OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 48, 3 (1986)
0305-9049 $3.00

PRACTITIONERS' CORNER

Professor Hendry's Econometric Methodology

Christopher L. Gilbert

I. THE AER PROCEDURE[1]

It is widely acknowledged that there is a substantially different outlook on both the teaching and practice of econometrics in Britain relative to that in the United States.[2] This paper is expositional and does not attempt to provide an historical account of these developments.[3] Instead, it aims to give a simple account of the British approach to econometrics which will be accessible to the non-specialist econometrician.

A clean comparison between the 'British' econometric methodology and the traditional ('North American') methodology is aided by focusing on a single 'representative' econometrician. David Hendry, who is widely known for his work on savings behaviour[4] and also as a critic of Milton Friedman,[5] is the obvious choice. This is, in part because of his substantial participation in these developments, but primarily because he has explicitly addressed the large methodological issues. However, the fact that I associate positions and developments with his name should not be taken as implying that he is the originator of these developments. The comparison is also aided by a certain degree of caricature of the opposing position. I trust that this will be understood as a pedagogic device.

Hendry is concerned with the issues of *model specification* and *validation* in a time series context. I first erect a 'straw man' which I shall oppose to Hendry's view about how one should do econometrics.

[1] This paper started as a lecture to final year Oxford PPE undergraduates in early 1984. I am grateful to David Hendry for detailed comments, but the views expressed remain my own, and I take full responsibility for errors and misrepresentations. I have received helpful comments from many friends, colleagues and students on earlier drafts of the paper. However, I should particularly mention Les Godfrey, John Knight, Steve Nickell, Adrian Pagan and Tessa van der Willigen.

[2] The geographical partitioning of views is of course far less precise than this. Many of the positions which I characterize as either 'British' or 'Northern America' are held elsewhere in the world.

[3] I address the historical development of British econometrics in Gilbert (1986).

[4] Davidson *et al.* (1978), Hendry and von Ungern-Sternberg (1981), Davidson and Hendry (1981), Hendry (1983).

[5] Hendry and Ericsson (1983).

This straw man will be familiar to many students and ex-students from econometric theory courses. I shall call this straw man the *Average Economic Regression* (AER) view of econometrics. On the AER view, we have a specification derived from theory which we *know* to be correct. For simplicity, we consider the single equation specification

$$y = X\beta + \epsilon.$$

The econometrician's task is very simple – it is just to obtain an estimate b of the coefficient vector β. What are the econometrician's problems? He must worry about the *pathology* of his estimators – so he will worry about serial correlation, multicollinearity, heteroscedasticity, simultaneity and so forth. These pathological *problems* generate the chapters of econometrics textbooks.[6] The problems manifest themselves in applied work in terms of *low* Durbin–Watson statistics, *wrong* signs, *insignificant* coefficients and so forth. The term 'wrong' is telling – we know what the right sign is; the estimates give us the wrong sign; and the econometrician's response to these pathological manifestations is to respecify his equation in some way – to add or subtract variables, change the definition of variables and so forth – until, eventually, he gets an equation which has all correct signs, statistically significant coefficients, a Durbin–Watson statistic of around 2, a relatively high R^2 and so forth.

The econometrician proceeds to publish his results, but this presents the reader with a problem. She/he reads two articles on the same topic, one written by a Professor A. Smith and the other by a Dr K. Marks. Smith claims that a particular data set supports his theory whilst Marks, who uses the same or a similar data set, estimating a different model, since he has started from a different theory, claims that the data set supports his theory. Which should the reader believe? He may very well conclude that either bourgeois or radical economics, depending on his preferences, is not very scientific, and he probably should conclude that neither is scientific. It certainly cannot be the case that both Smith and Marks are correct. But the AER view of econometrics, shared by bourgeois and radical alike, has not been of much help.

The problem with the AER approach is that we are using econometrics to *illustrate* the theories which we believe independently. The alternative might be to use econometrics to *discover* which views of the economy (or market) are tenable and to test, scientifically, the rival views. Our normal concept of science tends to be based on the natural, and in particular, the physical sciences, and does not obviously translate to the social sciences. This alternative programme for econometrics

[6] For example, chapter 6 of Wonnacott and Wonnacott (1970) is entitled 'Serial correlation and other *problems*' and chapter 6 of Surrey (1974) is 'Miscellaneous single-equation *problems*' (italics added).

raises the question of what is required for economics to be scientifc. Much of Hendry's work may be seen as an attempt to grapple with this issue.

II. DATA GENERATNG PROCESSES AND MODELS

Hendry's approach to econometrics is grounded in the concept of the *Data Generating Process* (DGP) [Hendry and Richard (1982), p. 10, and Hendry and Richard (1983), p. 115]. This is nothing more than the joint probability of all the sample data (i.e. on both endogenous and exogenous variables). Thus x_t is the vector of observations on all these variables in period t, and $X_{t-1} = (x_1, \ldots, x_{t-1})'$. The joint probability of the sample x_t may be written as

$$\prod_{t=1}^{T} D(x_t | X_{t-1};\ \theta) \qquad (1)$$

where θ is a vector of parameters of the joint density function D. This joint density function is on the one hand uncontroversial, but on the other hand so general as to be useless. Econometric modelling consists of judicious simplification of this DGP. The econometrician will do this in four ways (generally iterating backwards and forwards):

(i) He will *marginalize* the DGP with respect to the variables that 'don't matter' (w_t say) in the determination of the variables of current interest;
(ii) He will *condition* the endogenous variables (y_t) on the (weakly) exogenous variables (z_t);
(iii) He will look for suitable *simple* representations of the conditioned marginalized DGP; and
(iv) He will replace the unknown parameters in this representation by estimated values.

Economic theory guides the econometrician in the first three of these steps; econometric estimation theory is concerned with the final operation. Provided that the marginalization and conditioning are valid, this allows the econometrician to replace the very general representation (1) by the much more specific[7]

[7] This requires that the DGP in (1) can be written as $D(x_t|X_{t-1};\ \theta) = D^1(W_t|X_t, Y_t; \lambda)\cdot D^2(y_t|Y_{t-1}, Z_t; \phi)\cdot D^3(Z_t|Y_{t-1}, Z_{t-1}; \psi)$ D^1 is the component of the DGP that depends on the irrelevant variables w_t; D^2 is the component of interest which relates the variable of interest y_t to the 'regressor' variables z_t, and D^3 is the component which describes the generation of these regressor variables. D^2 is the main object of investigation, but for certain purposes (e.g. forecasting) one will need to know (or to be able to approximate) D^3. Marginalization with respect to W_t involves two components: marginalization with respect to the current values w_t, which is generally unproblematic; and marginalization with respect to the lagged values W_{t-1} which requires Granger non-causality (see footnote 15).

$$\prod_{t=1}^{T} D(y_t \mid Y_{t-1}, Z_t ; \phi). \tag{2}$$

We do not know in advance whether the simplifications we have made in moving from (1) to (2) are valid. Furthermore, the concept of *validity* is not straightforward. Economies are complicated organizations, and this must be reflected in complexity of the general representation of the DGP given as (1). This complexity is reinforced by our inability in macroeconomics, and to a lesser extent more generally, to perform controlled experiments. Despite this complexity, limited observation sets typically force us to consider relatively simple models. If one uses test statistics with constant size (i.e. a constant degree of confidence), almost any simple model will be rejected given a sufficiently large data set. The fact that the hypothesis that a particular coefficient is equal to zero cannot be rejected only implies that insufficient data are available to permit this rejection; and any hypothesis can be maintained by testing it on a sufficiently short time series.[8]

This implies that we know for certain that the simplified representation of the DGP (2) cannot be strictly valid. The question therefore becomes one of *adequacy* rather than *validity*. Hendry proposes that we look for a *Tentatively Adequate Conditional Data Characterization* or what might be called a model which is *congruent* with all the evidence. The conditionality is on the data, and the characterization is only tentatively adequate because we should not be so vain as to suppose that it is impossible to improve upon our work. Models are not right or wrong but are useful or misleading for particular purposes; non-congruent models are open to constructive improvement.

This discussion of the derivation of the model (2) as a tentative simplification of the overall DGP (1) is to be contrasted with the AER procedure in which the model (1) is simply asserted as correct. This has major implications for econometric procedure – in the AER procedure poor test statistics imply problems in consistently and efficiently estimating the parameters of the (axiomatically correct) model; in the DGP approach the same statistics, obtained from the same regressions, imply model misspecification.

The AER response to a poorly fitting equation is to add variables or parameters (e.g. autocorrelation parameters) in an attempt to 'patch' the original 'theoretical' model. We may represent this approach as *simple → general*. In direct opposition, Hendry advocates *general → simple*. Take the consumption function as an example. The AER procedure starts from the textbook representation

$$C_t = \beta_0 + \beta_1 Y_t. \tag{3}$$

[8] For this reason, Leamer (1978) proposes that the econometrician should reduce the significance level he uses as his sample size increases.

When this appears inadequate, the specification is altered by, for example, adding C_{t-1} as an additional regressor, or by adding inflation or interest rate regressors. The difficulty with this approach is that, if two investigators start from different simple hypotheses (as is likely to be the case with Smith and Marks) there is no reason to suppose that they will converge on the same final equation and even with the same starting point, investigators may diverge if they adopt different respecifications.

The general-to-simple methodology is less vulnerable to this objection. The investigator starts with a very general hypothesis that is acceptable to all the adversaries and then narrows it down by looking for simplifications that are acceptable on the data. Thus, if it is agreed that we should confine our attention, in the consumption function example, to income, inflation, and liquid asset variables, we might initially specify

$$\ln C_t = \sum_{j=1}^{5} \alpha_j \ln C_{t-j-1} + \sum_{j=0}^{5} \beta_j \ln Y_{t-j} + \sum_{j=0}^{5} \gamma_j \ln P_{t-j} + \sum_{j=0}^{5} \delta_j \ln L_{t-j-1} + \psi' q_t + \epsilon_t \qquad (4)$$

where P is the consumption deflator and L is end-period holdings of liquid assets. q_t is a vector consisting of the constant term, three seasonals and a time trend. (A limit on the length of the lag distributions is forced by degrees of freedom considerations.)[9] One then conducts what Leamer has called a *simplification search* with respect to this general specification (Leamer, 1978). Trivedi (1984) refers to this iterative estimation and testing procedure as 'testimation'.

In moving from the general to the simple, therefore, we confine our attention to specifications that are acceptable, using the classical F-test, as simplifications of the general specification. I shall call such simplications *F-acceptable*. This criterion alone is likely to rule out the contributions of Professor Smith and Dr Marks – since they have not estimated the general specification they will not have been able to ask whether or not their results might be rejected by more general hypotheses. The estimated general equation may suggest simplifications (if for example, a number of coefficients are near zero), and eliminate (as implausible) other simplifications (those involving deletion of variables with estimated coefficients which are relatively large in rela-

[9] The greater the number of variables included in the equation, the shorter the lag length that must be imposed. But why five periods? Hendry would argue that this is natural in the analysis of quarterly data, since it allows simplification to include both fourth differenced variables (e.g. $\Delta_4 \ln Y_t$) and the first difference of these fourth differences ($\Delta_1 \Delta_4 \ln Y_t$) – see Davidson *et al.* (1978) for an example.

tion to their estimated standard errors). Moreover, the long run elasticities calculated from the estimated general equation will (as the consequence of collinearity amongst the regressors) have much smaller standard errors than will the individual coefficients, and will not be subject to misspecification bias through the incorrect omission of variables contained in the general data set.

Even within the general-to-simple approach, there are likely to be alternative *F*-acceptable congruent simplifications of the general representation, and we will be forced to choose between these. Which does one choose? There are four distinct questions here:

What criteria must a satisfactory model satisfy?
How do we discover such models?
What features should we design into our models?
What do acceptable models typically look like?

These questions form the agenda of the remainder of this paper. But there is a major issue which we need to address first.

The AER procedure is open to the objection that, since the published regression has been accepted in part because its coefficients all have the correct signs and are statistically significant, these significant coefficients cannot be taken as evidence for or against the hypotheses under investigation.[10] What, one is tempted to ask, about the other 999 regressions which have been consigned to Smith's (Marks') waste bin? Do these also confirm/reject the same hypotheses? It is this problem which prompts Leamer (1983) to propose that econometricians confine themselves to publishing mappings from prior to posterior distributions, rather than actually making statements about the economy.[11]

The important question in the present context is whether the Hendry methodology is robust in this respect. Leamer (1983) argues that we need to invoke the *axiom of correct specification* in order to justify use of classical statistical methodology on a regression which has been chosen on the basis of regression estimates of competing specifications. This view implies that the use of classical methodology in econometrics is largely an act of faith. A more appealing way to recast Leamer's proposition is as the *axiom of route independence*: the validity or invalidity of a particular specification cannot be dependent on the research strategy that generated the specification. This axiom implies that testing must proceed as if the specification were the first selected. But the implication of Leamer's axiom is reversed: we do not assume that the specification is correct, but test to see whether it is congruent

[10] A regression estimate of an equation specification selected on the basis of a previous regression is known as *pre-test* estimator. The distribution theory of such estimators is complicated – see Judge and Bock (1978).

[11] See McAleer *et al.* (1985) for a critical discussion of this view.

with the data. Tests of the specification relative to the general formulation (perhaps (4)), together with other tests to be discussed in Section 3, establish whether or not this can be maintained. If this is established, then we may tentatively proceed to conduct tests within the specification without the support of the axiom of correct specification.[12]

III. MODEL ACCEPTANCE CRITERIA

Hendry and Richard (1983) give six criteria for model selection.[13]

(i) Models must be *data admissible*. This is a logical criterion – is it logically possible for the data to have been generated from this model? This concept is best illustrated by giving examples of a data-inadmissible specification. Consider the case of modelling a proportion of agents taking a certain action – for example the proportion of voters voting Conservative. Proportions must lie in the interval [0, 1]. If the dependent variable of the equation is defined so that with non-zero probability it will lie outside this range, it is possible to predict that less than 0 per cent or more than 100 per cent of the electorate will vote Conservative. This violates the data admissibility criterion.

(ii) Models must be *consistent with theory*. There may of course be alternative theories, but a satisfactory model must be consistent with at least one theory.

(iii) Satsifactory models should have regressors that are (at least) *weakly exogenous*. (In a simultaneous context, this requirement refers to the regressor variables in the reduced form.) Technically, with weakly exogenous regressors, it is valid to condition on the regressor set. If the regressors are not weakly exogenous, then they are endogenous by default and must be jointly modelled (in a simultaneous system). Rather obviously, there is not much point in claiming that one has 'explained' y in terms of x if x, in turn, is 'explained' in terms of y. What is required is a joint explanation of x and y.

For certain purposes weak exogeneity may not be sufficient. Consider the following simple model:

$$y_t = \alpha + \beta x_t + u_t \qquad u_t \sim \text{NID}(0, \sigma_u^2) \qquad (5)$$

$$x_t = \gamma + \delta y_{t-1} + v_t \qquad v_t \sim \text{NID}(0, \sigma_v^2) \qquad (6)$$

[12] A problem with sequential testing is that if each test is conducted at, say, a 5 per cent significance level, the size of the overall sequence of tests is likely to be considerably in excess of 5 per cent. Each component test should therefore be conducted at a lower level of significance if the overall probability Type I errors is to be controlled – see Mizon (1977).

[13] Their order of presentation is somewhat different, but I do not believe that this implies any ordering of priority. Hendry and Richard (1982) give the same list but omit (ii), which is discussed earlier in the article, and might be considered to be implied by a reformulation of (iv) as the requirement that the parameters *of interest* be constant.

and where $E(u_s v_t) = 0 \ \forall s, t$. x_t is weakly exogenous in relation to β which may therefore be consistently estimated, by for example, OLS[14] but if one proposes to use (5) to forecast more than one period ahead, one will also need to know (6), because any change in ones forecast y_f will alter the value of the 'explanatory' variable x_{f+1} needed to forecast y_{f+1}. Weak exogeneity is sufficient (in most cases) for testing hypotheses, but for forecasting one in general requires *strong exogeneity* which is weak exogeneity plus an absence of feedback.[15] Forecasting models therefore tend to be larger, and often much larger, than the models used to test hypotheses about behaviour in particular sectors of the economy.

An even stronger requirement is *super exogeneity*, which may be required in policy analysis. Suppose, for example, that in the regression

$$y = X\beta + \epsilon \tag{7}$$

one of the theoretical variables relates to agents' expectations about the money stock (or its rate of growth), and that this is modelled by including in the set of X variables lagged values of the money stock. If the monetary authority now changes its monetary growth rule, it is plausible that the relationship between the expectational variables and the lagged money stock variables will be altered, and consequently, the policy change will result in a change in the β coefficients. Given strong exogenity of X, (7) may forecast perfectly well, but it will in principle be misleading if used to analyse the implications of alternative monetary policies. For this we require that the parameter vector β be independent of the process generating the X variables, and not just of the values of the X variables. This will be the case if (7) is respecified in terms of more fundamental super exogenous variables.[16]

[14] Note that variables are not weakly exogenous with respect to other variables, but relative to the parameters of interest. (The same point applies to the less useful concept of *predetermination*). For example, if y_t and z_t are jointly normally distributed, it is valid to regress y_t on z_t, taking z_t as weakly exogenous, but by symmetry, one may equally take z_t to be weakly exogenous and regress z_t on y_t. Whether one decides to follow one or other or neither of these alternatives depends on the parameters of interest, and this is a question which relates to the economic theory relating y_t and x_t, and not to their statistical relationship.

[15] If the z variables are to be strongly exogenous, then D^3 in footnote 7 (above) must specialize to

$$D^3(z_t | Z_{t-1}; \psi)$$

implying that z_t is independent of all (past and present) values of y_t. This latter requirement is the condition that y does not *Granger-cause* z. Strong exogeneity is equivalent to weak exogeneity plus Granger-non-causality. These relationships are very clearly explained in section 5 of Engle *et al.* (1983).

[16] This is related to the famous *Lucas critique* (Lucas 1976)). Super exogeneity requires weak exogeneity, and in addition that the parameter vectors ψ and ϕ be independent. Technically, this states that they must lie in non-intersecting spaces Ψ and Φ respectively, but an alternative way of putting this is that there does not exist a lower dimensional vector μ such that $\psi = \psi(\mu)$ and $\phi = \phi(\mu)$. Super exogeneity is a sufficient condition for a parameter vector to be invariant, but it is not necessary since particular parameters of interest may turn out to be

(iv) Satisfactory models must exhibit *parameter constancy*. This is essential if the model is to be of any use in forecasting or in policy simulation, both of which require that the same parameter values apply inside and outside the sample period. A consistent feature of Hendry's applied work is that he uses a number of observations at the end of the sample period for testing the constancy of the estimated model. This may be done using a standard Chow test (Harvey, 1981, p. 181). In the single equation context

$$\xi = (\hat{y}_f - y_f)'[I + X_f(X'X)^{-1}X_f]^{-1}(\hat{y}_f - y_f)/s^2 \sim F(n, T-k) \qquad (8)$$

where $\hat{y}_f$ is the vector of forecasts for n post-sample periods conditional upon the values X_f of the regressor variables, y_f are the outcomes, and s^2 is the standard estimate of the error variance. If this test is satisfied, then, however poorly it may forecast, the forecast errors do not betray any evidence of parameter non-constancy. In this case, efficiency then demands that the model be re-estimated over the entire sample (including the final n observations). However, if the model fails this test, it cannot adequately represent the Data Generating Process, and respecification is required.

(v) A satisfactory model must be *data coherent*. This is the requirement that the differences between the fitted values generated by the model and the actuals should be random, in the sense that they should not be predictable from their own past history. For a model to be satisfactory, it should not be possible to predict how the model will misforecast. If one can make such predictions, then there exists a superior model, which is simply the original model augmented by the process for forecasting the original model's misforecasts.

This has implications for serial correlation, since serial correlation of a model's residuals is just such a systematic departure of a model's forecasts from the out-turns. Here, then, is a radical difference between the AER view and Hendry's methodology. On the AER view, the absence of serial correlation is tested for using the Durbin–Watson statistic (which is most powerful against the presence of first order autocorrelation) and, if found, is corrected by application of the Cochrane–Orcutt transformation. On Hendry's view, such correction is entirely inappropriate since the presence of serial correlation will generally imply a model which cannot be taken to represent the Data Generating Process, and the remedy to this is a revised specification and not re-estimation.

Most frequently, the required revision is the incorporation of one or more of the lagged regressor variables. The specification

$$y_t = \alpha + \beta' x_t + u_t \qquad (9)$$

---

invariant with respect to changes in, say, expectations generation, even though other parameters are affected.

with

$$u_t = \rho u_{t-1} + \epsilon_t$$

where the $\{\epsilon_t\}$ are white noise, may be transformed to

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + \beta' x_t - \rho\beta' x_{t-1} + \epsilon_t \quad (10)$$

The unrestricted version of this equation may be written as

$$y_t = \gamma_o + \gamma_1 y_{t-1} + \gamma_2' x_t + \gamma_3' x_{t-1} + \epsilon_t \quad (11)$$

The hypothesis of first order serial correlation is only acceptable if the restriction

$$\gamma_1\gamma_2 + \gamma_3 = 0 \quad (12)$$

is satisfied. If this turns out to be the case serial correlation is 'a convenient simplification, not a nuisance,' but more usually the restriction will be rejected, and the statistic indicating serial correlation would be correctly interpreted as pointing to the omission of the relevant lagged variable [Hendry and Mizon (1978)].[17] As will be apparent, this is just an application of the general to simple methodology discussed in Section 2 (serial correlation implies the presence of a 'common factor' in the lag distributions in (11)).

Hendry's view of serial correlation follows from the fact that the properties of the equation disturbance are implied by the model specification. On this view, the disturbance ϵ_t is derived from the model. This is to be contrasted with the AER view where one posits $y = X\beta + \epsilon$ and then proceeds to make assumptions about ϵ. This implies a particular process for y, and is indeed how we proceed in Monte Carlo analysis. However, in empirical work, the dependent variable already follows a particular process, and the assumption that $y = X\beta + \epsilon$ implies (if β is constant and X follows a known and independent process) a process for the disturbance ϵ. Any additional assumptions about ϵ are likely to lead to over-determination. Rather, in looking for data coherence, one is looking to see whether the derived ϵ process is close to white noise. If it is not, this suggests augmentation or respecification of the equation.[18]

[17] There is a complication that some components of x_{t-1} may be exactly collinear with x_t (for example if x_t contains a time trend or seasonal dummies). Let x_{t-1}^* be the sub-vector of x_{t-1} which is not exactly collinear with x_t and partition x_t and γ_2 conformably into x_t^* and x_t^{**} and γ_2^* and γ_2^{**} respectively. Then (9) becomes

$$\gamma_1\gamma_2^* + \gamma_3 = 0$$

The seminal paper is Sargan (1964), extended in Sargan (1980).

[18] 'Most estimation theory in econometric textbooks is predicated on knowing the "correct" model, while simultaneously, most practitioners are well aware that their models are inevitably inadequate in many respects. Consequently, it seems more useful to try and *design* empirical models such that their residuals are white-noise *innovations* relative to the available information, ...' (Hendry and Richard (1983, p. 17, authors' italics).

(vi) A satisfactory model should *encompass* a wide range of rival models.[19] One model may be said to encompass a second model if it can explain the second model's results. A good model should not only explain the data, but should also explain both the successes and the failures of rival models in accounting for the same data. The Special Theory of Relativity encompasses Newtonian mechanics (which could not explain the Michelson–Morley results), but is itself encompassed by the General Theory of Relativity (as it cannot explain the visibility of the planet Mercury behind the Sun during a solar eclipse).

Suppose Smith believes

$$H_{\text{Smith}}: y = X\alpha + u \qquad u \sim N(0, \sigma_u^2 I) \tag{13}$$

whilst Marks asserts

$$H_{\text{Marks}}: y = Z\beta + v \qquad v \sim N(0, \sigma_v^2 I) \tag{14}$$

If $X \neq Z$, at most one of these theories can give both a correct and complete description of the economy.[20] In the actual sample, there is also a relationship between the X and Z variables, which may be summarized as

$$Z = X\Pi' + w \tag{15}$$

This is uncontentious. Π is simply the set of regression coefficients obtained from regressing the Z variables on the X variables over this sample, and may have no theoretical interpretation. This relationship implies that, on H_{Marks}

$$y = X\Pi'\beta + (v + w\beta). \tag{16}$$

This allows Marks to attempt to predict Smith's results, since equating coefficients in the two equations (14) and (16) relating y to X

$$\alpha = \Pi'\beta \tag{17}$$

and

$$\sigma_u^2 = \sigma_v^2 + \beta'\Sigma_w\beta \tag{18}$$

where Σ_w is the sample variance covariance matrix of w.[21] Note that this implies $\sigma_u^2 > \sigma_v^2$, so there is no point in attempting to encompass a

[19] In Hendry and Richard (1982) *all* rival models are to be encompassed! In practice the data are unlikely to be decisive in all cases. See Mizon (1984) and Mizon and Richard (1985) for more extended discussion of encompassing.

[20] Both may be 'correct' in the limited sense that it can be true that $E(y_t|x_t) = \alpha' x_t$ and $E(y_t|z_t) = \beta' z_t$. However, it cannot be the case that $E(y_t|x_t, z_t) = \alpha' x_t$ and $E(y_t|x_t, z_t) = \beta' z_t$ (for arbitrary values of α and β) unless $x_t = z_t$ (Mizon, 1984).

[21] This requires that v and w be independent. By (weak) exogeneity of Z, $E(Z'v) = 0$. In addition, Smith must claim (weak) exogeneity of X implying $E(X'v) = 0$. But (11) then implies $E(W'v) = 0$.

theory with a lower error variance.[22] If H_{Marks} is correct, it can predict the coefficients that will be obtained by fitting the incorrect specifica- H_{Smith}, and it will also predict how badly H_{Smith} will fit the data.

It is interesting that both of these conditions may be tested by the classical F test on the common nesting hypthesis

$$H_{\text{nesting}}: y = Z\beta + X^*\gamma + e \tag{19}$$

where X^* is a submatrix of X which is not exactly collinear with Z. H_{Marks} simply implies $\gamma = 0$. Thus encompassing may also be seen as an application of general to simple methodology. The role of H_{nesting} is to provide a common distributional framework for H_{Smith} and H_{Marks}, and it is important to note there need be no theory which gives it an interpretation (Mizon, 1984, pp. 142–43). If both H_{Marks} and H_{Smith} are rejected (the latter by formulation of H_{nesting} in terms of Z^* and X), then each theory rejects the other, and more work is required. (It is of course possible that neither theory will be rejected on a given data set.)

I have now listed the six criteria which an acceptable model must satisfy. A model which does satisfy all six criteria is said to be *congruent* with the evidence. It is notable that the major criterion in the AER methodology is absent from this list. This is the criterion of goodness of fit. It is possible to characterize an alternative approach to specification as that of selecting *parsimoniously undominated* models (e.g. Amemiya, 1980). A readily recognizable caricature of the econometrician's objective is the maximization of $\bar{R}^2$ subject to Durbin-Watson near 2. In the Hendry methodology, goodness of fit is not an explicit criterion for accepting a model as a good representation of the Data Generating Process. However, the criterion is implicitly present in the encompassing principle, since in a linear framework, encompassing implies *variance-dominance* (i.e. the criterion which states that one should accept the model with the lowest variance). Hendry simply notes that this is one of several criteria that one would wish to adopt (Hendry and Richard, 1982, p. 14).

That a model can be acceptable despite having a poor fit is clear. Similarly, the fact that a rival model has a better fit does not *ipso facto* make it a better model. The latter point is perhaps obvious through the phenomenon of over-fitting: R^2s can always be increased by the addition of further 'explanatory' variables, but in general this will result in apparent structural non-constancy as additional data points are added and so parameter constancy will be violated. To illustrate the former point, consider modelling asset prices generated by an efficient market (stock market or foreign exchange prices for example). There is

[22] This comment requires the qualification that in small samples one might have $\hat{\sigma}_v^2 > \hat{\sigma}_u^2$ despite (18) as the consequence of sampling variation.

every reason to suppose that, at least to a very good approximation, such prices follow a martingale process, and that the changes in these prices are therefore unpredictable. A random walk model will have zero explanatory power in relation to these price changes, but may nevertheless be a perfectly acceptable statistical representation.[23]

More generally, however, goodness of fit is implied by use of the encompassing principle. One representation can only encompass other representations if it has lower error variance than do these alternatives. And, because we are looking for simplifications and have thus implicitly adopted a parsimony principle, if one F-acceptable simplification nests another, we adopt the second (more parsimonious) specification. Uniqueness is not guaranteed, but the set of possible representations will be much smaller, and these will all be sustainable in relation to the given data set. Thus it might be possible to claim that a particular asset price follows a martingale process relative to one information set, and is therefore unpredictable from that information set, but this explanation may be encompassed by an alternative that makes the prices predictable relative to an enlarged data set (which includes, for example, insider information).

IV. MODEL DESIGN AND MODEL DISCOVERY

How should the econometrician set about discovering congruent simplifications of the general representation of the DGP (1) (i.e. simplifications which satisfy the six criteria in Section 3)? Hendry offers no advice (Hendry, 1985). Scientific discovery is necessarily an innovative and imaginative process, and cannot be automated.[24]

Nevertheless, the same scientific insight may be expressed in a number of different ways; and there will be an element of choice as to how a particular hypothesis is incorporated in an econometric model. This is the issue of model design. In looking for counterparts to Economics among the natural sciences, Engineering Science has at least as great a claim as Physics, and I shall use an engineering analogy.[25]

One of the most obvious feats of mechanical engineers is bridge-building, and their success in this activity, together with the occasional

[23] The random walk model strengthens the martingale model by the additional imposition of constant variance.

[24] Does the common factor approach, discussed in relation to serial correlation in Section 3, provide a method for automating model discovery? A general procedure based on the Wald test for testing for the number of common factors in an unrestricted representation (like (4)) was developed by Sargan (1980), and is implemented in Hendry and Mizon (1978). The difficulty with this approach is that common factor tests have very low power in samples of the size normally employed in time series econometrics (Mizon and Hendry (1980)). Hendry appears to have retreated from this pure approach, if indeed he ever whole-heartedly adopted it. He sees the general to simple approach as providing a testing methodology but not a method of discovery.

[25] This analogy is taken from Hendry's graduate lectures.

failure, is apparent to the entire world. Hendry's bridges are econometric models. Bridges are built to serve particular purposes, and so are econometric models. Bridge design rests on the sciences of mechanics and the properties of materials; similarly, the structure of econometric models relies on economic and econometric theory. Bridge designs must be thoroughly tested, for example in wind tunnel experiments, before the bridge can be built; similarly, econometric models should be thoroughly tested before being employed in forecasting or policy analysis. Builders of bridges must submit their plans and their test results to detailed professional and public scrutiny, and the same should be true of econometric model-builders. Cost is important in building bridges (steel is used where tungsten would undoubtedly be stronger), and similarly econometric model-builders must operate within budget constraints. And whilst it is true that the failure of a bridge is more dramatic than mistaken policies resulting from the use in forecasting or policy formation of incorrect econometric models, the welfare loss from the latter may also be large. What then, should a good econometric bridge look like?

Partly, this must depend on what sort of traffic it is to carry. Heavy lorries require a stronger construction than do pedestrians and cyclists. In the econometric design process, a vital issue is that the appropriate concept of exogeneity depends on the use to which the model is to be put – weak exogeneity will in large measure be sufficient for hypothesis testing; strong exogeneity is in general required for forecasting; and for policy analysis, we will require super-exogeneity. If one is involved in forecasting or policy analysis, then the time horizon will be important. Responses which may be negligible if one's horizon is 12 months may dominate the outcome if one's horizon is 12 years. In addition, different problems may require different degrees of disaggregation.

A second criterion that Hendry would adopt is *robustness*. Econometric models are tested in the wind tunnels of sample data, and, in good professional hands, are also validated against immediate post-sample data (8). But no firm of civil engineers would gain a bridge contract unless they could assure the authorities that the bridge would be able to withstand all reasonable eventualities. What changes from sample conditions should worry the econometrician? OLS estimates reflect the sample data covariance matrix $T^{-1}X'X$, and the same matrix, projected into some other space or otherwise transformed will be important in other estimation procedures.[26] Robustness requires that the parameter estimates, and hence model forecasts and policy prescrip-

[26] $X'\Omega^{-1}X$ in GLS, and $X'P_ZX$ in Instrumental Variables (where P_Z is the projection matrix into the space spanned by the instruments). LIML is asymptotically equivalent to Instrumental Variables estimation, and so the same point holds as an asymptotic approximation.

tions, are unaffected if, in the future, the world throws up a different data covariance matrix. If the weight limit for lorries rises from 32 tonnes to 36 tonnes, bridges should not need to be strengthened.

This all relates to the chapter in the AER pathologist's handbook entitled *multicollinearity*. Collinearity is often seen as a problem relating to the data, and it is obviously true that the correlations in the data are what they are. But equally, the econometrician is free to define his regressor matrix X in any way he likes in relation to this data. In experimental applications, collinearity reflects bad experimental design, and in necessarily non-experimental time-series econometrics, collinearity implies failure to exploit efficiently the data which are available. The regression

$$y = W\alpha + \epsilon \qquad (20)$$

where $W = XA$ for A square and non-singular, is statistically indistinguishable from the regression

$$y = X\beta + \epsilon \qquad (21)$$

However $X'X$ may achieve an arbitrarily high degree of collinearity on any proposed measure, whilst $W'W$ can be as close to orthogonal as one likes.[27] In looking for simplifications of the general hypothesis (e.g. (17)), robustness dictates that we should look for simplifications that involve near-orthogonal variables. Specifically, consider the regression

$$y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \epsilon_t \qquad (22)$$

It is likely that X_t and X_{t-1} will be quite highly correlated, but the regression may be equivalently reformulated as

$$y_t = \alpha_0 + \alpha_1 \Delta X_t + \alpha_2 X_{t-1} + \epsilon_t \qquad (23)$$

where $\alpha_0 = \beta_0, \alpha_1 = \beta_1$ and $\alpha_2 = \beta_1 + \beta_2$. The level variable X_{t-1} and the change variable ΔX_t will be nearly independent when X_t is highly autoregressive. Similarly, a regression involving a distributed lag on X_t, X_{t-1} and X_{t-2} may be reformulated in terms of a level effect (X), a first difference (ΔX) and a second difference ($\Delta^2 X$). This is essentially the procedure followed by Davidson *et al.* (DHSY, 1978).[28] The resulting

[27] Setting A equal to the matrix of eigenvectors of $X'X$ makes $W'W$ diagonal. The columns of W are then the principal components of X. This is not however a sensible transformation of the X variables since the parameters (α) of the principal components in the y regression will not in general be interpretable.

[28] The simplification in Davidson *et al.* (1978) is in terms of seasonal differences ($\Delta_4 Y_t = Y_t - Y_{t-4}$) and the first difference of these differences ($\Delta_1 \Delta_4 Y_t$). Modelling in terms of seasonal differences may be appropriate when using quarterly data if the same quarters in successive years are more closely related than successive quarters in the same year. This would be the case, for example, if families plan their vacation expenditures over a period of years in relation to their permanent income, rather then take whatever vacations they can afford in a given year in relation to their income that year (so that a week in Blackpool in 1985 may be followed by a fortnight in the Bahamas in 1986). But note that in an error correction equation,

near-orthogonality of the regressor variable set allows valid inferences to be drawn from individual estimated parameters taken in isolation; and permits inference about the effects of omitting particular regressors from the equation.

Simplification in terms of near-orthogonal variables may conflict with a further requirement that it is reasonable to seek in econometric models, namely that the parameters of the models be *interpretable* in terms of economic theory. This is not simply a matter of compatibility with theory. Consider for example the contrast between structural and reduced form versions of a model. If the former is compatible with a particular theory, then, *a fortiori*, so must be the latter. But the coefficients of the reduced form model are not (at least directly) interpretable. In relation to collinearity, we can always find data transformations that will give complete orthogonality, but there is no guarantee that these coefficients will be interpretable, and in general they will not.

Why is interpretability important? Two reasons may be proposed. Firstly, although reduced form coefficients may be compatible with theory, it is difficult to know whether or not this is true of any particular set of reduced form coefficients. Secondly, parameters sometimes change, and indeed, in a policy context, the policy may consist of changing parameters (e.g. tax parameters). In order to exploit our knowledge of these changes, we must parameterize in terms of the theoretical (structural) parameters. The DHSY simplification mentioned above must therefore be rationalized not only in terms of the near-orthogonality of the regressor variables, but also in terms of the theoretical interpretability of the DHSY relationship. This is the subject of the next section.

A third criterion which Hendry adopts is *parsimony*. Parsimonious specification is the econometric application of Occam's razor – if two competing explanations of the same phenomenon have the same explanatory power, then select the simpler. The epistemological justification of the parsimony principle has provoked long debates. There is no convincing reason to suppose that simplicity is an in-built feature of creation; and neither biological nor social evolutionary theories imply a tendency towards simplicity. It is, therefore, unreasonable to suppose that economic DGPs will be simple. But if the DGP is not simple, why should simplicity be a desirable feature of a congruent model?

There are two reasons for valuing parsimony in econometrics, and these both relate to the charaterization of econometric models as *tentatively* adequate. At a practical level, degrees of freedom considera-

---

in which lagged levels variables are included in the regression, the Davidson *et al.* argument that seasonal differencing allows the seasonal dummies to be dropped (on the grounds that $\Delta_4 q_t = 0$ if q_t is a seasonal dummy) is incorrect – see Hendry and von Ungern-Sternberg (1981).

tions force simplicity. A given time series will not allow precise estimation of more than a small number of parameters; and, even with the adoption of Hendry's orthogonalization procedure, the econometrician will be forced to choose between omitted variable bias and imprecise estimates resulting from inclusion of too many explanatory variables. Secondly, it is arguable that the human intellect possesses a limited ability to comprehend complex phenomena. Complexity permits *degenerate* research programmes to defend their central propositions by a *protective belt* of auxiliary hypotheses (Lakatos (1970)).[29] These auxiliary hypotheses may be correct, but it is more plausible to suppose that vested professional interests are being protected. Parsimony, on this view, forces scientific honesty.

The econometric modeller must therefore decide on what is important in relation to (i) theory, (ii) the characteristics of his sample and (iii) the use to which the model is to be put, and seek, within this context, a simple specification satisfying the criteria of Section 3. The resulting model should be specified, according to Hendry, in terms of a near-orthogonal regressor variable set, satisfy exogeneity assumptions sufficiently strong for the uses to which the model will be put, and have coefficients which are both constant and interpretable in terms of economic theory.

V. DO ACCEPTABLE MODELS HAVE A COMMON STRUCTURE?

In his applied econometric work over the past decade, Hendry has consistently adopted a specification which has become known as the *error correction model* (ECM). The best known instance is the Davidson *et al.* (1978) (DHSY) consumption function.

The ECM, which was introduced to economists by Phillips (1954), was first explicitly adopted in applied econometrics in Sargan's (1964) very influential model of UK wage determination. Sargan modelled the quarterly change in wage rates as

$$\Delta \ln w_t = \beta_0 + \beta_1(\ln p_{t-1} - \ln p_{t-4}) - \beta_2 u_{t-1} - \beta_3(\ln w_{t-1} - \ln p_{t-1}) + \beta_4 t + \psi q_t + \epsilon_t \quad (24)$$

where w_t is the index of wage rates, p_t is the Retail Price Index, u_t is the unemployment rate, and q_t is now a vector of seasonal and other dummies. (The model also included an equation for p_t.) The important term here is $-\beta_3(\ln w_{t-1} - \ln p_{t-1})$, and if we incorporate the time trend, we may write this composite term as

[29] Hendry (1985) appears to place himself deliberately within a Lakatosian framework. See Cross (1982) for a Lakatosian discussion of modern economic theory.

$$-\beta_3\left[\ln\left(\frac{w_{t-1}}{p_{t-1}}\right)-\ln\omega_t\right] \tag{25}$$

where (with suitable adjustment to the intercept)

$$\omega_t=\left(\frac{w_0}{p_0}\right)\exp\left(\frac{\beta_4 t}{\beta_3}\right)$$

We may interpret ω_t as the *target real wage* towards which the actual real wage, w_t/p_t, is adjusting, and, as was natural in a model estimated on a sample which terminated in the mid-60's, Sargan saw the target real wage as rising steadily over time. To the extent that the actual real wage was in excess of the target, wage settlements would be lower (because of the negative coefficient β_3) but if the actual real wage fell below the target (through the effects, for example, of incomes policy or Sterling devaluation) wage increases would be higher than otherwise. The error correction term in the Sargan wage model is therefore directly interpretable in terms of the target real wage hypothesis. Hendry intoduces analogous terms into his consumption functions, where their presence might have been less expected.

Consider an annual version of the DHSY model (with the inflation terms omitted for simplicity):

$$\Delta\ln C_t=\beta_0+\beta_1\Delta\ln Y_t-\beta_2(\ln C_{t-1}-\ln Y_{t-1})+\epsilon_t \tag{26}$$

This is the most simple example of an ECM. The change in the dependent variable (the logarithm of consumption) is related to the change in the explanatory variable (log income), and the lagged discrepancy between the dependent and the explanatory variable. In a quarterly version of the model, lagged or higher differences of the explanatory variable are likely to appear, and the final error correction term will often appear with a longer lag length (typically four).

It is straightforward to obtain the long-run consumption-income ratio along a steady state growth path from (26).[30] This is the ratio towards which the actual rate will tend. If the rate of income growth is set to g, this ratio is

$$\ln\left(\frac{C}{Y}\right)=\frac{\beta_0-(1-\beta_1)g}{\beta_2} \tag{27}$$

It is sometimes argued that the ECM should be seen as defining a class of model parameterizations rather than a class of models. To see this, consider a second order general distributed lag linking the logarithm of consumption to the logarithm of income

[30] The presence of growth rates in the steady state solution has provoked some discussion – see Currie (1981) and Nickell (1985).

$$\ln C_t = \alpha + \sum_{j=1}^{2} \beta_j \ln C_{t-j} + \sum_{j=0}^{2} \gamma_j \ln Y_{t-j} + \epsilon_t \tag{28}$$

Trivially, this may be rewritten, without imposing any restrictions, in terms of current and lagged income and consumption changes, and a set of lagged level variables.

$$\begin{aligned}\Delta \ln C_t = {} & \alpha + (\beta_1 - 1)\Delta \ln C_{t-1} + (2\gamma_0 + \gamma_1)\Delta \ln Y_t - (\gamma_0 + \gamma_1)\Delta^2 \ln Y_t \\ & - (1 - \beta_1 - \beta_2)(\ln C_{t-2} - \ln Y_{t-2}) \\ & + (\beta_1 + \beta_2 + \gamma_0 + \gamma_1 + \gamma_2 - 1)\ln Y_{t-2}\end{aligned} \tag{29}$$

Equation (29) is of the ECM form, but it is observationally equivalent to the unrestricted specification (28). No restrictions have been imposed. However, the ECM parameterization may be useful because the near-othogonality of the regressors (here $\Delta \ln Y$ and $\Delta^2 \ln Y$) facilitates parsimonious representation of the general distributed lags. This is because the t-statistics on the variables in the unrestricted ECM equation (29) provide a good guide as to the effects of omission of these variables in a simplified equation. But if this were all, the ECM would only define a useful regression strategy, as discussed in the previous section.

In fact, more is implied. In the ECM strategy the investigator is permitted to eliminate and otherwise restrict the influence of the differenced variables at will, but cannot elminate the lagged level terms since it is these terms that define the equilibrium solution of the model. Equation (29) implies that across steady state equilibria, the elasticity of consumption to income will be $(\gamma_0 + \gamma_1 + \gamma_2)/(1 - \beta_1 - \beta_2)$. On the Permanent Income Hypothesis, this ratio should be equal unity, and this constitutes a test of that model within this framework. In (29) it is immediately testable through the coefficient of $\ln Y_{t-2}$. The premise of the ECM approach to applied econometric modelling is that long run proportionality relationships of this sort (with or without unit coefficients of proportionality) are features that we should expect. If we model entirely in terms of differenced variables, as in the influential Box–Jenkins approach, we cannot say anything about long run relationships; and it is likely that, if there are such relationships, our models will forecast very poorly as the forecast horizon is extended.

The claim that there generally exist long-run relationships between variables may now be recognized as the claim that economic variables are *co-integral* (Engle and Granger 1985). The statistical theory of co-integral processes postdates much of Hendry's work on ECMs, which is an example of application preceding theory. A variable x is said to be $I(n)$ (integral of order n) if its nth difference $\Delta^n x$ is *weakly stationary* (i.e. if its mean and variance are constant over time). In general, provided we work in logarithms, economic variables are either $I(1)$ or $I(0)$. Consumption and income both grow over time and thus cannot have constant means; we would therefore expect $\ln C$ and $\ln Y$ to be $I(1)$. If two variables x and y are both $I(1)$, then in general any combination of these variables, say $x-\alpha y$, will also be $I(1)$. However, there may exist a singularity, say α^*, such that $x-\alpha^* y$ is $I(0)$. If such a singularity does exist, x and y are co-integral. This implies that, in the long run, although x and y can be arbitrarily high or low, they must be proportional to each other with factor of proportionality α^*. The Permanent Income Hypothesis implies exactly this sort of proportionality relationship.[31] Moreover, it is arguable that much of economic theory consists of comparative static results which allow comparison of alternative equilibrium positions or paths, and which have little implication for dynamic adjustment. On this view, restrictions should typically be imposed on the equilibrium solutions of econometric equations, and this is facilitated by the ECM formulation.

The appeal of the ECM formulation is that it combines flexibility in dynamic specification with (apparently) desirable long run properties. Engle and Granger (1985) have shown that, if two or more variables are co-integral, there must exist an ECM linking these variables. Hendry's applied work, both on the consumption-income relationship, and on money demand,[32] has assumed cointegration. Recent work on methods for testing for unit roots have brought us to the position where we can test for cointegration prior to embarking on the modelling exercise. The generality of the ECM will become clear once such tests are routinely performed.[33]

This account of the ECM leaves open the economic mechanisms which generate cointegration. A standard way to obtain models of this kind is by invoking *adjustment costs*. Hendry and von Ungern-Sternberg (1981) provide a particularly simple example of this. Consumers have a long-run desired consumption-income ratio K and a long-run desired asset-income ratio of H (which clearly cannot be independent of K). They minimize a myopic loss function[34]

$$\Lambda_t = \lambda_1(a_t - y_t - h)^2 + \lambda_2(c_t - y_t - k)^2 + \lambda_3(c_t - c_{t-1})^2 \qquad (30)$$

where $h=\ln H$, $k=\ln K$, $c_t=\ln C_t$, $y_t=\ln Y_t$ and $a_t=\ln A_t$, and where A_t asset holdings satisfy

$$A_t = A_{t-1} + (Y_t - C_t) \qquad (31)$$

[31] Long run proportionality is dropped in Hendry (1983a).

[32] Hendry and Mizon (1978), Hendry (1979), Hendry and Ericsson (1983).

[33] Fuller (1976, pp. 366–82), Dickey and Fuller (1981), Sargan and Bhargava (1983).

[34] Hendry and von Ungern-Sternberg include an additional term $\lambda_4(C_t - C_{t-1})(Y_t - Y_{t-1})$ in their loss function, but this does not alter the form of the control solution (32), although it removes the restriction that $\theta_1 = \theta_2 + \theta_3$.

The adjustment cost component of Λ_t is given by the term $\lambda_3(C_t - C_{t-1})^2$. The adjustment path implied by this formulation may be written in the error correction form

$$\Delta c_t = \theta_0 + \theta_1 \Delta y_t - \theta_2(c_{t-1} - y_{t-1}) + \theta_3(a_{t-1} - y_{t-1}) \tag{32}$$

where the θ_i parameters are functions of the λ_i parameters of (30) together with h and k.

The objective function specified in (30) is for a single period. More generally one would expect agents to minimize a loss function specified as

$$\Lambda_t = \sum_{i=0}^{\infty} \delta^i \Lambda_{t+i} \tag{33}$$

where $\delta = 1/(1+\rho)$ and ρ is the discount rate. The solution of this model is somewhat more complicated since planned consumption in period t must depend on planned consumption in period $t+1$, and so forth. Nickell (1985) has provided a systematic analysis of models of this sort.

Suppose that the individual's target consumption in period t is $C_t^* = KY_t$, and that he minimizes a loss function

$$\Lambda = \sum_{i=0}^{\infty} \delta^i [\lambda(c_{t+i} - c_{t+i}^*)^2 + (c_{t+i} - c_{t+i-1})^2] \tag{34}$$

The first term in (34) is the penalty from having an consumption level differing from one's target; the second term is the adjustment cost penalty. Nickell shows that the solution to this minimization problem gives as optimal consumption level for period t

$$\Delta c_t = (1-\mu)\left[(1-\delta\mu) \sum_{i=0}^{\infty} (\delta\mu)^i c_{t+i}^* - c_{t-i}\right] \tag{35}$$

where $0 < \mu < 1$. Note the absence of any term in Δc_t^* in (35). As it stands, equation (35) defines a partial adjustment model and not an ECM. The change in the current consumption level is the fraction $1-\mu$ of the difference between the weighted sum of future target consumption levels and the lagged level. With static income expectations, (35) simplifies to the familiar:

$$\Delta c_t = (1-\mu)(k + y_t - c_{t-1}) \tag{36}$$

In general, however, income expectations would not reasonably be static. Indeed, under rational expectations, the relationship between the current expectation of future income levels and current and past income realizations will depend on the actual income generation process. It turns out that the most satisfactory simple approximation to such processes is obtained by supposing that income follows a second order autoregression with drift (i.e. that the difference in income follows a first order autoregression with a non-zero intercept):

$$\Delta y_t = \frac{g}{(1-\theta)} + \theta \Delta y_{t-1} + \epsilon_t \quad (37)$$

where ϵ_t is white noise. Forming rational expectations of future income levels using (37) and substituting into (35) gives an ECM consumption specification

$$\Delta c_t = \frac{1-\mu}{1-\delta\mu}\delta\mu g + \frac{1-\mu}{1-\theta\delta\mu}\Delta y_t + (1-\mu)(k + y_{t-1} - c_{t-1}) \quad (38)$$

Hence the ECM may be expected to arise from partial adjustment of consumption, money stock etc., to desired values, in conjunction with rational expectations and plausible income generating processes. There may, of course, be other routes, but adjustment costs and second order autoregressions occur sufficiently widely in economics to suggest a wide application for this class of model.

VI. CONCLUSIONS

David Hendry's views on econometrics may be summarized as follows:

(1) The data generating process underlying a data set is never known. Applied econometrics is therefore in part a process of discovery, and all proposed models must be presumed to be misspecified.

(2) Applied economists should aim to provide structural models which may be taken as approximations to the unknown data generating process.

(3) Restrictions derived from economic theory will typically relate to the equilibrium solutions to these models, rather than to the short term dynamics. Such restrictions may frequently be imposed and tested within an error correction framework. The generality of this model follows from the fact that many important economic series are cointegral.

(4) Any proposed model must be tested within and outside the sample. One needs to ask whether the model adequately characterizes the data, and also whether this characterization is superior to that provided by rival specifications.

(5) Equation specification should be conducted by testing alternative simplifications against a common nesting maintained hypothesis. Tests may be carried out using standard classical statistical procedures, or straightforward generalizations of these procedures.

To a large extent, these views are shared with other practicing British econometricians, and are part of a tradition which originated at the London School of Economics in the 1950's and 1960's. In Gilbert (1986), I argued that this tradition derives its power from the combination of the Cowles emphasis on structural modelling with statistical time series analysis methodology. In the United States, by contrast, these two traditions have tended to remain distinct, with the result that structural models compete with VAR (vector autoregression) formulations without any middle ground.

Because much recent British applied econometrics, and in particular Hendry's contributions, differ considerably in style from American studies on the same topics, there is a tendency to suppose that the British approach is, in some sense, extreme. The foregoing analysis suggest that, on the contrary, David Hendry's econometric methodology constitutes a sensible middle ground. The methodology is grounded in classical statistical methods but, at the same time, makes sense of the prevalence of properly conducted specification searches in econometric practice. It rationalizes much of what econometricians have always done, but also provides guidance as to how these things can be done more efficiently. The most important test of any methodology is, however, the applied research that it generates. In this respect, the accumulating body of published articles which utilize the methods discussed in this paper establish this methodology as among the most powerful of the current approaches to applied economics.

Institute of Economics and Statistics,
University of Oxford

Date of Receipt of Final Manuscript: May 1986

REFERENCES

Amemiya, T. (1980). 'Selection of Regressors', *International Economic Review*, Vol. 21, pp. 331-54.

Currie, D. (1981). 'Some Long Run Features of Dynamic Time Series Models', *Economic Journal*, Vol. 91, pp. 704-15.

Davidson, J. E. H. and Hendry, D. F. (1981). 'Interpreting Econometric Evidence: the Behaviour of Consumers' Expenditure in the UK', *European Economic Review*, Vol. 16, pp. 177-592.

Davidson, J. E. H., Hendry, D. F., Srba, F. and Yeo, S. (1978). 'Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the UK', *Economic Journal*, Vol. 88, pp. 661-92.

Dickey, J. A. and Fuller, W. A. (1981). 'Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root', *Econometrica*, Vol. 49, pp. 1057-72.

Engle, R. F. and Granger, C. W. H. (1985). 'Dynamic Model Specification with Equilibrium Constraints: Cointegration and Error Correction', paper presented to the Econometric Society Fifth World Congress, MIT.

Engle, R. F., Hendry, D. F. and Richard, J-F. (1983). 'Exogeneity', *Econometrica*, Vol. 51, pp. 277-304.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, New York, Wiley.

Gilbert. C. L. (1986). 'The Development of British Econometrics 1945-85', University of Oxford, *Applied Economics Discussion Paper*, no. 8.

Hall, R. E. (1978). 'Stochastic Implications of the Life Cycle-Permanent Income Hypotheses: Theory and Evidence', *Journal of Political Economy*, Vol. 86, pp. 971-87.

Harvey, A. C. (1981). *The Econometric Analysis of Time Series*, Deddington, Philip Allan.

Hendry, D. F. (1979). 'Predictive Failure and Econometric Modelling in Macroeconomics: the Transactions Demand for Money', in Ormerod, P. (ed.), *Modelling the Economy*, London, Heinemann.

Hendry, D. F. (1983). 'Econometric Modelling: the Consumption Function in Retrospect', *Scottish Journal of Political Economy*, Vol. 30, pp. 193–220.

Hendry, D. F. (1985). 'Econometric Methodology', paper presented to the Econometric Society Fifth World Congress, MIT.

Hendry, D. F. and Ericsson, N. R. (1983). 'Assertion Without Empirical Basis: an Econometric Appraisal of "Monetary Trends in ... the United Kingdom" by Milton Friedman and Anna Schwartz', in Bank of England Academic Consultants, *Monetary Trends in the United Kingdom*, paper no. 22, pp. 45-101.

Hendry, D. F. and Mizon, G. E. (1978). 'Serial Correlation as a Convenient Simplification, not a Nuisance: a Comment on a Study of the Demand for Money by the Bank of England', *Economic Journal*, Vol. 88, pp. 549-63.

Hendry, D. F., Pagan, A. R. and Sargan, J. D. (1984). 'Dynamic Specification', ch. 18 of Griliches, Z. and Intriligator, M. D. (eds.) *Handbook of Econometrics*, Vol. 2, Amsterdam, North-Holland.

Hendry, D. F. and Richard, J-F. (1982). 'On the Formulation of Empirical Models in Dynamic Econometrics', *Journal of Econometrics*, Vol. 20, pp. 3-33.

Hendry, D. F. and Richard, J-F. (1983). 'The Econometric Analysis of Economic Time Series', *International Statistical Review*, Vol. 51, pp. 111-63.

Hendry, D. F. and von Ungern-Sternberg, T. (1981). 'Liquidity and Inflation Effects on Consumers' Expenditure', in Deaton, A. S. (ed.), *Essays in the Theory and Measurement of Consumer Behaviour*, Cambridge, Cambridge University Press.

Hendry, D. F. and Wallis, K. F. (eds.) (1984). *Econometrics and Quantitative Economics*, Oxford, Basil Blackwell.

Judge, G. G. and Bock, M. E. (1978). *The Statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics*, Amsterdam, North-Holland.

Lakatos, I. (1970). 'Falsification and the Methodology of Scientific Research Programmes', in Lakatos, I. and Musgrave, A. (eds.), *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press.

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Non-Experimental Data*, New York, Wiley.

Leamer, E. E. (1983). 'Let's Take the Con Out of Econometrics', *American Economic Review*, Vol. 73, pp. 31–43.

Lucas, R. E. (1976). 'Econometric Policy Analysis: A Critique', in Brunner, K. and Meltzer, A. H. (eds.), *The Phillips Curve and Labor Markets*, (*Journal of Monetary Economics*, Supplement, Vol. 1, pp. 19–46).

McAleer, M., Pagan, A. R. and Volker, P. A. (1985). 'What will Take the Con Out of Econometrics?', *American Economic Review*, Vol. 75, pp. 293-307.

Mizon, G. E. (1977). 'Model Selection Procedures', in Artis, M. J. and Nobay, A. R. (eds.), *Studies in Modern Economic Analysis*, Oxford, Basil Blackwell.

Mizon, G. E. (1984). 'The Encompassing Principle in Econometrics', in Hendry and Wallis (1984), pp. 135-72.

Mizon, G. E. and Hendry, D. F. (1980). 'An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification', *Review of Economic Studies*, Vol. 47, pp. 21-46.

Mizon, G. E. and Richard, J-F. (1985). 'The Encompassing Principle and its Application to Non-Nested Hypotheses', *Econometrica*, Vol. 54, pp. 657-78.

Nickell, S. (1985). 'Error Correction, Partial Adjustment and All That: an Expository Note', *BULLETIN*, Vol. 47, pp. 119-31.

Phillips, A. W. (1954). 'Stabilization Policy in a Closed Economy', *Economic Journal*, Vol. 64, pp. 290-323.

Sargan, J. D. (1964). 'Wages and Prices in the United Kingdom: a Study in Econometric Methodology', in Hart, P. E., Mills, G. and Whittaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, London, Butterworths; reprinted in Hendry and Wallis (1984), pp. 275-314.

Sargan, J. D. (1980). 'Some Tests of Dynamic Specification for a Single Equation', *Econometrica*, Vol. 48, pp. 879-97.

Sargan, J. D. and Bhargava, A. (1983). 'Testing Residuals from Least Squares Regression for being Generated by the Gaussian Random Walk', *Econometrica*, Vol. 51, pp. 153-74.

Surrey, M. J. C. (1974). *An Introduction to Econometrics*, Oxford, Oxford University Press.

Trivedi, P. K. (1984). 'Uncertain Prior Information and Distributed Lag Analysis', in Hendry and Wallis (1984), pp. 173-210.

Wonnacott, R. J. and Wonnacott, T. H. (1970). *Econometrics*, New York, Wiley.