The Economic Journal, **106** (*September*), 1398–1413. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA.

REVIEW ARTICLE

METHODOLOGY: ALCHEMY OR SCIENCE?*

Bruce E. Hansen

In *Econometrics: Alchemy or Science?* David Hendry has collected 19 of his best papers. As described by the author, 'This is a collection of my main essays on econometric methodology from the period 1974–85 during which the approach developed into its present form...' One of Hendry's passions is methodology, and *Econometrics: Alchemy or Science?* is an attempt to show how it developed over a fifteen-year period during which he struggled to produce a new empirical research programme.

The papers selected for the volume exposit one aspect of Hendry's methodology or another: some directly through exhortation, and others indirectly by example. The volume allows the reader to see how the ideas developed and changed over the period. In addition, each chapter is prefaced by a brief preamble in which Hendry describes the genesis of the paper, its relationship to the other papers in the volume, and what flaws he sees in the analysis (looking back with hindsight). These preambles are an ingenious innovation, and are quite helpful at keeping the reader focused on the theme of the methodology and its development.

Students of econometric methodology will find the volume valuable as a history of Hendry's personal methodology. Students of time series econometrics will also find the volume valuable as a collection of excellent research papers.

In the next section of this paper, I describe the contents of the volume in detail. In Section II, the subject matter – Hendry's methodology itself – is reviewed and critiqued. A word of caution may be appropriate at the outset. Any attempt to build a 'methodology' is bound to 'fail', simply because the goal is quite ambitious. While Hendry's methodology may be one of the most successful and attractive methodologies extant (the other current methodologies might include those of Leamer, Sims, and Prescott), the bottom line is that each fails the test of providing a coherent and complete set of guidelines for empirical research. Hence the title of this review article. A correct 'methodology' might indeed lead us to wisdom and science, but a false one might be alchemy.

I. CONTENTS OF 'ALCHEMY'

For a complete list of the papers published in *Econometrics: Alchemy or Science?* see the Appendix.

* I thank Adrian Pagan and three referees for very helpful comments on an earlier draft. I also thank the National Science Foundation and Sloan Foundation for research support.

I.1. Part I: Roots and Route Maps

Econometrics: Alchemy or Science? is divided into four parts, with Part I setting the stage. It contains two early papers (the 'Roots') in which Hendry begins to struggle with the general issues of model specification, and two later papers (the 'Route Maps') which summarise important aspects of the methodology.

The first paper gives the volume its name, and is based on Hendry's inaugural lecture at the London School of Economics. In my opinion, it is one of the best-written papers in the history of econometrics. It gracefully takes the audience through the hazards and pitfalls of nonsense regressions, complete with an elegant portrayal of rainfall as a better predictor of prices than money (!), and then shows how rigorous econometric testing can eliminate these hazards. Econometrics is alchemy since econometricians can create nearly any result desired, but it is also science because econometricians also know how to reject and avoid spurious models. Hence his exhortation that 'the three golden rules of econometrics are test, test and test.' While published back in 1980, this paper displays a remarkable understanding of late-1980s time-series theory, such as the implications of unit roots, spurious regressions, cointegration and error-correction.

The second and third papers are early attempts at econometric modelling. Hendry includes them to illustrate the connections and differences between his first attempts at empirical research and his later mature research. These early papers are more conventional in the sense that they take a theoretical specification more-or-less at face value. In this early work, Hendry also moved from a specific-to-general (StoG) modelling strategy to a general-to-specific (GtoS) approach.[1] The GtoS modelling strategy eventually became a pillar of the methodology.

The issues of dynamic specification are laid in detail in the fourth paper, taken from the *Handbook of Econometrics*. This chapter is a clear description of this aspect of the methodology. The authors show how most common time series equations are special cases of a general dynamic relationship, so that these specifications can be *tested* against the general alternative. This provides a theoretical case for the GtoS approach.

I.2. Part II: The Development of Empirical Modelling Strategies

Part II contains the core material of the volume. These papers, published (with one exception) in the narrow period 1977–81, document the development of Hendry's methodological views during the mid-to-late 1970s. The main issues are: (1) Time-series versus econometrics; (2) Serial correlation and common factors; (3) General to specific modelling; and (4) Predictive failure. The two primary economic examples explored are aggregate consumption and money demand.

[1] This move appears to have been influenced by Box and Jenkins (1970), who introduced a general-to-specific approach for univariate time series models.

The sixth paper of the volume, 'Serial Correlation as a Convenient Simplification, not a Nuisance', is particularly insightful. The authors drive home the simple idea that serially correlated errors are a special case of a general dynamic relationship, so that the former specification can (and should!) be tested against the latter. While this idea is not original to Hendry, his repeated and consistent attacks on naive dynamic specifications have helped to move the profession away from autoregressive error modelling to unrestricted dynamic models. Indeed, for many applied econometricians, the starting point is now an unrestricted VAR.

A major turning point in Hendry's methodological approach came with the eighth paper, often called DHSY after the authors: Davidson, Hendry, Srba and Yeo. In this paper, Hendry assembled the basic concepts which were to become the Hendry methodology. These involved general-to-specific modelling, error-correction, conditioning, and parameter constancy. The empirical relationship discovered in DHSY appeared to survive relatively unscathed into the 1980s, a remarkable achievement for an empirical time-series equation. It may not be too unfair to suggest that much of Hendry's subsequent formal work was an attempt to justify the somewhat *ad hoc* approach followed by DHSY. This included the development of the concept of encompassing – the need for a chosen model to explain the findings of rival empirical models.

In his introduction, Hendry describes the ideas behind the eleventh paper as a 'Gestalt Shift'. Rather than viewing the methods of DHSY as particular to that individual paper, Hendry started to view the modelling tools as generic. This paper explores a money demand equation using a GtoS modelling strategy. The issues of predictive failure and error correction play important roles in this work, illustrating that they had taken a firm ground in Hendry's vision by the late 1970s.

I.3. Part III: Formalisation

Even though Hendry characterises himself as an applied (rather than a theoretical) econometrician, he has made important contributions to econometric theory, and some of this work (five papers) is presented in the third part of the volume. In 'Exogeneity', written jointly with Robert Engle and Jean-François-Richard, Hendry introduces three new definitions of exogeneity. 'On the Formulation of Empirical Models in Dynamic Econometrics' (written with Jean-François-Richard) formalises several of the ideas inherent in Hendry's methodology, and carefully outlines the idea of encompassing – that a good model must be able to 'explain' the empirical findings of other models. 'The Econometric Analysis of Economic Time Series' (also written with Richard) is an accessible review of the methodology.

I.4. Part IV: Retrospect and Prospect

The volume closes on a rather unusual note. The final chapter is an edited version of the manual to PC-GIVE, Hendry's econometric software package. The basics of the methodology are summarised in the context of econometric estimation.

II. HENDRY'S METHODOLOGY

Hendry uses *Alchemy* to showcase his methodology. It is therefore appropriate in this review article to present a critical appraisal of his methodology. Based on my reading, I believe the methodology has four main components. First, Hendry argues that any empirical model should satisfy a set of specific *criteria*, which he labels: (1) Data Coherency; (2) Valid Conditioning; (3) Parameter Constancy; (4) Data Admissibility; (5) Theory Consistency; and (6) Encompassing. Secondly, the appropriate method to discover the correct model is through general-to-specific (GtoS) modelling. Thirdly, empirical research should be *progressive*, in the sense that research should build on past empirical discoveries. These three components are interrelated, as the GtoS selection approach is intended to produce empirical models which are data coherent, and encompassing is a criteria designed to ensure progressive research. The fourth component which I assign to the Hendry methodology is perhaps more controversial, as you will find no clear statement of it in *Alchemy*. It is a *preference* for linear conditional models with homoskedastic Gaussian errors and constant coefficients. While Hendry does not make this preference an official part of his methodology, I believe that it emerges as an unofficial component through his empirical work in practice.

We will discover as we investigate Hendry's methodology that there is tension between these goals. Most notably, we will find that it is impossible to adhere rigorously to GtoS modelling. Hendry's practical solution is to internalise the preference for linear conditional models with homoskedastic Gaussian errors and constant coefficients, setting up models in this class as his 'general' specification. Tests of these assumptions imply specific-to-general (StoG) modelling, violating the second component of the methodology. There is no clear solution to this dilemma, but it points out an important limitation with the Hendry methodology as a general guideline for empirical research.

II.1. Model Selection and Data Mining

The subsequent sections will outline a set of criteria which Hendry believes that empirical models should satisfy. Several can be assessed by statistical tests, and Hendry emphasises that such tests should be used to determine whether a given model satisfies the stated criteria. Thus his methodology implies a search across models, where each is tested by a variety of statistical procedures. Only those models which pass the battery of tests can be considered as valid empirical representations.

The fact that Hendry openly acknowledges that data-based searches are a necessary part of empirical research has been quite controversial, as it conflicts with traditional statistical methodology as presented in our major textbooks. Many economists pejoratively call this aspect of the methodology 'data mining' – the view is that since a specification search gives the researcher the ability to select a model which passes the desired statistical tests, the end results will be biased. Hendry counters this objection with two arguments. First, he does not claim to test the model by the criteria used to select the model. If an econometrician reports test statistics for criteria used in model selection, Hendry argues that the tests demonstrate only that indeed the model meets these criteria, *not* that the model is adequate. True testing will require new data or new test criteria. Secondly, in Hendry's own words, 'The validity of a model is a property of the model in relation to the evidence and so cannot be affected by how that model is selected.' Thus specification searches are okay.

I have never been completely persuaded by the arguments of either Hendry or his detractors. On the one hand, it would be dishonest to claim that specification searches have no role in econometric research. Economic theory is not sufficiently rich to fully specify the details of most empirical models. On the other hand, it is unclear how to interpret conventional tests used at the end of a specification search, since the distribution theory is derived under the classical paradigm of no specification search. Furthermore, while Hendry is obviously correct that the 'validity' of any given model cannot be affected by a specification search, the relevant questions are whether a specification search will lead us closer to the truth, and whether a specification search will distort the usefulness of test statistics used to assess the 'validity' of a selected model.

After some reflection, I believe that we can partially defend Hendry's position with a formal argument. Let $\psi \in \Psi$ index the set of models under consideration, and let ψ_0 denote the true model. Suppose that for any model ψ, we have some test statistic $\eta_T(\psi)$ with a classical 5 % critical value C. By the latter I mean that $P\{\eta_T(\psi_0) > C\} = 0{\cdot}05$.[2]

A model ψ is 'Hendry-admissible' (my label) if the model passes the statistical test; in our notation this occurs if $\eta_T(\psi) < C$. In general, there might be more than one model ψ which is Hendry-admissible; we can denote this set by $\Psi_T^* = \{\psi \in \Psi : \eta_T(\psi) < C\}$. My interpretation of the Hendry methodology is that only models which are elements of Ψ_T^* should be reported or used in practice. A specification search is typically necessary to discover the class Ψ_T^*, and it is this specification search which disturbs Hendry's critics. If a model $\hat{\psi}$ is selected so that it satisfies $\eta_T(\hat{\psi}) < C$, what is the statistical interpretation? Is $\hat{\psi}$ biased in some sense?

By thinking of Ψ_T^* as a class, we can give it a statistical interpretation. We can say that Ψ_T^* has coverage probability β if $P\{\psi_0 \in \Psi_T^*\} = \beta$. Since $\psi_0 \in \Psi_T^*$ if and only if $\eta_T(\psi_0) < C$, we see that

$$P\{\psi_0 \in \Psi_T^*\} = P\{\eta_T(\psi_0) < C\} = 0{\cdot}95, \qquad (1)$$

so Ψ_T^* has coverage probability 95 %. Thus there is no 'bias' in the statistical test implied by the specification search. We conclude that so long as the Hendry methodology simply divides models into 'Hendry-admissible' and 'Hendry-non-admissible', then the reporting of conventional critical values is appropriate. The difficulty, of course, is that the full set Ψ_T^* is not actually reported; only one element is typically reported by a single researcher. Perhaps the proper view is that the set Ψ_T^* will be eventually discovered by the

[2] We have simplified the analysis a bit by assuming that the criteria can be assessed by a single test statistic; in reality a set of test statistics are used, but this abstraction does not appear to be essential to the analysis, and is not unique to the present debate.

profession over time, as each researcher reports their own model, only constrained by the requirement that it satisfies the condition of 'Hendry admissibility'.

What happens to Ψ_T^* as $T \to \infty$? Let Ψ^* denote its limit. It should not include any model ψ for which $\eta_T(\psi)$ has asymptotic power. It may be possible to design tests η_T which can reject all models except the true model. If this is the case, then Ψ^* will consist only of the true model ψ_0. If ψ_0 is not an element of Ψ, then Ψ^* will be the empty set.

II.2. Data Coherence

A cornerstone of the Hendry methodology is that the chosen model must adequately yet parsimoniously describe the conditional distribution of the dependent variable. Throughout his career, Hendry has warned of the dangers of incomplete specification, and has recommended that an empirical model be able to pass a test against a general alternative. This is one of the great strengths of the Hendry methodology, for it forces researchers to take the data seriously.

There are two levels to this part of the methodology. The first is the general prescription that all models must match the conditional distribution. The second is a set of specific ideas which will help lead to correct formulations of dynamic econometric models. On the first issue there is undoubtedly general agreement in the econometrics profession (and probably always has been). Hendry should be given particular credit, however, for consistently emphasising the issue in his research and teaching.

The second issue of relevance are Hendry's specific suggestions to deal with issues of dynamic specification. There are two suggestions of note. The first, which occupies much space in *Alchemy*, is the suggestion of replacing autoregressive error specifications by general dynamics. Simply put, a traditional autoregressive error specification takes the form

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t, \quad u_t = \rho u_{t-1} + e_t \tag{2}$$

with e_t iid$(0, \sigma^2)$. An unrestricted dynamic relationship takes the form

$$y_t = \alpha y_{t-1} + \mathbf{x}_t' \boldsymbol{\beta}_1 + \mathbf{x}_{t-1}' \boldsymbol{\beta}_2 + e_t. \tag{3}$$

As now discussed in most econometrics textbooks, (2) is a special case of (3), obtained by the restriction $\boldsymbol{\beta}_2 = -\boldsymbol{\beta}_1 \alpha$. In his early work, Hendry advocated testing this restriction. In his later work, he simply moved away from models such as (2), since this restriction is seldomly satisfied in actual empirical practice.

It appears that on this point, conventional wisdom has emerged with Hendry's views. Autoregressive error specifications such as (2) are hard to find in top-flight econometric work.

Hendry's second important suggestion was to make extensive use of error-correction models (ECMs). While the idea of ECMs is present in the work of Phillips (1954) and Sargan (1964), it was Hendry and his co-authors (especially DHSY) who pushed the ECM specification to the front stage of econometric practice. Even before the concept of cointegration had been invented, Hendry correctly understood that an ECM merged short-run dynamic specifications with long-run restrictions. Furthermore, Granger's (1981) invention of cointegration was partly motivated by Hendry's work.

To close our discussion, let us turn to the broader issue of the class of relevant models. When I read the papers in *Alchemy* I am struck by the repeated focus on linear models with normal errors. No doubt this was typical of applied research of the 1970s and early 1980s. While not central to Hendry's vision in the general sense, the normal linear model appears important to the implementation of Hendry's methodology in empirical practice. Only within the class of linear normality can we restrict attention to a sufficiently narrow class of models so that we can meaningfully entertain the notion of 'testing one's model against the general alternative'.

This is not a trivial point. Today, time series econometricians are experimenting with an extremely large class of models. Latent variables, non-linearities, complex interactions and structural change are routinely considered. Hendry's methodology suggests that these researchers should each test their model against 'the' general alternative, in order to guarantee that their model is data coherent. But this is not possible, since the general alternative is far too broad to conceivably write down, let alone estimate. We are forced to restrict attention to limited tests against specific alternatives. This violates a central dictum of the Hendry methodology, and calls into question its usefulness as a general guideline for empirical practice.

II.3. Valid Conditioning

An important paper in *Alchemy* is 'Exogeneity', written with Robert Engle and Jean-François-Richard. In this paper, the authors (EHR) clarify the concept of exogeneity, focusing in particular on dynamic models. Their discussion makes use of the following concepts. Suppose there are a pair of random vectors $(\mathbf{Y},\mathbf{X})$ with the joint distribution $J(\mathbf{Y},\mathbf{X})$, with $F(\mathbf{Y}|\mathbf{X})$ the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$, and $G(\mathbf{X})$ the marginal distribution of $\mathbf{X}$. Let β and λ denote the parameters of F and G, respectively.

I believe that the authors make three important contributions. First, they argue that the purpose of exogeneity is to know whether it is appropriate to model the conditional distribution $F(\mathbf{Y}|\mathbf{X})$ alone rather than the joint distribution $J(\mathbf{Y},\mathbf{X})$ (which is equivalent to jointly modelling $F(\mathbf{Y}|\mathbf{X})$ and $G(\mathbf{X})$). Thus exogeneity is not merely about estimate consistency – while a violation of exogeneity may lead to inconsistent estimates for parameters of interest, it is not intrinsic to the concept.

Secondly, EHR point out that 'exogeneity' is a concept which pertains to parameters, not dependent variables. That is, it is not meaningful to ask whether $\mathbf{X}$ is exogenous for $\mathbf{Y}$, for the answer will depend on the definition of β. Instead, one may ask whether $\mathbf{X}$ is exogenous for some parameter γ. Indeed, the authors define $\mathbf{X}$ to be *weakly exogenous* for γ if γ is a function of β alone, and β and λ are variation free (meaning that β can be varied without imposing constraints on λ, and *vice-versa*). If this condition is satisfied (along with regularity conditions) then β can be efficiently estimated from the conditional likelihood of $\mathbf{Y}$ given $\mathbf{X}$, which is a function of F and not G.

Thirdly, EHR clarify the relationship between exogeneity and causality, which is especially important in a time series context. The definition of weak exogeneity given above does not impose any restrictions on the nature of causality between **Y** and **X**, but an exercise which attempts to forecast **Y** given fixed values of **X** requires that **Y** does not Granger-cause[3] **X**. If this additional requirement is satisfied, then EHR define **X** to be *strongly exogenous* for γ.

The authors also have a final concept. If **X** is weakly exogenous for γ, and γ is invariant to changes in the marginal distribution of **X** (i.e. λ), then **X** is said to be *super exogenous*. This condition is important for policy analysis, where the goal is to consider the effect of changes in **X** upon the distribution of **Y**.

While a clarified set of definitions of exogeneity is useful for theory, it appears that the concepts introduced in 'Exogeneity' have had very little effect upon econometric practice, including Hendry's own work. This is for two reasons. First, EHR's definition of weak exogeneity is not particularly useful in linear models. This is because the conventional definition is equivalent (and somewhat easier to understand). Consider the simple example

$$y_t = x_t \beta + e_{1t}, \quad x_t = e_{2t},$$

where

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right],$$

independent across t. Conventionally, we would say that x_t is exogenous for β if $\mathrm{E}(x_t e_{1t}) = 0$, and it is easy to see that this holds if and only if $\rho = 0$. To use the EHR definition, we need to construct the conditional distribution of y_t given x_t, which is $\mathrm{N}(x_t \beta^*, v^2)$, where $\beta^* = \beta + \sigma\rho$ and $v^2 = \sigma^2(1-\rho^2)$. Since only β^* and v^2 are identified from the conditional distribution, β is identified only when $\rho = 0$, and thus EHR define x_t as weakly exogenous for β only when $\rho = 0$, which is equivalent to the conventional definition.

A second reason why 'Exogeneity' has had only a small impact on econometric practice is that no new test statistics were introduced. Tests for weak exogeneity,[4] Granger causality and parameter constancy were known (for linear models) before the paper was written.

At a deeper level, I worry that there is a divergence in the Hendry methodology between theory and practice. The theory says that we should work with conditional models only when the appropriate concept of exogeneity holds for the parameters of interest. But in practice, Hendry routinely works with linear equations with contemporaneous variables on the right-hand-side; for example, in Hendry's equations for the growth rate in consumption and the growth rate in the money stock, he includes on the right-hand-side the growth rate in income, which would be typically considered endogenous in the economics literature. Despite the obvious endogeneity problem, Hendry estimates his equations by OLS. Occasionally, he checks the validity of the implicit weak exogeneity assumption by reestimating using lagged variables as

[3] **Y** does not Granger-cause **X** if the conditional distribution of X_t, conditional on $\{X_{t-1}, X_{t-2}, \ldots\}$ and $\{Y_{t-1}, Y_{t-2}, \ldots\}$, does not depend on $\{Y_{t-1}, Y_{t-2}, \ldots\}$.

[4] Wu (1973) and Hausman (1978).

instruments, but this is neither his typical practice nor consistent with his views on general-to-specific (GtoS) modelling. Since a model with weakly exogenous contemporaneous regressors is a restriction on the class of models considered, the GtoS method suggests starting with the joint model, and working towards the conditional, rather than starting with a conditional model as typical in Hendry's work. Instead, Hendry's empirical practice is simple-to-general (StoG). Perhaps this is a sensible position in this context, but it is inconsistent with his methodological principles.

Recently, Hendry introduced a novel 'test' for weak exogeneity. If the parameters of the marginal distribution (λ in the above notation) have changed within sample, but the parameters of the conditional distribution (β in our notation) have remained constant, it appears that λ and β are variation free, and hence the regressors are weakly exogenous for β. From this argument, Hendry claims that a test for parameter constancy in β, given that λ has changed, can be interpreted as a test for weak exogeneity.

Unfortunately, the claim is false. Indeed, the test does not provide information about whether the parameters of the conditional distribution are actually the structural parameters of interest (which is the key issue in discussions about exogeneity!) This can be seen in a simple constructed example. Suppose that a structural equation of interest is

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + e_t, \tag{4}$$

where y_t is real-valued and $\mathbf{x}_t$ is $k \times 1$. The marginal distribution of $(\mathbf{x}_t, e_t)$ is

$$\begin{pmatrix} e_t \\ \mathbf{x}_t \end{pmatrix} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 + \boldsymbol{\gamma}'\boldsymbol{\Omega}\boldsymbol{\gamma} & \boldsymbol{\gamma}'\boldsymbol{\Omega} \\ \boldsymbol{\Omega}\boldsymbol{\gamma} & \boldsymbol{\Omega} \end{pmatrix}, \tag{5}$$

where σ^2 is real-valued, $\boldsymbol{\gamma}$ is a $k \times 1$ vector and $\boldsymbol{\Omega}$ is a $k \times k$ matrix. If $\boldsymbol{\gamma} \neq \mathbf{0}$, $\mathbf{x}_t$ is not weakly exogenous for $\boldsymbol{\beta}$, and OLS estimation of (4) will be inconsistent for $\boldsymbol{\beta}$.

From (5) we can derive the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$, and the marginal distribution of $\mathbf{X}$. Since $e_t \mid \mathbf{x}_t \sim \mathrm{N}(\mathbf{x}_t'\boldsymbol{\gamma}, \sigma^2)$, the conditional distribution of y_t given $\mathbf{x}_t$ is

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}^* + u_t, \quad u_t \sim \mathrm{IN}(0, \sigma^2), \tag{6}$$

where $\boldsymbol{\beta}^* = \boldsymbol{\beta} + \boldsymbol{\gamma}$, and the marginal distribution of $\mathbf{x}_t$ is

$$\mathbf{x}_t \sim \mathrm{IN}(\mathbf{0}, \boldsymbol{\Omega}).$$

Now suppose that the parameters of the marginal distribution of $\mathbf{X}$ have changed, and all others remain constant. This means that $\boldsymbol{\Omega}$ has changed, but σ^2, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ remain constant. Since $\boldsymbol{\Omega}$ does not appear in the conditional distribution (which only depends on $\boldsymbol{\beta}^*$ and σ^2), the parameters of the conditional distribution will remain constant. It follows that statistical tests will support the hypothesis that the estimated model (6) is constant, and a researcher following Hendry's suggestion will erroneously infer that $\mathbf{x}_t$ is weakly exogenous for $\boldsymbol{\beta}$! The reason why Hendry's idea fails is that it forgets one of the central claims of 'Exogeneity' – that exogeneity is a concept which is relevant for parameters of interest, which need not be functions only of the parameters of the conditional distribution.

The only correct conclusion is that parameter constancy and exogeneity are distinct issues and require distinct tests. In the present case, the researcher needs to find a valid instrument – one which is correlated with $\mathbf{x}_t$ yet not with e_t – in order to find a valid test for the weak exogeneity for $\mathbf{x}_t$ for $\boldsymbol{\beta}$. This requires specification of a joint distribution of $(\mathbf{x}_t, e_t, \mathbf{z}_t)$, for which (5) is the marginal. If no such instrument $\mathbf{z}_t$ is available, then no test of weak exogeneity is possible.

II.4. Parameter Constancy

In his empirical work, Hendry emphasises the fact that only models with constant parameters are useful. This is an important point and cannot be over-emphasised. Hendry has found that parameter constancy tests are typically quite successful at detecting invalid specifications, as 'generic' economic relationships tend to be non-constant, rather than time invariant. Hendry's empirical work has been very creative in his attempt to detect non-constant parameters, hitting his models with a battery of formal and informal statistical tests. His econometric package PC-GIVE encourages the use of informal graphical analysis of predictive failure tests, and such graphs appear in much of Hendry's recent empirical work. When other attempts to eliminate non-constant parameters fail, he is willing to use intervention dummy variables to describe the remaining non-constancy. I applaud this concern with parameter stability, and believe that it has had an impact on empirical practice.

If we discover that an estimated model fails a parameter stability test, what then? One constructive approach would be to use a time-varying coefficient model such as

$$\left.\begin{aligned} y_t &= \mathbf{x}_t' \boldsymbol{\beta}_t + e_{1t}, \quad e_{1t} \sim \mathrm{N}(0, \sigma^2), \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \mathbf{e}_{2t}, \quad \mathbf{e}_{2t} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}). \end{aligned}\right\} \tag{7}$$

The model (7) is quite flexible, contains the conventional linear regression model as a special case, and properly accounts for coefficient uncertainty in forecasting exercises. Yet my reading of *Alchemy* is that the Hendry methodology would be opposed to this modelling strategy. Instead of modelling coefficients as time-varying, Hendry would recommend re-formulation of the model. There appears to be a belief that the *true* model will exhibit coefficient stability, so we should limit our search to models which possess this characteristic. The trouble of course is that model (7) is indeed a model with constant parameters – $(\beta_0, \sigma^2, \boldsymbol{\Sigma})$ – so it is not intrinsically inferior to a linear model with constant parameters. It would also be appropriate for Hendry to start with (7) and move towards a constant coefficient model, since Hendry strongly believes in GtoS modelling. It appears that there is no single, unified, correct approach to econometric modelling.

II.5. Theory Consistency

When discussing his methodology, Hendry argues that empirical models should be consistent with economic theory. In his applications, however, Hendry places a higher priority on data coherency than theory consistency. Since it is almost impossible to find a tightly specific macroeconomic model which is data coherent, there is an inherent need to recognise that there will be a trade-off between these goals in practice.

The distinguishing characteristic of Hendry's attitude towards the role of economic theory is his handling of 'long-run' or static information. Early on, Hendry recognised that error-correction models place restrictions on the steady-state behaviour of the variables, and that this was a useful method to incorporate at least some implications of economic theory. This work helped pave the way to the important literature on cointegration.

While many macroeconomists see Hendry's work as far too atheoretical, there is a large school of empirical time series analysts whose work is even farther removed from economic theory. Indeed, Hendry frequently argues that it is necessary to merge theory and practice in applied macroeconomics. In several of his papers, including the early 'On the Time-Series Approach to Econometric Model Building', Hendry rejects the call to abandon econometric modelling in favour of pure Box–Jenkins analysis. He argues that the Box–Jenkins models are more successful only because they have more complete dynamics. Hence econometric models could be made more successful if only they could add more dynamics while retaining the economic structure, for which he suggests error-correction models. On this dimension, Hendry's efforts should be commended, as he is trying to find a middle ground between atheoretical macroeconomics and an empirically irrelevant theory.

Despite his efforts to merge theory and data, Hendry errs by equating 'theory' with simple static models. Hendry does not motivate his empirical applications by appeals to contemporary theory; his motivations tend to rely on simple static equations which are used only to motivate the steady-state implications. Yet modern macroeconomics is rich with models and implications for dynamic empirical modelling. Hendry's excessive scepticism of this theory is unwarranted and possibly harmful, as it tends to turn away potential readers from his papers.

The truth is that it is exceedingly difficult to develop empirically successful models which are firmly based in contemporary theory. In my view, there is no reason for all researchers to try. The principle of comparative advantage suggests that it is perfectly acceptable for some researchers to specialise in data description (reduced form analysis), others in pure theory (stylised models), and others in bridging the gap. While Hendry is perfectly justified to build models in the style suggested by the Hendry methodology, it does not seem prudent for the remainder of the profession necessarily to follow his recommendation. We as economists need a wide range of information. First, we need to know the 'stylised facts' about the world. These can best be discovered through non-structural data analysis and reduced-form model estimation. Some researchers will find this work to be their comparative advantage. Secondly, we also need to know how specific economic models behave. This requires analytical and numerical evaluation of often-complex systems. Such tasks may best be left to specialists in economic theory, and will undoubtedly include exercises in the 'calibration' spirit currently fashionable among real-business-cycle theorists. Thirdly, we need to know which economic models are consistent with the data. If they are not consistent, we want to know in which dimensions they are consistent, and in which dimensions they are inconsistent. This requires an intermediate – econometric – treatment, and may be the most challenging task of the three. While model evaluation is intermediate between reduced form analysis and pure theory, it is not clear to me that there is a unique balance between the two which is ideal in all situations. Different researchers will lean towards different ends of the spectrum, some working with tightly specified theoretically rigorous models which are not completely data coherent, and others working with empirically tight but theoretically loose models. I suspect that we can learn from many different approaches, not just one 'correct' path.

I would also like to point out that a major challenge that macroeconometrics faces in the coming decade is to evaluate the current crop of stochastic general equilibrium models empirically. Hendry's ideas may partially help us meet this challenge. Many modern growth models, when linearised, imply that the state variables are individually integrated and mutually cointegrated. The economic model may imply a restricted set of dynamics: these restrictions can be tested against a reduced-form vector autoregression to see if the model is 'Hendry-admissible'. If the model fails the specification tests, then the question is how to modify the model. One of Hendry's points is that it is desirable to modify only the inessential features of the model, retaining the central (say, steady-state) features. The difficulty is how to make such modifications successfully without leaving the framework of the economic model. In this dimension, I am afraid that Hendry's methodology will not provide much insight.

II.6. *Encompassing*

The criteria of encompassing seems largely unique to Hendry and his co-authors. While the idea was present in some earlier papers, it seems to have been first carefully explored in 'On the Formulation of Empirical Models in Dynamic Econometrics', with Jean-François-Richard. The idea is that the true DGP should be able to explain any empirical result, so a chosen model should similarly be able to explain the results obtained from any competing specification. When two models are nested, encompassing reduces to a likelihood ratio test. When two models are non-nested, then encompassing is a particular non-nested test. In fact, it is a generalisation of the original Cox likelihood ratio test. In linear Gaussian models, the Cox statistic can be viewed as a 'variance-encompassing' statistic, since it asks whether the null model can explain the residual variance in the alternative model.

As pointed out in 'The Econometric Analysis of Economic Time Series' (also with Jean-François-Richard), in linear models the encompassing test is equivalent to an 'embedding' test. Specifically, if the null is $\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta}+\mathbf{e}$ and the alternative is $\mathbf{Y} = \mathbf{X}\boldsymbol{\delta}+\mathbf{u}$, the encompassing test attempts to explain the point estimate $\hat{\boldsymbol{\gamma}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ using $\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{Y})$ and $\hat{\boldsymbol{\Pi}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z})$. The embedding test is the F statistic for $\hat{\boldsymbol{\delta}}$ from the regression model

$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\delta} + \mathbf{e}$. Hendry and Richard show that in fact these two tests are equivalent. This is strong evidence that the encompassing test is a good choice for testing non-nested hypotheses.

While the above argument is sufficient to justify the encompassing test for non-nested models, I find Hendry's preferred justification more confusing. He claims that the Cox test 'emphasizes a "negative" rather than a "positive" approach since all contending models could end being rejected'. I find this confusing since the encompassing test is simply a generalisation of the Cox test, and not fundamentally different. Both are traditional 'specification tests': they set up the null hypothesis that one's chosen model is correct, and then test that model (and hypothesis) by taking the alternative to be a specific non-nested model. One finds that the chosen null model is incorrect (misspecified) if the test statistic rejects the null hypothesis, and that the chosen model is tentatively acceptable if the test fails to reject the null.

I think where Hendry believes that the 'tests' differ is not in the mathematics, but in the metaphysics surrounding the test. Hendry argues that encompassing is a principle which models should satisfy, and that the encompassing statistic is a test for encompassing. The Cox test is not set up as a general principle; instead it was proposed as a specific test to compare two non-nested alternatives. Quite frankly, this is not a very meaningful distinction and might be more misleading than illuminating. One distinction might be that Cox suggested that we should run the test in both directions (taking both models as the null) while Hendry recommends taking only our selected model as the null, which is a distinction concerning the implementation of the tests, not of the tests themselves.

The Hendry methodology suggests that any proposed model needs to be able to encompass all rival models. Since 'encompass a rival model' means 'pass an encompassing test with the rival model as the alternative', this means that a proposed model must be able to pass a large number of specification tests, taking every potential model very seriously. This seems to be a very strict requirement. Suppose that the proposed model is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, and there are m possible rival models: $\mathbf{Y} = \mathbf{Z}_j \boldsymbol{\gamma}_j + \mathbf{e}_j, j = 1, \ldots, m$. The encompassing test against a particular rival model consists of testing the significance of $\boldsymbol{\delta}_j$ in the regression $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{\delta}_j + \mathbf{e}^*$. Let the p-value for this statistic be $\eta_T(j)$. Hendry's methodology suggests that we use only models for which $\eta_T(j) > 0.05$ for all j. This methodology seems prone to error. If the tests $\eta_T(j)$ are not perfectly correlated and m is sufficiently large, then the probability that at least one of the $\eta_T(j)$ will be smaller than 0.05 can be made arbitrarily large, even when the proposed model is the correct DGP. A modified proposal to control size would be to use an appropriate critical value c which satisfies

$$P[\min_{1 \leqslant j \leqslant m} \eta_T(j) \leqslant c] = 0.05.$$

Unfortunately, finding the correct value of c is quite difficult. (All we know is that it is not 0.05 unless $m = 1$!)

An alternative way to control overall size is to use a conventional approach to the testing of joint hypotheses. Note that the encompassing test $\eta_T(j)$ is a conditional moment test of the hypothesis

$$\mathrm{E}(\mathbf{Z}_j'\,\mathbf{e}) = 0. \tag{8}$$

The joint test suggested by the Hendry methodology takes the alternative to be (8) for all $j \leqslant m$. The standard conditional moment test for this alternative is just the F-test for significance of $\boldsymbol{\delta}$ in the regression $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{\delta}+\mathbf{e}^*$ where $\mathbf{Z}$ consists of the non-redundant columns of $[\mathbf{Z}_1, \ldots, \mathbf{Z}_m]$. This can be interpreted as either the encompassing test against the rival model $\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma}+\mathbf{e}$ or as an omnibus data coherency test as described in Section II.2. Thus if we take the conventional approach to unify the distribution theory for the proposed set of encompassing tests, we find that the methodology of encompassing leads us right back to the practice of testing against general alternatives, and is therefore not a distinct testing issue from that of data coherency.

In sum, I think it is important to distinguish between two operating concepts in Hendry's vision of encompassing: (i) as a non-nested hypothesis test; and (ii) as a model-selection method. I agree with Hendry's analysis that encompassing tests are useful for comparing non-nested models, and that non-nested hypothesis testing is too frequently ignored in econometric practice. But the methodological idea that all good models should survive a battery of encompassing tests against a range of conceivable rival models does not seem to make good statistical sense. Implementation of this methodology either results in a test with unknown size, or with a simple F-test for a general specification.

III. CONCLUSION

Central to Hendry's methodology is the GtoS model selection approach. At first glance, it appears unassailable. Yet as we have explored the details of the methodology, we have found several inconsistencies. Hendry appears to recommend that researchers start with models which are unrestricted in the dynamics, but are limited to the set of linear conditional models with Gaussian errors and constant coefficients. Deviations from this class are not considered *a priori*, but can be tested using specification tests. This practice is GtoS as regards linear dynamics, but is StoG in every other dimension. Thus Hendry himself fails to follow his own methodology.

Indeed, it is easy to see that it is impossible to implement the GtoS approach fully. This would require an enormously complex exercise, with a complete model of the joint distribution of all variables, allowing for non-linearities, heteroscedasticity, coefficient drift and non-Gaussian errors. It is clear that this would be too costly in terms of parameterisation. The only practical solution is to mix-and-match the GtoS and StoG methods. Hendry advocates one such mixture, but others are surely conceivable. If we accept this premise, then other researchers should be free to make alternative choices for their initial 'general' formulation, which may not nest Hendry's 'general' formulation as a special case. This leads to possible diversity in econometric research, rather than the homogeneity suggested by Hendry's vision of a progressive research tradition.

In this review I have tried to take a stern look at the methodology of econometrics developed by David Hendry and illustrated in the collection of papers in *Econometrics: Alchemy or Science?* While arguing that the methodology is incomplete and partly self-contradictory, I hope that my message has not been too negative. While imperfect, Hendry's methodology has made major improvements to econometric understanding and practice. I know that his writings have made a major impact on my own way of thinking about econometric research, and I believe this is true of other econometricians who have read his work. Yet I reserve the right to be a sceptic; and in this manner have approached his book. Others will no doubt come to different conclusions. I hope that the publication of *Econometrics: Alchemy or Science?* will provide fuel for a energetic debate about econometric method and practice.

References

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control.* Holden-Day.

Granger, C. W. G. (1981). 'Some properties of time series data and their use in econometric model specification'. *Journal of Econometrics*, vol. 16, pp. 121–30.

Hausman, J. A. (1978). 'Specification tests in econometrics.' *Econometrica*, vol. 46; pp. 1251–71.

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 *Econometric Analysis for National Economic Planning*, (ed. G. Mills, P. E. Hart and J. K. Whitaker), pp. 25–63. Butterworths.

Wu, D. M. (1973). 'Alternative tests of independence between stochastic regressors and disturbances.' *Econometrica*, vol. 41, pp. 733–50.

APPENDIX: *Papers Republished in* Econometrics: Alchemy or Science?

Part I: Roots and Route Maps

1. 'Econometrics – alchemy or science.' *Economica* (1980).
2. 'Stochastic specification in an aggregate demand model of the United Kingdom.' *Econometrica* (1974).
3. 'Testing dynamic specification in small simultaneous systems: an application to a model of building society behavior in the United Kingdom', with Gordon J. Anderson. *Frontiers of Quantitative Economics*, vol. III A (1977).
4. 'Dynamic specification', with Adrian R. Pagan and J. Denis Sargan. *Handbook of Econometrics*, vol. II (1984).

Part II: The Development of Empirical Modelling Strategies

5. 'On the time-series approach to econometric model building.' *New Methods in Business Cycle Research* (1977).
6. 'Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England', with Grayham E. Mizon, ECONOMIC JOURNAL (1978).
7. 'An empirical application and Monte Carlo analysis of tests of dynamic specification', with Grayham E. Mizon. *Review of Economic Studies* (1980).
8. 'Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom', with J. E. H. Davidson, F. Srba and S. Yeo. ECONOMIC JOURNAL (1978).
9. 'Liquidity and inflation effects on consumers' expenditure', with Thomas von Ungern-Sternberg, *Essays in the Theory and Measurement of Consumers' Expenditure* (1981).

10. 'Interpreting econometric evidence: the behaviour of consumers' expenditure in the United Kingdom', with James E. H. Davidson. *European Economic Review* (1981).

11. 'Predictive failure and econometric modelling in macroeconomics: the transactions demand for money.' *Economic Modelling* (1979).

12. 'Monetary economic myth and econometric reality.' *Oxford Review of Economic Policy* (1985).

Part III: Formalisation

13. 'The structure of simultaneous equations estimators.' *Journal of Econometrics* (1976).

14. 'AUTOREG: a computer program library for dynamic econometric models with autoregressive errors', with Frank Srba. *Journal of Econometrics* (1980).

15. 'Exogeneity', with Robert F. Engle and Jean-François-Richard. *Econometrica* (1983).

16. 'On the formulation of empirical models in dynamic econometrics', with Jean-François-Richard. *Journal of Econometrics* (1982).

17. 'The econometric analysis for economic time series', with Jean-François-Richard. *International Statistical Review* (1983).

Part IV: Retrospect and Prospect

18. 'Econometric modelling: the "consumption function" in retrospect.' *Scottish Journal of Political Economy* (1983).

19. 'Postscript: the econometrics of PC-GIVE.'