\Omega}_1$, are given by

$$\hat{\omega}_{ij0} = \frac{1}{T}\{y_{\cdot i} - f_{\cdot i}(\hat{\theta}_0)\}'\{y_{\cdot j} - f_{\cdot j}(\hat{\theta}_0)\}, \tag{11}$$

$$\hat{\omega}_{ij1} = \frac{1}{T}\{y_{\cdot 1} - g_{\cdot i}(\hat{\theta}_1)\}'\{y_{\cdot j} - f_{\cdot j}(\hat{\theta}_1)\}, \tag{12}$$

where $\hat{\omega}_{ij0}$ and $\hat{\omega}_{ij1}$ denote the (i, j)th elements of matrices $\hat{\Omega}_0$ and $\hat{\Omega}_1$, respectively. The estimates $\hat{\theta}_0$ and $\hat{\theta}_1$, used in (11) and (12), are the maximum likelihood estimates of θ_0 and θ_1 which are derived by solving the equations

$$\left(\frac{\partial f(\hat{\theta}_0)}{\partial \theta_0}\right)'(\hat{\Omega}_0^{-1}\otimes I)\{y - f(\hat{\theta}_0)\} = 0, \tag{13}$$

$$\left(\frac{\partial g(\hat{\theta}_1)}{\partial \theta_1}\right)'(\hat{\Omega}_1^{-1}\otimes I)\{y - g(\hat{\theta}_1)\} = 0. \tag{14}$$

Equations (11)–(14) together define $\hat{\Omega}_0$, $\hat{\Omega}_1$, $\hat{\theta}_0$, and $\hat{\theta}_1$. Substituting (11) and (12) in (9) and (10), we have

$$\hat{L}_{10} \equiv L_0(\hat{\alpha}_0) - L_1(\hat{\alpha}_1) = \frac{T}{2}\log\left\{\frac{|\hat{\Omega}_1|}{|\hat{\Omega}_0|}\right\}. \tag{15}$$

Thus, from (5),

$$T_0 = \frac{T}{2}\log\left\{\frac{|\hat{\Omega}_1|}{|\hat{\Omega}_{10}|}\right\}, \tag{16}$$

where $\hat{\Omega}_{10}$ is a maximum likelihood estimator of $\Omega_{10} = \operatorname{plim}_0 \hat{\Omega}_1$. In order to

obtain estimates of Ω_{10} and θ_{10}, we follow Cox and solve for α_{10} the asymptotic equation

$$\text{(17)} \qquad \operatorname{plim}_0\left(\frac{1}{T}\frac{\partial L_1(\alpha_{10})}{\partial \alpha_{10}}\right)=0.$$

Clearly, from (10), we have

$$\text{(18)} \qquad L_1(\alpha_{10})=-\frac{nT}{2}\log(2\pi)-\frac{T}{2}\log|\Omega_{10}|-\tfrac{1}{2}\{y-g(\theta_{10})\}'(\Omega_{10}^{-1}\otimes I)\times \{y-g(\theta_{10})\}.$$

Differentiating with respect to a typical element of Ω_{10}^{-1},

$$\frac{\partial L_1(\alpha_{10})}{\partial \omega_{10}^{ij}}=\frac{T}{2}\omega_{ij10}-\tfrac{1}{2}\{y_{\cdot i}-g_{\cdot i}(\theta_{10})\}'\{y_{\cdot j}-g_{\cdot j}(\theta_{10})\},$$

which, given that H_0 is true, becomes, from (1),

$$\frac{\partial L_1(\alpha_{10})}{\partial \omega_{10}^{ij}}=\frac{T}{2}\omega_{ij10}-\tfrac{1}{2}\{f_{\cdot i}(\theta_0)+u_{\cdot i0}-g_{\cdot i}(\theta_{10})\}'\{f_{\cdot j}(\theta_0)+u_{\cdot j0}-g_{\cdot j}(\theta_{10})\}.$$

Hence in view of assumptions (i) to (iii)

$$\operatorname{plim}_0\left\{\frac{1}{T}\frac{\partial L_1(\alpha_{10})}{\partial \omega_{10}^{ij}}\right\}=\tfrac{1}{2}\omega_{ij10}-\frac{1}{2T}\{f_{\cdot i}(\theta_0)-g_{\cdot i}(\theta_{10})\}'\{f_{\cdot j}(\theta_0)-g_{\cdot j}(\theta_{10})\} -\tfrac{1}{2}\operatorname{plim}_0\left(\frac{1}{T}u'_{\cdot i0}u_{\cdot j0}\right).$$

Equating to zero, and rearranging the terms, we have an estimate of Ω_{10},

$$\text{(19)} \qquad \hat{\omega}_{ij10}=\hat{\omega}_{ij0}+\frac{1}{T}\{f_{\cdot i}(\hat{\theta}_0)-g_{\cdot i}(\hat{\theta}_{10})\}'\{f_{\cdot j}(\hat{\theta}_0)-g_{\cdot j}(\hat{\theta}_{10})\}.$$

Similarly, differentiating (18) with respect to θ_{10} we get

$$\frac{\partial L_1(\alpha_{10})}{\partial \theta_{10}}=\left(\frac{\partial g(\theta_{10})}{\partial \theta_{10}}\right)'(\Omega_{10}^{-1}\otimes I)\{y-g(\theta_{10})\}.$$

Taking probability limits gives, if H_0 is true,

$$\operatorname{plim}_0\left(\frac{1}{T}\frac{\partial L_1(\alpha_{10})}{\partial \theta_{10}}\right)=\frac{1}{T}\left(\frac{\partial g(\theta_{10})}{\partial \theta_{10}}\right)'(\Omega_{10}^{-1}\otimes I)\{f(\theta_0)-g(\theta_{10})\}.$$

If we equate this to zero, we get

$$\text{(20)} \qquad \left(\frac{\partial g(\theta_{10})}{\partial \theta_{10}}\right)'(\Omega_{10}^{-1}\otimes I)\{f(\hat{\theta}_0)-g(\hat{\theta}_{10})\}=0.$$

Equations (19) and (20) may be solved together for $\hat{\theta}_{10}$ and $\hat{\Omega}_{10}$ in terms of $\hat{\theta}_0$ and $\hat{\Omega}_0$.

The steps for the calculation of T_0 can now be clearly seen and are the obvious analogue of the calculation in the single equation linear case. At the first step, H_0

and H_1 are estimated by full information maximum likelihood and estimates of $\hat{\theta}_0$, $\hat{\theta}_1$, $\hat{\Omega}_0$, and $\hat{\Omega}_1$ are calculated in the usual way, i.e. according to formulae (13), (14), (11) and (12). We then take the predicted values from H_0, $f(\hat{\theta}_0)$, and use these as dependent variables in a second FIML regression of H_1. The estimated variance-covariance matrix of this regression is then added to $\hat{\Omega}_0$ to derive $\hat{\Omega}_{10}$ according to (19). T_0 is then immediately derived from (16). We thus require only one additional nonlinear regression, and numerical procedures are much eased in that the parameters $\hat{\theta}_1$ make excellent starting values for the estimation of $\hat{\theta}_{10}$.

If we have only one equation, rather than n, the key expressions can easily be simplified in the following manner. The matrices $\hat{\Omega}_1$ and $\hat{\Omega}_{10}$ are replaced by scalars $\hat{\sigma}_1^2$ and $\hat{\sigma}_{10}^2$ and we have

$$(16') \qquad T_0 = \frac{T}{2}\log\left\{\frac{\hat{\sigma}_1^2}{\hat{\sigma}_{10}^2}\right\}.$$

Corresponding to (11)–(14) and (19)–(20) we have

$$(11') \qquad \hat{\sigma}_0^2 = \frac{1}{T}\{y - f(\hat{\theta}_0)\}'\{y - f(\hat{\theta}_0)\},$$

$$(12') \qquad \hat{\sigma}_1^2 = \frac{1}{T}\{y - g(\hat{\theta}_1)\}'\{y - g(\hat{\theta})\},$$

$$(13') \qquad \left(\frac{\partial f(\hat{\theta}_0)}{\partial \theta_0}\right)'\{y - f(\hat{\theta}_0)\} = 0,$$

$$(14') \qquad \left(\frac{\partial g(\hat{\theta}_1)}{\partial \theta_1}\right)'\{y - g(\hat{\theta}_1)\} = 0,$$

$$(19') \qquad \hat{\sigma}_{10}^2 = \hat{\sigma}_0^2 + \frac{1}{T}\{f(\hat{\theta}_0) - g(\hat{\theta}_{10})\}'\{f(\hat{\theta}_0) - g(\hat{\theta}_{10})\},$$

$$(20') \qquad \frac{\partial g(\hat{\theta}_{10})'}{\partial \theta_{10}}\{f(\hat{\theta}_0) - g(\hat{\theta}_{10})\} = 0.$$

These may immediately be recognized as the nonlinear equivalents of the linear single equation results of Pesaran [**9**].

Derivation of $\hat{V}_0(T_0)$

We first calculate the variance of $L_{10} = L_0(\alpha_0) - L_1(\alpha_{10})$. From (9) and (18)

$$L_{10} = \frac{T}{2}\log\left\{\frac{|\Omega_{10}|}{|\Omega_0|}\right\} + \tfrac{1}{2}\{f((\theta_0) - g(\theta_{10})\}'(\Omega_{10}^{-1} \otimes I)\{f(\theta_0) - g(\theta_{10})\}$$
$$- \tfrac{1}{2}u_0'\{(\Omega_0^{-1} - \Omega_{10}^{-1}) \otimes I\}u_0 + \{f(\theta_0) - g(\theta_{10})\}'(\Omega_{10}^{-1} \otimes I)u_0.$$

Since u_0 is distributed as multivariate normal, it is a relatively simple exercise to calculate the variance of L_{10}. Some manipulation yields

$$V(L_{10}) = \{f(\theta_0) - g(\theta_{10})\}'(\Omega_{10}^{-1}\Omega_0\Omega_{10}^{-1} \otimes I)\{f(\theta_0) - g(\theta_{10})\} + \frac{T}{2}\operatorname{trace}[\{(\Omega_0^{-1} - \Omega_{10}^{-1})\Omega_0\}^2]. \tag{21}$$

In order to compute $V_0(T_0)$ we also need to derive η; see (8). Using (15) and (19), we have

$$\eta = \frac{T}{2}\frac{\partial}{\partial \alpha_0}\{\log|\Omega_{10}| - \log|\Omega_0|\}, \tag{22}$$

where $\alpha_0' = (\theta_0', \operatorname{st}(\Omega_0)')$, and

$$\omega_{ij10} = \omega_{ij0} + \frac{1}{T}\{f_{\cdot i}(\theta_0) - g_{\cdot i}(\theta_{10})\}'\{f_{\cdot j}(\theta_0) - g_{\cdot j}(\theta_{10})\}. \tag{23}$$

Differentiating with respect to θ_0,

$$\frac{\partial}{\partial \theta_0}\{\log|\Omega_{10}| - \log|\Omega_0|\} = \frac{\partial \log|\Omega_{10}|}{\partial \theta_0} = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{10}^{ij}\frac{\partial \omega_{ij10}}{\partial \theta_0}.$$

Substituting from (23) we obtain

$$\frac{\partial \log|\Omega_{10}|}{\partial \theta_0} = \frac{2}{T}\left\{\left(\frac{\partial f(\theta_0)}{\partial \theta_0}\right)' - \left(\frac{\partial \theta_{10}}{\partial \theta_0}\right)'\left(\frac{\partial g(\theta_{10})}{\partial \theta_{10}}\right)'\right\} \times (\Omega_{10}^{-1} \otimes \mathrm{I})\{\mathrm{f}(\theta_0) - \mathrm{g}(\theta_{10})\}.$$

But, by virtue of (17) and especially (20),

$$\left(\frac{\partial g(\theta_{10})}{\partial \theta_{10}}\right)'(\Omega_{10}^{-1} \otimes I)\{f(\theta_0) - g(\theta_{10})\} = 0.$$

Hence

$$\frac{\partial \log|\Omega_{10}|}{\partial \theta_0} = \frac{2}{T}\left\{\frac{\partial f(\theta_0)}{\partial \theta_0}\right\}'(\Omega_{10}^{-1} + I)\{f(\theta_0) - g(\theta_{10})\}. \tag{24}$$

The derivatives in (22) with respect to Ω_0 can be evaluated as follows:

$$\frac{\partial}{\partial \omega_{ij0}}\{\log|\Omega_{10}| - \log|\Omega_0|\} = \sum_{r=1}^{n}\sum_{s=1}^{n}\frac{\partial \log|\Omega_{10}|}{\partial \omega_{rs10}}\frac{\partial \omega_{rs10}}{\partial \omega_{ij0}} - \frac{\partial \log|\Omega_0|}{\partial \omega_{ij0}}.$$

But from (23) $\partial \omega_{rs10}/\partial \omega_{ij0} = \delta_{ri}\delta_{sj}$, so that

$$\frac{\partial}{\partial \omega_{ij0}}\{\log|\Omega_{10}| - \log|\Omega_0|\} = \omega_{10}^{ij} - \omega_0^{ij},$$

i.e.,

$$\frac{\partial}{\partial \operatorname{st}(\Omega_0)}\{\log|\Omega_{10}| - \log|\Omega_0|\} = \operatorname{st}(\Omega_{10}^{-1} - \Omega_0^{-1}). \tag{25}$$

Substituting (24) and (25) in (22) we reach

$$\eta = \begin{bmatrix} \left(\dfrac{\partial f(\theta_0)}{\partial \theta_0}\right)' (\Omega_{10}^{-1} \otimes I)\{f(\theta_0) - g(\theta_{10})\} \\ \dfrac{T}{2} \operatorname{st}(\Omega_{10}^{-1} - \Omega_0^{-1}) \end{bmatrix}. \tag{26}$$

The derivation of the information matrix Q (see (7)) is straightforward; again we consider derivatives with respect to θ_0 and Ω_0, respectively. Differentiation of L_0 twice yields

$$\begin{aligned} \operatorname{plim}_0 \left\{ -\frac{1}{T} \frac{\partial^2 L_0(\alpha_0)}{\partial \theta_0 \partial \theta_0'} \right\} &= \operatorname{plim}_0 \left\{ \frac{1}{T} \left(\frac{\partial f(\theta_0)}{\partial \theta_0}\right)' (\Omega_0^{-1} \otimes I) \left(\frac{\partial f(\theta_0)}{\partial \theta_0}\right) \right\} \\ &= \Sigma_{00}, \end{aligned} \tag{27}$$

which, by assumption (iv), is nonsingular. We also have

$$\operatorname{plim}_0 \left\{ -\frac{1}{T} \frac{\partial^2 L_0(\alpha_0)}{\partial \omega_{ij0} \partial \theta_0} \right\} = 0 \qquad (i, j = 1, \ldots, n), \tag{28}$$

and

$$\operatorname{plim}_0 \left\{ -\frac{1}{T} \frac{\partial^2 L_0(\alpha_0)}{\partial \omega_{ij0} \partial \omega_{rs0}} \right\} = \tfrac{1}{2} \omega_0^{ir} \omega_0^{js} \qquad (i, j, r, s = 1, \ldots, n). \tag{29}$$

Combining relations (27) to (29), the asymptotic information matrix of H_0 becomes

$$Q = \begin{pmatrix} \Sigma_{00} & 0 \\ 0 & \tfrac{1}{2}(\Omega_0^{-1} \otimes \Omega_0^{-1}) \end{pmatrix}. \tag{30}$$

Consequently, from (26) and (30),

$$\eta' Q^{-1} \eta = R' \Sigma_{00}^{-1} R + \frac{T^2}{2} \{\operatorname{st}(\Omega_{10}^{-1} - \Omega_0^{-1})\}' (\Omega_0 \otimes \Omega_0) \{\operatorname{st}(\Omega_{10}^{-1} - \Omega_0^{-1})\}, \tag{31}$$

where

$$R = \left(\frac{\partial f(\theta_0)}{\partial \theta_0}\right)' (\Omega_{10}^{-1} \otimes I)\{f(\theta_0) - g(\theta_{10})\}.$$

The second expression on the right hand side of (31) is equal to $\frac{1}{2}T^2$ trace $[\{(\Omega_0^{-1} - \Omega_{10}^{-1})\Omega_0\}^2]$, so that combining (31) and (21) according to (6) gives

$$\begin{aligned} V_0(T_0) = \{f(\theta_0) - g(\theta_{10})\}' \{&\Omega_{10}^{-1} \Omega_0 \Omega_{10}^{-1} \otimes I \\ &- (\Omega_{10}^{-1} \otimes I) \frac{1}{T} \frac{\partial f(\theta_0)}{\partial \theta_0} \Sigma_{00}^{-1} \frac{\partial f(\theta_0)'}{\partial \theta_0} (\Omega_{10}^{-1} \otimes I)\} \{f(\theta_0) - g(\theta_{10})\}. \end{aligned} \tag{32}$$

This is the basic result for the variance; $\hat{V}_0(T_0)$ is derived by replacing each of the expressions by consistent, i.e. maximum-likelihood, estimates.

To simplify expression (32), define

$$h = (\Omega_0 \Omega_{10}^{-1} \otimes I)\{f(\theta_0) - g(\theta_{10})\}$$

so that

$$V_0(T_0) = h'\{(\Omega_0^{-1} \otimes I) - (\Omega_0^{-1} \otimes I)\frac{1}{T}\frac{\partial f(\theta_0)}{\partial \theta_0}\Sigma_{00}^{-1}\frac{\partial f(\theta_0)'}{\partial \theta_0}\hat{\Phi}(\Omega_0^{-1} \otimes I)\}h$$

and its maximum likelihood estimate can be simply written as

$$\hat{V}_0(T_0) = d'\hat{\Phi}_0^{-1}d, \tag{33}$$

where

$$d = \{I - \hat{F}(\hat{F}'\hat{\Phi}_0^{-1}\hat{F})^{-1}\hat{F}'\hat{\Phi}_0^{-1}\}\hat{h},$$

$$\hat{F} = \left(\frac{\partial f(\theta_0)}{\partial \theta_0}\right)_{\theta_0 = \hat{\theta}_0},$$

$$\hat{h} = (\hat{\Omega}_0 \hat{\Omega}_{10}^{-1} \otimes I)\{f(\hat{\theta}_0) - g(\hat{\theta}_{10})\},$$

$$\hat{\Phi}_0 = \hat{\Omega}_0 \otimes I.$$

If we define $N_0 = T_0/\sqrt{\hat{V}_0(T_0)}$, we know from Cox's results that, given the truth of H_0, N_0 is asymptotically distributed as $N(0, 1)$. Similarly, we may compute T_1, $\hat{V}_1(T_1)$, and N_1 when H_1 is assumed to be correct.

The form (33) is particularly convenient for seeing clearly how the estimate of the variance may be calculated. Note that $\hat{h}$ is the residual vector of the regression of $f(\hat{\theta}_0)$ on H_1, transformed by premultiplication by $(\hat{\Omega}_0 \hat{\Omega}_{10}^{-1} \otimes I)$. Using $\hat{h}$ and the (known) covariance matrix $\hat{\Phi}_0$, we compute a generalized least squares regression of $\hat{h}$ on $\hat{F}$ yielding residuals d. The variance is then given by (33) immediately. In other words, as in the single equation linear case, the estimator of $V_0(T_0)$ can be evaluated in a straightforward manner by performing one additional linear regression.

In the case of only one nonlinear equation the result given in (32) and (33) can be written in the simpler form

$$\hat{V}_0(T_0) = \frac{\hat{\sigma}_0^2}{\hat{\sigma}_{10}^4}\{f(\hat{\theta}_0) - g(\hat{\theta}_{10})\}'\{I - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'\}\{f(\hat{\theta}_0) - g(\hat{\theta}_{10})\}$$

where, as before, $\hat{F}$ is the derivative $\partial f(\theta_0)/\partial \theta_0$ evaluated at $\theta_0 = \hat{\theta}_0$. Clearly then, apart from the scalar $\hat{\sigma}_0^2/\hat{\sigma}_{10}^4$, $\hat{V}(T_0)$ is calculated by regressing on $\hat{F}$ the residuals from the regression of $f(\hat{\theta}_0)$ on H_1 and calculating the residual sum of squares. This is the direct analogue of the linear case. We can thus see that the computational problems involved are not severe given that H_0 and H_1 have to be estimated in any case. In addition to the original nonlinear regressions, in order to calculate both T_0 and T_1 and their variances, we need only compute two extra nonlinear regressions (for both of which we have excellent starting values) and two linear regressions. The artificial variables required for the latter are in most cases calculated automatically by the numerical algorithms used for the

nonlinear estimation. There are thus no computational grounds to prevent widespread use of the test.

3. AN APPLICATION

In this section we give an example of how the tests described above may be used in practice. The problem we have chosen is the analysis of the relationship between consumption and income using U.S. quarterly data. The equations which have been estimated should not be regarded as a serious contribution to the analysis of the consumption function; rather, we have deliberately kept the analysis as simple as possible in order to illustrate how the test can be applied in practice. It should also be emphasized that these examples are not a substitute for simulation experiments with the technique; these we hope to present at a later date. Inevitably then, our experiments can only illustrate a few of the potential applications.

In any given study, the applied econometrician has a wide range of tests, formal and informal, which can be used to help choose among alternative specifications and functional forms. The tests which we are proposing are in no way a replacement for current practices; they are an addition to them, and, we believe, a valuable addition. Consequently, in the analysis below, we shall proceed very much as an applied econometrician might, bringing in, as we go along, the additional information provided by the tests.

The data we use are quarterly, seasonally adjusted, observations on real, 1958 price, consumers' expenditure on non-durable goods and on personal disposable income. These were collected from the *Survey of Current Business* [**12**] and are presented in Appendix, Table A. Observations from 1954, second quarter, to 1974, third quarter were used (82 observations in all) although the additional 29 observations on real personal disposable income from the first quarter of 1947 were used in the construction of lags.

We shall consider a variety of models, embodying alternative functional forms and alternative specifications of the lag structure between income and expenditure.

The first, and simplest, model postulates a linear relationship between consumption, income, and wealth; the influence of lagged income is thus indirect, and operates entirely through the wealth term. Denoting consumption by c, income by y, and wealth by w, we postulate

$$H_1: \qquad c = \alpha_1 + \beta_1 y + \gamma_1 w + u, \tag{34}$$

where u is normally distributed as $N(0, \sigma^2)$. For simplicity, we follow the practice of Stone [**11**] in defining wealth as the accumulated value of real saving, i.e.,

$$w - Bw = B(y - c), \tag{35}$$

where B denotes the backward shift operator. We may thus construct a series for w around some base point (the variable was taken to be zero in 1954 II) and

absorb into the intercept the (constant) value of w in that quarter. Estimation gives the following results:

$$(36)\qquad \underset{(9.766)}{c = 26.510} + \underset{(0.03592)}{0.84960y} + \underset{(.0057178)}{.0084700w}, \qquad R^2 = 0.997959 \quad \text{and} \quad \hat{\sigma}^2 = 17.3915.$$

The figures in brackets are *asymptotic* standard errors; $\hat{\sigma}^2$ is also the asymptotic, i.e. maximum likelihood, estimate of σ^2.

The wealth term in this equation is not very well determined and, whatever its validity on other grounds, it seems very unlikely to be capturing the full effects of lagged incomes. An obvious alternative is to estimate an equation containing lagged consumption as an explanator. We thus have a partial adjustment model of consumption of the type first estimated by Brown [**2**]; this may also be thought of as a natural variant of Duesenberry's [**5**] relative income hypothesis. We thus have

$$(37)\qquad H_2: \qquad c = \alpha_2 + \beta_2 y + \gamma_2 Bc + u,$$

where again u is distributed as $N(0, \sigma^2)$. Note that this last assumption rules out interpreting (37) as the Koyck transformation of an equation linking consumption to a declining geometric lag function of income; a variant of this latter hypothesis will be considered below.

Estimation of (37) yields

$$(38)\qquad \underset{(1.7888)}{c = 5.6294} + \underset{(0.07794)}{0.33838y} + \underset{(0.08673)}{0.62827Bc}, \qquad R^2 = 0.998722, \quad \hat{\sigma}^2 = 10.8887.$$

This yields a long run marginal propensity to consume of 0.91 which is in line with the usual estimates. In other respects, the results are more satisfactory than those for H_1. On conventional informal grounds, although H_1 and H_2 are non-nested, we should thus expect the N test to lead us to reject H_1 in favor of H_2. Taking H_1 as the maintained hypothesis first, we fit the predicted values of (36) to equation (37) and calculate $\hat{\sigma}^2_{21}$ to be 17.8116, which, with 82 observations, gives a T value, conditional on the validity of H_1, of -20.1772. The variance of this statistic is estimated at 0.1837 so that the N ratio, which is asymptotically distributed as $N(0, 1)$ under H_1, takes a value of -47.08. Clearly, we cannot maintain the validity of H_1 on this evidence. We may now reverse the procedure and take H_2 as the maintained hypothesis. This, by a similar sequence of calculations, leads to an N ratio of 0.37, i.e., to the conclusion that H_2 cannot be rejected against the evidence of the data and H_1 combined. Thus in this simple example, the N test supports the judgement that would be made informally in practice, that H_1 should be rejected in favor of H_2.

We may now turn from the choice of variables to the choice of functional form. The linearity of (37) was not chosen on any strong theoretical grounds but rather for convenience. We shall thus examine an alternative possibility, that the

relationship is multiplicative. This can be derived from long-run proportionality between consumption and income, coupled with an adjustment procedure formulated in terms of ratios rather than differences. We thus write

$$(38) \qquad H_3: \qquad c = e^{\alpha_3} y^{\beta_3} (Bc)^{\gamma_3} + u,$$

with $u \sim N(0, \sigma^2)$. Note that the constant of proportionality is written as an exponential; this guarantees that c be positive, and by narrowing the range of search for α_3, improves rapidity of convergence of the numerical algorithm. The parameter estimates for H_3 are

$$(39) \qquad \underset{(0.02506)}{\alpha_3 = 0.075693}, \quad \underset{(0.08351)}{\beta_3 = 0.38351}, \quad \underset{(0.08591)}{\gamma_3 = 0.60029}, \quad R^2 = 0.998756, \quad \hat{\sigma} = 10.6016.$$

The long-run *elasticity* of consumption to income is thus estimated to be 0.96 which is comparable with the long-run *marginal propensity* to consume of 0.91 estimated from H_2. Note that H_3 is very similar to H_2: the fit is marginally better but hardly enough to suggest any strong preference for one rather than the other. Applying the N test, if H_2 is the maintained hypothesis we calculate a value of -3.38 against H_3, while if H_3 is maintained against H_2, the value is 2.68. Each of these values is significantly different from zero and, on the face of it, we should reject both H_2 and H_3. The negative value when H_2 is maintained suggests that the true model deviates from H_2 in the direction of H_3, presumably 'beyond' it given the positive and significant ratio when H_2 is maintained. Note however that we are using an asymptotic statistic in a small sample situation and these two ratios are not so large as to justify confident rejection of both models. Consequently, it would not be difficult to defend a slight preference for H_3, again giving a formal result in support of our informal supposition.

It is also possible to consider H_1 against H_3, and by doing so, we can draw up Table I for all possible pairs. Each row relates to a particular maintained hypothesis while each column relates to the alternative. We have filled in the matrix by listing along the diagonals the values of $\hat{\sigma}^2$ for each model to give an idea of absolute fit as well as of comparative performance.

TABLE I

N-STATISTICS AND $\hat{\sigma}^2$ FOR H_1, H_2, AND H_3

Alternative hypothesis:		H_1	H_2	H_3
Maintained hypothesis:	H_1	17.3915	−47.08	−29.30
	H_2	0.37	10.8887	−3.38
	H_3	1.08	2.68	10.6016

Note from the table that testing H_1 against H_3 and *vice versa* gives very similar results to testing H_1 and H_2 as a pair; this reflects further the close correspondence of H_2 and H_3.

So far, the formal results have simply supported our intuition. This is satisfactory enough; indeed we should have been justified in being suspicious of

our formal test if it had not been so. However, we now consider two more complex formulations of lag structures and we shall see that intuition is often neither sufficiently precise nor sufficiently well-informed. The first more complex hypothesis is that, from the second period onwards, the weight of lagged income in the consumption function is subject to a declining geometric progression. Hence

$$(40) \qquad H_4: \qquad c = \alpha_4 + \beta_4 y + \frac{\gamma_4 By}{1-\delta_4 B} + u,$$

where division by the expression involving the lag operator is short-hand for the infinite series of weights. As written, (40) is highly nonlinear in the parameter δ_4 but it can be directly estimated using the values of y which are available prior to the sample period. This yields estimates:

$$(41) \qquad \underset{(17.715)}{\alpha_4 = 31.668}, \quad \underset{(0.045096)}{\beta_4 = 0.75293}, \quad \underset{(.0022297)}{\gamma_4 = .0031090}, \quad \underset{(.022041)}{\delta_4 = .98528},$$
$$R^2 = 0.998332, \qquad \hat{\sigma}^2 = 14.2151.$$

These results should be handled with some care. The lag parameter, δ_4, is sufficiently close to unity to cast doubt on the estimation procedure since the omitted terms in the lag will have sizable weights, even with 29 additional observations preceding the estimation period on y. The method could be improved by appropriate treatment of the truncation remainder (for example as suggested by Pesaran [**8**]) but in any case it is clear that the third term on the right-hand side of (40) is effectively acting as a large constant; note also the large standard error of α_4 and the small value of γ_4. Clearly then, whatever is the true lag structure on y, it cannot adequately be represented by H_4.

The final model we shall consider is the hypothesis that the lag structure can be approximated by a second-order Almon polynomial, i.e.,

$$(9) \qquad H_5: \qquad c = \alpha_5 + \beta_5 y + \sum_{i-1}^{20} (\gamma_5 + \delta_5 i + \varepsilon_5 i^2) B^{i+1} y + u,$$

and we have taken a lag stretching back altogether 21 periods, i.e., $5\frac{1}{4}$ years. The parameter estimates are

$$(10) \qquad \underset{(2.291)}{\alpha_5 = 11.868}, \quad \underset{(0.06809)}{\beta_5 = 0.95274}, \quad \underset{(.030001)}{\gamma_5 = -0.030902},$$
$$\underset{(.0074149)}{\delta_5 = .0073125}, \quad \underset{(.00034966)}{\varepsilon_5 = -.00032576}, \quad R^2 = 0.997933,$$
$$\hat{\sigma}^2 = 17.6131.$$

(The sum of lags is $-.043423$, so the long-run marginal propensity to consume is again estimated to be 0.91.)

In (10) none of the Almon lag parameters is significantly different from zero and it is evident that this formulation is no better at capturing the shape of the lag than was the geometric formulation. Indeed H_5 has an absolute fit which is worse even than the original model H_1.

However, once H_5 is brought into the comparisons, the N tests yield a rather more complex picture. Table II reproduces the results of Table I and completes all the pairwise comparisons.

TABLE II

N-STATISTICS β AND $\hat{\sigma}^2$ FOR H_1 TO H_5

Alternative hypothesis:		H_1	H_2	H_3	H_4	H_5
Maintained hypothesis:	H_1	17.39	−47.08	−29.30	−28.30	−6.42
	H_2	0.37	10.89	−3.38	−2.58	−6.03
	H_3	1.08	2.68	10.60	−1.86	−5.84
	H_4	2.19	−11.20	−12.09	14.22	−6.21
	H_5	−14.89	−73.04	−39.66	−121.3	17.61

Taking the comparison between H_4 and H_5 first, we see that each model rejects the other. The extremely poor fit of H_5 is reflected in the N ratios in the final row; maintaining H_5 against any alternative generates a large negative ratio. H_4 is not greatly superior; with the exception of H_1, which is itself rejected against all alternatives (see row 1), all the hypotheses provide evidence against the truth of H_4. More surprisingly, H_5, which is clearly the least satisfactory model, contains evidence sufficient to reject both H_2 and H_3 ($N = -6.03$, and -5.84), the two models which so far have not been conclusively falsified. This forcefully makes the point that the N statistics are quite different from comparisons of R^2 statistics or likelihoods. The model which fits the worst (H_5) can yield sufficient information to reject a model (H_3) which not only fits better than any of the other alternatives considered, but which cannot be firmly rejected in pairwise comparisons with any of the other models, all of which fit better than the one responsible for the rejection. The N test is not a measure of relative fit; it is a measure of whether a given hypothesis can or cannot explain the performance of an alternative hypothesis against the evidence. There is no requirement that that alternative should itself yield a satisfactory explanation of the phenomena under consideration.

In conclusion, it will be noticed that we are left without a satisfactory model; every formulation we have considered has been rejected against one or more alternatives. Considering the simplicity of the models, this is not surprising. Even so, the use of the technique in applied econometric work is likely to lead to this situation quite frequently in actual practice. We do not consider this to be to its disadvantage. It is our belief that too many hypotheses are accepted too readily in econometric work and it is our hope that the techniques discussed here will prove sufficiently powerful in practice to convince more researchers of the difficulties of establishing economic hypotheses. This can only enhance the status of those models which survive the tests.

Bank Markazi Iran, Tehran
and
University of Bristol

Manuscript received November, 1976; revision received May, 1977.

APPENDIX

THE DATA

TABLE A

EXPENDITURE, INCOME & WEALTH; U.S. 1954 II–1974 III
(Seasonally adjusted annual rates, $1958 billions)

1954	II	253	270	0	1964	I	366	387	760
	III	257	274	17		II	371	397	781
	IV	262	279	34		III	379	402	807
1955	I	268	282	51		IV	379	407	830
	II	273	289	66	1965	I	388	411	858
	III	276	294	82		II	393	416	881
	IV	280	299	100		III	400	430	904
1956	I	280	300	118		IV	409	438	933
	II	280	302	138	1966	I	415	442	962
	III	281	303	160		II	415	443	989
	IV	285	308	182		III	421	450	1017
1957	I	287	308	205		IV	421	454	1045
	II	287	309	226	1967	I	424	459	1079
	III	289	311	249		II	430	463	1114
	IV	290	310	271		III	432	468	1147
1958	I	286	307	291		IV	434	472	1183
	II	288	308	312	1968	I	445	480	1220
	III	292	315	333		II	448	486	1255
	IV	295	319	356		III	458	488	1293
1959	I	302	323	380		IV	460	491	1323
	II	307	328	401	1969	I	466	492	1354
	III	310	326	422		II	469	496	1381
	IV	310	328	437		III	470	504	1408
1960	I	314	332	455		IV	472	508	1442
	II	318	334	473	1970	I	474	511	1478
	III	316	334	489		II	478	522	1514
	IV	316	332	507		III	481	528	1559
1961	I	316	334	522		IV	478	524	1605
	II	320	340	540	1971	I	490	535	1652
	III	324	345	559		II	494	541	1697
	IV	330	352	581		III	498	542	1744
1962	I	333	355	603		IV	504	547	1789
	II	336	359	624	1972	I	513	552	1832
	III	340	360	647		II	523	559	1871
	IV	345	363	667		III	531	567	1906
1963	I	349	367	685		IV	542	584	1942
	II	351	369	703	1973	I	553	599	1984
	III	356	374	721		II	554	602	2030
	IV	358	379	739		III	555	605	2078
						IV	546	606	2128
					1974	I	540	594	2187
						II	543	587	2242
						III	547	587	2286

TABLE B

REAL INCOME 1947 I–1954 I

1947	I	217	1950	I	245	1953	I	269
	II	213		II	243		II	272
	III	218		III	247		III	271
	IV	216		IV	249		IV	271
1948	I	220	1951	I	248	1954	I	271
	II	226		II	253			
	III	231		III	254			
	IV	231		IV	254			
1949	I	227	1952	I	254			
	II	227		II	256			
	III	228		III	262			
	IV	230		IV	265			

REFERENCES

[1] ATKINSON, A. C.: "A Method for Discriminating between Models," *Journal of the Royal Statistical Society*, 32 (1970), Series B, 323–344.

[2] BROWN, T. M.: "Habit Persistence and Lags in Consumer Behavior," *Econometrica*, 20 (1962), 355–371.

[3] COX, D. R.: "Tests of Separate Families of Hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley: University of California Press, 1961.

[4] ———: "Further Results on Tests of Separate Families of Hypotheses," *Journal of the Royal Statistical Society*, Series B, 24 (1962), 406–424.

[5] DUESENBERRY, J. S.: *Income, Saving, and the Theory of Consumer Behavior*. Cambridge, Mass.: Harvard University Press, 1949.

[6] GAVER, K. M., AND M. S. GEISEL: "Discriminating among Alternative Models: Bayesian and non-Bayesian Methods," in *Frontiers in Econometrics*, ed. by P. Zarembka. Academic Press, 1974, 49–77.

[7] HACKING, I.: *Logic of Statistical Inference*. Cambridge: Cambridge University Press, 1965.

[8] PESARAN, M. H.: "The Small Sample Problem of Truncation Remainders in the Estimation of Distributed Lag Models with Autocorrelated Errors," *International Economic Review*, 14 (1973), 120–131.

[9] ———: "On the General Problem of Model Selection," *Review of Economic Studies*, 41 (1974), 153–171.

[10] QUANDT, R. E.: "A Comparison of Methods for Testing Non-Nested Hypotheses," *Review of Economics and Statistics*, 56 (1974), 92–99.

[11] STONE, J. R. N.: "Personal Spending and Saving in Post-War Britain," in *Economic Structure and Development: Essays in Honour of Jan Tinbergen*, ed. by H. C. Bos, H. Linnemann, and P. de Wolff. Amsterdam: North Holland, 1973, 79–98.

[12] U. S. DEPT. OF COMMERCE: *Survey of Current Business*. Washington, D.C.: Government Printing Office, 1975.