The econometrics of the Holy Grail

A critique

Hugo A. Keuzenkamp

Department of Economics
Tilburg University

and

Centre for the Philosophy
of the Natural and Social Sciences
London School of Economics

Revised: 16 February 1995

Abstract.
This paper discusses the methodological views of David Hendry. A critique is given on the ‘reification’ of the so-called Data Generating Process. The merits of general to specific modelling are analyzed. Hendry’s neo-Popperian philosophy is examined. It is argued that this philosophy is not able to transform econometrics from ‘alchemy’ to ‘science’.

Keywords:
Econometric methodology; data generation process; testing; general to specific.

E-mail:
h.a.keuzenkamp@kub.nl

‘What seek ye in this country?’
‘Sir,’ said Launcelot, ‘I go to seek the adventures of the Sangrail.’
‘Well,’ said he, ‘seek it ye may well, but though it were here ye shall have no power to see it no more than a blind man should see a bright sword, and that is due to your sin, and else ye were more abler than any man living.’
And then Sir Launcelot began to weep.

Sir Thomas Malory, *Le Morte d’Arthur*, Book XV, Chapter 2.

1. Introduction[1,2]

During the last two decades, David Hendry has been a leading participant in the debate on econometric methodology. His influence in applied econometrics, especially in the UK and Europe, is large. This is due to original insights, a penetrating style of writing and a zeal that has made many scholars enthusiastic about econometrics. Hendry’s ‘econometrics in action’ seminars, presented during the mid-eighties, were not only thought-provoking but also highly entertaining. They illustrated and advanced the breakthrough in computer-aided econometrics education. Different versions of PC-GIVE reflect this development.

Hendry belongs to the small group of econometricians with very explicit methodological views. Those views are scattered throughout many of his publications, but have been made readily accessible by means of a sample of 18 of Hendry’s contributions to econometrics, collected in the volume ‘*Econometrics: Alchemy or Science*’ (Hendry,

[1] This paper was written during a visit to the Centre for the Philosophy of the Natural and Social Sciences at London School of Economics. I would like to thank Mark Blaug, Michael McAleer, Les Oxley, Stephen Pollock, Ron Smith, Mark Steel and participants at seminars at the London School of Economics and Exeter University for helpful comments on previous drafts. Financial support from the Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged.

[2] This paper is a review essay of David F. Hendry (1993), *Econometrics: Alchemy or Science? Essays in Econometric Methodology*, Blackwell, Oxford; ISBN 1-55786-264-8; £50.00. References without author and year apply to this volume of papers.

1993).[3] The papers, published between 1971 and 1985, are accompanied by new preambles and a postscript. The first essay, from which the title of the book originates, is Hendry's inaugural lecture at the London School of Economics. It serves, in a sense, as an introduction to the whole volume. Like many other inaugural lectures, it is like a mission statement. The preambles of the other papers explain, with hindsight, the context and innovations, but also the shortcomings, of the respective papers. The book is divided in four parts: (1) Roots and route maps; (2) empirical modelling strategies (including 'DHSY'); (3) formalization (including the classic Exogeneity paper, co-authored with Engle and Richard); and (4) retrospect and prospect (including 'the econometrics of PC-GIVE').

As the papers themselves have appeared in respected journals and books, the purpose of this essay is not to assess Hendry's ability as an academic econometrician. Indeed, there is no doubting Hendry's skills and the technical quality of his methods. But Hendry aims higher, at methodology. Here, he is on shaky ground. Not surprisingly, Hendry's methodological views have been controversial, in particular outside the UK. The value added of *Alchemy or Science* is primarily its effort to provide an anthology of his methodological thinking. The preambles to the essays particularly serve this purpose. Hence, I will focus on methodology.[4]

Before doing so, consider briefly the paper with the mission statement, *Alchemy or Science*, itself. The example on which it builds is not especially different from that in a paper by George Udny Yule (1926), in which the pitfalls of time series analysis are illustrated by means of correlating marriages and mortality in England and Wales from 1866 to 1911. Likewise, Hendry uses cumulative rainfall in order to explain inflation in the UK. We all are (or should be) familiar with the emerging point, which is the same

---

[3] Hendry's bibliography, presented at the end of the book, contains 86 scientific publications.

[4] Some useful reviews of Hendry's econometric methodology have been published in the past (Gilbert, 1986; Pagan, 1987, 1994). Another source of interest is the *Econometric Theory* dialogue (Hendry, Leamer and Poirier, 1990). Somewhat surprisingly, Hendry (1987) has been omitted in the collection of essays, but the arguments given in that paper recur in the papers that are included.

as that of Yule and of Granger and Newbold (1974). The fact that this point had to be reiterated more than fifty years after Yule's investigations says much about the state of modern econometrics!

Whereas Yule was preoccupied with the question of whether differencing or detrending would be the best way forward (a question still bothering time series analysts today), Hendry's remedy for nonsense correlations ('alchemy') are the three golden rules, 'test, test and test'. This trinity was not really available in the time of Yule, and has been problematic in statistical research since its development.[5] Those three cheers for statistical testing are Hendry's way of transforming econometrics from alchemy to science. Econometrics, Hendry writes in the general introduction, 'is potentially scientific precisely because alchemy is creatable, detectable and refutable' (1993, p. 1).[6] I will review Hendry's methodological views, starting with the general and then considering some specific issues--a fruitful routine, it is said.

Section 2 contains a brief discussion of alchemy. Section 3 considers the Holy Grail, better known as the Data Generation Process. Section 4 provides a discussion on the merits of a general-to-specific modelling strategy. Section 5 deals with Hendry's references to the philosophy of science, in particular Popper and Lakatos. Section 6 discusses the importance of model design and the consequences for inference. Section 7 concludes the paper.

---

[5] Keynes (1939) refers to 'the mine Yule sprang' under time series statistics (see Hendry, p. 20). The question is whether there has been significant progress beyond the stage of Yule (1926). An enormous literature on the AR(1) model, yielding an extravagance of ever more complicated test statistics (a cursory look at issues of *Econometrica* around 1990 shows the point), has not helped to raise time series econometrics to a more scientific (credible?) level. The recent fad of calibrating dynamic models rather than estimating them is a '*j'accuse*' of some applied economists against the incredibility of modern macroeconometrics. Of course, calibration is more likely a symptom of the disease than the cure (ingredients in calibrated models tend to be microeconometric findings).

[6] Ironically, the alchemist philosophy (influenced by Aristotelian thought) in a nutshell is to go from *potentials* to *actuals*.

2. Alchemy

The early econometricians, in particular Jan Tinbergen, were regularly accused by John Maynard Keynes of practising alchemy (see *e.g.* Hendry's quotation, p. 14) or black magic. The alchemists tried to transmute lead to gold, but not quite successfully, as we know. As Chauser wrote about the 'Oxford Cleric' - a student in alchemy - in the Prologue to his Canterbury Tales, 'But al be that he was a philosophre, Yet hadde he but litel gold in cofre'.[7] Alchemy has a bad connotation. However, the fact is that the alchemists gathered impressive empirical knowledge about basic chemical facts. This knowledge enabled the growth of chemistry. What was it that made chemistry, unlike alchemy, a science? This question is hard to answer (and it may well be that the question itself is flawed). It is certainly not a lack of 'the experimental method', as alchemists experimented like mad.[8] In popular writings, alchemy is considered to be measurement without theory (or with a wrong theory - but are not all theories wrong?). Perhaps a more convincing view is that alchemy is science with wrong (or, more precisely, unattainable) aims.[9] Perhaps the unintended success in activities other than making gold, and finally discarding the unattainable aim itself, yielded chemistry the respect that (time series) econometrics has yet to gain.

One of the salient features of alchemy is the use of a mystical and symbolic language. Econometrics is not much different. We have become familiar with the notion of an hypothetical infinite population used in an investigation of money-inflation data for the UK from 1952-I to 1992-IV. *Alchemy or Science* adds some modernisms. The reader

[7] The (1977) Penguin Books edition of the Canterbury Tales, translated in modern English by Nevill Coghill, reads (p. 27):
'Though a philosopher, as I have told,
He had not found the stone for making gold.'

[8] It may be, however, that a characteristic of experiments in alchemy is the lack of replicability, and replicability is an important feature of science. I owe this suggestion to Mark Blaug.

[9] The fact that the aims are 'wrong' or unattainable is, of course, reasoning with hindsight. It should not be understood as a critique of the alchemists.

is treated to ‘congruence’, the Haavelmo distribution, parsimonious encompassing and other mystifying idiosyncrasies.[10] Where the alchemists spoke of panacea or arcanum (the medicine against all ills), Hendry prescribes tests. But those fine words do not suffice to transfer econometrics from alchemy to science. Repeatedly, Hendry advocates testing on ground that it is criticism that should be the core of scientific method in econometrics. So let us practise criticism!

3. The Data Generation Process

The ‘Data Generation Process’ (DGP) is the Holy Grail of Hendry’s econometrics.[11] It is a highly dimensional probability distribution for a huge vector of variables, ***w***,

$$DGP \equiv f(\boldsymbol{W}^1_T | \boldsymbol{W}_0, \omega^1_T) = \prod_{t=1}^{T} f(\boldsymbol{w}_t | \boldsymbol{W}^1_{t-1}, \omega_t) \ , \qquad \textbf{(1)}$$

conditional on initial conditions ($\boldsymbol{W}_0$), parameters $\omega_t \in \Omega_t$, continuity (needed for a density representation) and time homogeneity ($f_t(\cdot) = f(\cdot)$). It is a recurring theme in Hendry’s writings as well as in the writings of his adepts (to use alchemist terminology). A DGP is a well defined notion in Monte Carlo studies, where the investigator is able to generate the data and knows exactly the characteristics of the generating process. However, this notion is transferred to applied econometrics, which is supposed to deal with the study of the properties of the DGP (p. 13). The theory of reduction intends to bring this incomprehensive distribution down to a parsimonious model, without loss of relevant information. Marginalizing and conditioning are two key notions of this theory of reduction.

In applied econometrics, Hendry argues, ‘the *data* are given, and so the distributions

[10] Just an example: ‘Congruence is a relation between the model and the Haavelmo distribution’ (p. 271).

[11] Both words, ‘generation’ and ‘generating’, are used (*e.g.* p. 393 and p. 394, respectively).

of the dependent variables are already fixed by what the data generating process created them to be--I knew that from Monte Carlo' (p. 73). Apart from the data, there are models (sets of hypotheses). Given the data, the goal is to make inferences concerning the hypotheses. This is reasoning from the sample to the population (in the terminology of frequentist probability theory), or inductive inference. In the theory of reduction, the idea is that the DGP (or population) can be used as a starting point as well as a goal of inference. Econometric models are regarded as reductions of the DGP (p. 247).

Although the DGP is sometimes presented as 'hypothetical' (*e.g.* p. 364--perhaps this is the influence of co-author Jean François Richard), there is a tendency in Hendry's writings to view 'the' DGP not as an hypothesis (of many possible hypotheses), but as fact or reality. Sometimes the DGP is presented as 'the relevant data generation process' (p. 74). In other instances, the DGP becomes the 'actual mechanism that generates the data' (Hendry and Ericsson, 1991, p. 18), or simply 'the actual DGP' (Spanos, 1986). Taking the DGP as a starting point or inference is, from a methodological point of view, highly objectionable, except perhaps in Monte Carlo studies. The related issue of general-to-specific modelling (and the theory of reduction) will be discussed in Section 6 below. Here, I will focus on the status of the DGP itself.

The DGP is reality and a model of reality at the same time.[12] Philosophers call this 'reification'. Once this position is taken, weird consequences follow. Consider the 'information taxonomy' that 'follows directly from the theory of reduction given the sequence of steps needed to derive any model from the data generation process' (p. 271; see also p. 358). Or, econometric models are 'derived and *derivable* from the DGP' (Hendry and Ericsson, 1990, p. 20; emphasis added). Is this science, metaphysics, or yet another version of the economist in the desert with an unopened can, who proposes 'let us assume we have a can opener'? Consider another example of reification: the proposition that 'it is a common empirical finding that DGP's are

[12] This conflation is explicitly recognized on p. 77; see also p. 86: 'a correct model should be capable of predicting the residual variance of an incorrect model and any failure to do this demonstrates that the first model is not the DGP'.

subject to interventions affecting some of their parameters' (p. 373). Rather, those empirical findings relate to statistical models of the data, not to the DGP or its parameters. Bruno de Finetti is known for his phrase 'probability does not exist', meaning to say that it is an invention of our mind. The same applies to the DGP or its parameters. The DGP does not exist.

Models are useful approximations to reality. Different purposes require different models (and often different data, concerning levels of aggregation, numbers of observations). The idea that there is one DGP waiting to be discovered is a form of medieval romanticism that suits the legend of the Holy Grail, but not modern science.[13] Neither is it, nor should it be, the task of the econometrician 'to model the main features of *the* data generation process' (p. 445; emphasis added). It may even be the purpose of the econometrician (or policymaker) to *change* the economic process in response to econometric inference or predictions. This may involve self-fulfilling (or self-denying) prophecies, which are hard (if not impossible) to reconcile with the notion of a DGP.

4. General to specific

An important characteristic of Hendry's methodology is his general-to-specific (GTS) approach to modelling. It was initially inspired by Denis Sargan's Common Factor (COMFAC) test, where a general model is needed as a starting point for testing a lagged set of variables corresponding to an autoregressive error process. Another source of inspiration was the reductionist approach developed at CORE (in particular, by Jean-Pierre Florens, Michel Mouchart and Jean-François Richard). The theory of reduction, of which Ronald A. Fisher may be regarded as a pioneer, deals with the conditions for valid marginalizing and conditioning (relying on sufficient statistics).[14]

---

[13] Alternatively, the belief in the DGP can be regarded as extreme scientific realism, whereas my critique stems from an instrumentalistic attitude.

[14] However, note that Fisher does not claim to reduce the DGP, but rather to reduce a large set of data without losing relevant information.

The basic idea of GTS is that there is a starting point for all inference, the data generation process. All models are reductions of the DGP, but not all of them are valid reductions. In the following, a brief outline of reduction will be sketched. Subsequently, a methodological critique is given. Reconsider the DGP, (1). The variable vector $\boldsymbol{w}$ contains only few variables that are likely to be relevant. These are labelled $\boldsymbol{y}^*$, the remainder is $\boldsymbol{w}^*$ (nuisance variables). Using the definition of conditional probability,[15] and assuming parameter constancy for simplicity, one can partition the DGP:

$$f(\boldsymbol{w}_t^*,\boldsymbol{y}_t^* \mid \boldsymbol{W}_{t-1}^*,\boldsymbol{Y}_{t-1}^*,\omega_1) = f_1(\boldsymbol{y}_t^* \mid \boldsymbol{W}_{t-1}^*,\boldsymbol{Y}_{t-1}^*,\omega_1) f_2(\boldsymbol{w}_t^* \mid \boldsymbol{W}_{t-1}^*,\boldsymbol{y}_t^*,\boldsymbol{Y}_{t-1}^*,\omega_2) \qquad \textbf{(2)}$$

where $\boldsymbol{W}_{t-1}^*=(\boldsymbol{W}_{t-1}^1)^*$. The interest is on the marginal density, f_1. If ω_1 and ω_2 are variation free ($(\omega_1,\omega_2) \in \Omega_1 \times \Omega_2$; i.e. f_2 does not provide information about Ω_1) then, a ‘sequential cut’ may be operated: consider only f_1. Assuming furthermore that $\boldsymbol{y}_t^*$ and $\boldsymbol{W}_{t-1}^*$ are independent conditionally on $\boldsymbol{Y}_{t-1}^*$ and ω_1, the nuisance variables can be dropped from f_1, which yields:

$$f(\boldsymbol{y}_t^* \mid \boldsymbol{Y}_{t-1}^*,\omega_1). \qquad \textbf{(3)}$$

This can be represented using a specific probabilistic and functional form (e.g. a Vector Autoregression, VAR). Note that this marginalizing stage of the reduction process proceeds entirely implicitly. It is impossible to consider empirically all conceivable variables of interest to start with. Note also that many hidden assumptions are made (e.g. going from distributions to densities, assuming linearity, parametric distributions, constant parameters). Some of them may be testable but most are taken for granted.[16]

The second stage, conditioning, aims at a further reduction by introducing exogeneity into the model. For this purpose, decompose $\boldsymbol{y}^*$ into $\boldsymbol{y}$ (endogenous) and $\boldsymbol{x}$ (exogenous) variables. Rewrite (3):

[15] $P(w|y) = P(w,y)/P(y)$.

[16] Hendry does not have a strong interest in non-parametric or non-linear models. This is hard to justify in a GTS approach. It rules out economic models with multiple equilibria.

$$f(\boldsymbol{y}^*_t \mid \boldsymbol{Y}^*_{t-1},\omega_1) = f(\boldsymbol{y}_t \mid \boldsymbol{X}_{t-1},\boldsymbol{x}_t,\boldsymbol{Y}_{t-1},\theta_1)f(\boldsymbol{x}_t \mid \boldsymbol{X}_{t-1},\boldsymbol{Y}_{t-1},\theta_2) \qquad \textbf{(4)}$$

where θ_1. are the parameters of interest. If $\mathbf{x}_t$ is 'weakly exogenous' (θ_1 and θ_2 are variation free), then 'valid' and efficient inference on θ_1 or functions thereof on the basis of

$$f(\boldsymbol{y}_t \mid \boldsymbol{X}_{t-1},\boldsymbol{x}_t,\boldsymbol{Y}_{t-1},\theta_1) \qquad \textbf{(5)}$$

is possible (where 'valid' means: no loss of information). This conditional model can be represented in specific form, usually the Autoregressive Distributed Lag model is chosen for this purpose (e.g. pp. 87-8):

$$A(L)\boldsymbol{y}_t = B(L)\boldsymbol{x}_t + \boldsymbol{u}_t \qquad \textbf{(6)}$$

(where *L* denotes the lag operator). Note again the comments on implicit and explicit assumptions mentioned with respect to the VAR-representation of (3), which equally apply to this case. The final stage of modelling is to consider whether further restrictions on (6) can be imposed without loss of information. Examples are common dynamic factors, or other exclusion restrictions.

Destructive testing (by digesting the (mis-)specification test menus of PC-GIVE) aims at weeding out the invalid models. Remaining rival models are still 'comparable via the DGP' (p. 463). Encompassing tests are supposed to serve this purpose (see below, Section 6). Wald tests are the preferred tools for, among others, testing common factors. One reason is computational ease (e.g. p. 152), but the nature of this test matches the GTS modelling strategy quite well. The Wald test is also used in the context of encompassing tests (e.g. p. 413).[17]

[17] A problem with the Wald test is that it is not invariant to mathematically identical formulations of non-linear restrictions (Godfrey, 1989, p. 65). It is usually assumed that this does not affect linear restrictions, such as $\beta_1=\beta_2$. However, an identical formulation of this restriction is $\beta_1/\beta_2 = 1$, which may lead to the same problem of invariance. Hence, using a Wald-type general to specific test may not always be advisable. Hendry is open minded with respect to the choice of test statistics: usually they are chosen on pragmatic grounds.

Hendry's justification for GTS is based on the invalidity of the test statistics if inference commences with the simple model:

> 'every test is conditional on arbitrary assumptions which are to be tested *later*, and if these are rejected all earlier inferences are invalidated, whether 'reject' or 'not reject' decisions. Until the model adequately characterizes the data generation process, it seems rather pointless trying to test hypotheses of interest in economic theory. A further drawback is that the significance level of the unstructured sequence of tests actually being conducted is unknown' (p. 255).

Four objections to this argument can be made.

First, even the most general empirical model will be 'wrong' or mis-specified. Because of this possibility, Hendry (p. 257) advises testing the general model, but does not explain how to interpret the resulting test statistics. The validity of the general model is in the eyes of the beholder.[18] It is revealing that Hendry's interest has shifted from systems of equations to single equations (see p. 3), which is hard to justify from a GTS perspective.

Second, one does not need an 'adequate' statistical representation of the DGP (using sufficient statistics) in order to make inferences. In many cases, stylized facts, a few figures or rough summary statistics (averages of trials in experimental economics; eyeball statistics) are able to do the work (note, for example, that statistical modelling in small particle physics is a rarity; the same appies to game theory). A crucial ingredient to a theory of reduction should be a convincing argument for the optimal

[18] For example, the practical starting point of most of Hendry's modelling, the autoregressive distributive lag model (in rational form $y_t = \{b(L)/a(L)\}x_t + \{1/a(L)\}u_t$,) is less general than a transfer function model, $y_t = \{b(L)/a(L)\}x_t + \{m(L)/r(L)\}u_t$. This arbitrary specific starting point may lead to invalid inferences, for example, in investigating the permanent income hypothesis (see Pollock, 1992).

level of parsimony (simplicity). No such argument is given.[19]

Third, pretest bias keeps haunting econometrics, whether they use stepwise regression, iterative data mining routines, or GTS modelling. Significance levels of the COMFAC test may be known asymptotically in a GTS modelling strategy (as COMFAC uses a series of independent tests) (see e.g. pp. 152-3), but the sampling properties of most other test statistics *and even this one* are highly obscure outside the context of repeated sampling (Neyman-Pearson) or valid experimental design (Fisher) (see Section 6 below).

A more general fourth objection can be raised. Would econometrics be better of if the GTS methodology were to be adopted? This is doubtful. Empirical econometrics is an iterative procedure, and very few econometricians have the discipline or desire to obey the GTS straitjacket. Not surprisingly, econometric practice tends to be gradual approximation, often by means of variable addition. Unlike Hendry's warnings, this practice may yield interesting and useful empirical knowledge. Consider one of the best examples in recent econometric research: the analysis of the Permanent Income Hypothesis. This literature, spurred on by publications of Hall, Campbell and Mankiw, Deaton, Flavin and others, is an example of fruitful variable addition. First, there was the random walk model. Then we had excess sensitivity and excess smoothness. These problems had to be explained and were so, among others, by liquidity constraints. New variables were added to the specifications, and better (more adequate) empirical approximations were obtained (where those approximations could be directly related to advances in economic theory). This literature, not hampered by an overdose of testing and encompassing (see Section 6), may not carry Hendry's praise as true examples of science, but the economics profession tends to disagree.[20] The empirical literature on the permanent income hypothesis is viewed as a rare success story in macroeconometrics, one of the few cases where

[19] See also Keuzenkamp and McAleer (1995).

[20] Of course, one may object that economists are not the best judges on the fruits of econometrics. This would violate a basic rule of economics (and political science): the consumer (voter) is always right.

econometric analysis actually added to economic understanding of macroeconomic phenomena--a case where theory and empirical research iterated to a better understanding of economics.[21]

5. Falsificationism and the three cheers for testing

'The three golden rules of econometrics are test, test and test;[22] that all three rules are broken regularly in empirical applications is fortunately easily remedied' (pp. 27-8). Testing is the main virtue of a scientific econometrics, Hendry claims. This idea is related to a philosophy of science that was taught during Hendry's years at the London School of Economics (LSE) by Karl Popper and Imre Lakatos. Their philosophy of falsificationism (Popper) and the methodology of scientific research programmes (Lakatos) frequently recur in Hendry's writings.

According to Popper, one can never prove a theory right, but one may be able to falsify it by devising a 'crucial experiment'. Falsifiability separates science from metaphysics. Real scientists should formulate bold conjectures and try not to verify but to falsify. There are numerous problems with this view.[23] I will mention some that are relevant here. A crucial experiment is probably rare in physics (it may even be a scientific myth), and more so in economics. In fact, economics has hardly any experimentation at all (and where there is experimentation -as in game theory- econometrics is rarely invoked for purposes of inference). Sometimes one can find the

[21] Hendry's contribution to this literature is contained as Chapter 10 in the volume of papers. It is shown that an error correction model is able to 'encompass' Hall's random walk model of consumption in the context of UK data. There is no doubt that Hendry's specification provides a better approximation to the data. However, it was not Hall's point to obtain the best possible approximation. A similar remark applies to the discussion between Hendry and Ericsson (1991) and Friedman and Schwartz (1991).

[22] A footnote is added here: 'Notwithstanding the difficulties involved in calculating and controlling type I and II errors' (p. 28).

[23] See *e.g.* the excellent treatment in Hacking (1983).

claim that econometric modelling may serve as a substitute for the experimental method,[24] but I am rather sceptical about such a claim (see also Section 6). The three cheers for testing may suggest that Hendry takes falsificationism seriously, but if so, then only without Popper's bold conjectures.[25]

The idea that scientists cannot prove theories, but may be able to falsify them, attributed to Popper, was a common sense notion in the statistical literature since (at least) the turn of this century. One can find this, explicitly, in the writings of Karl Pearson, Ronald Aylmer Fisher, Harold Jeffreys, Jerzy Neyman and Egon Pearson,[26] Frederick Mills, Jan Tinbergen, Tjalling Koopmans, and probably many others.[27] They did not need philosophical consultation for gaining this insight, neither did they render it a philosophical dogma according to which falsification becomes the highest virtue of a scientist.[28] Econometricians are, in this respect, just like other scientists: they rarely aim at falsifying, but try to construct satisfactory empirical models.[29] Of course, 'satisfactory' needs to be defined, and this is difficult. Philosophers like Van Fraassen have dealt with this issue but econometricians may find more wisdom in the *Oxford Bulletin of Economics and Statistics*, to mention just one research outlet that owes much to Hendry's stimulus, rather than tracts in the philosophy of science literature in order to clarify the operational meaning of 'satisfactory'.

---

[24] For example, expressed in Goldberger (1964). See also Morgan (1990, pp. 9-10 and p. 259).

[25] Darnell and Evans (1990), who (unlike me) make an effort to uphold the Popperian approach to econometrics (meanwhile having a quasi-Bayesian interpretation of probability) similarly complain that Hendry does not deliver the Popperian good.

[26] Although with a twist in their case, as they are interested in behaviour rather than inference.

[27] Citations available on request.

[28] According to Mark Blaug, an important difference between Popper and those statisticians is Popper's banning of immunizing strategems. Whether an absolute ban would benefit science is doubtful. See Keuzenkamp and McAleer (1995) for a discussion of 'ad hocness' and inference, in particular the references to Jeffreys.

[29] See Keuzenkamp and Barten (1995).

So far for Popper, Hendry's favourite philosopher. His second idol is Lakatos. Lakatos (1970) noted that, in practice, few scientific theories are given up because of a particular falsification. Moreover, scientists may try to obtain support for their theories. He accepted a 'whiff of inductivism'. In order to appraise scientific theories, Lakatos invented the notion of 'progressive' and 'degenerative' scientific research programmes, which can be on theoretical, heuristic and empirical grounds. A research programme is theoretically progressive if it predicts some novel, unexpected fact. It is empirically progressive if some of those predictions are confirmed. Finally, a research programme is heuristically progressive if it is able to avoid or even to reduce the dependency on auxiliary hypotheses that do not follow from the 'positive heuristic', the general guidelines of the research programme.

It is very difficult to encounter a single novel fact in Lakatos' sense in the econometric literature, and the papers of Hendry are no exception. Moreover, it is totally unclear what kind of heuristic in an economic research programme drives Hendry's empirical investigations, apart from a few rather elementary economic relations concerning consumption or money demand. Hendry does not predict novel facts. What he is able to do is provide novel interpretations of given facts, which is something entirely different. Lakatos (1970, p. 176) once criticized the

> 'patched-up, unimaginative series of pedestrian "empirical" adjustments which are so frequent, for instance, in modern social psychology. Such adjustments may, with the help of so-called "statistical techniques", make some "novel" predictions and may even conjure up some irrelevant grains of truth in them. But this theorizing has no unifying idea, no heuristic power, no continuity.'

In order to invoke Lakatos, one needs to provide the necessary ingredients of a research programme, in particular the driving 'heuristic'. This is very difficult to do, and the notion that there are some long run relationships in economics which hardly go beyond what can be found in introductory macroeconomics textbooks, together with the view that people adjust according to an error-correcting scheme, seems to be an insufficient heuristic for a well-defined research programme.[30]

However, Hendry also applies the notion of progressiveness to his methodology itself, rather than to the economic implementations:

> 'The methodology itself has also progressed and gradually has been able both to explain more of what we observe to occur in empirical econometrics and to predict the general consequences of certain research strategies' (p. 417).

Here a 'meta research programme' is at stake: not neo-classical economics, for example, but the *method* of inference is regarded as a research programme, with a heuristic, and all the other Lakatosian notions (hard core, protective belt, novel facts). Hendry's positive heuristic is then to 'test, test and test'. The hard core might be the notion of a DGP to be modelled using 'dynamic econometrics'. Probably the most promising candidate for a 'novel fact' is the insight that, if autocorrelation in the residuals is removed by means of a Cochrane-Orcutt AR(1) transformation, this may impose an invalid common factor: the transformed residuals may behave as a white noise process but they are not innovations. This is an important insight. Whether the stated heuristic is sufficient for defining a research programme, and whether this novel fact satisfies the desire for progression, is a matter on which disagreement may (and does) exist.

Apart from the dubious relation to the philosophy of science literature, there is a very different problem with Hendry's golden rules, which is that the statistical meaning of his tests is unclear. Standard errors and other 'statistics' presented in time series econometrics do not have the same interpretation for statistical inference as they have in situations of experimental data. In econometrics, standard errors are just

[30] Note moreover that the recent shift to co-integration, which can also be found in Hendry's writings, weakens the error correcting view of the world. The links between co-integration and error correction are less strong than at first sight may seem (see Pagan, 1994).

measurements of the precision of estimates, given the particular measurement system (econometric model) at hand. They are not test statistics. At times, Hendry seems to agree with this interpretation. For example,

> ‘Genuine *testing* can therefore only occur after the design process is complete and new evidence has accrued against which to test. Because new data has been collected since chapter 8 was published, the validity of the model could be investigated on the basis of Neyman-Pearson (1933) "quality control" tests’ (p. 420).

This statement also suggests that encompassing tests (discussed in the next section) are not ‘genuine’ tests (indeed, sometimes encompassing is presented as a form of mis-specification testing).

The reference to Neyman-Pearson in the quotation is only partly valid. Neyman-Pearson emphasizes the context of decision making in repetitive situations (hence, the theory is based on repeated sampling). In Hendry’s case, a number of additional observations has been collected, but Hendry’s aim is not decision making but inference.[31]

Occasionally, instead of the words ‘statistical tests’, the words ‘diagnostic checks’ are used. But not in the trinity, the three golden rules. Would they have the same aural appeal if they were ‘check, check and check’, or ‘measure, measure and measure’? It is doubtful, although as rules they seem more appropriate. Econometricians measure rather than test, and the obsession with testing is rather deplorable. Indeed, a final problem with excessive testing is that, once the number of test statistics exceeds the number of observations used to calculate them, one may wonder how well the aim of reduction has been served, and what the meaning of the tests really is. Hendry and Ericsson (1991) provide an example. Given 93 annual observations, 46 test statistics

---

[31] See Keuzenkamp and Magnus (1995) for a discussion of Neyman-Pearson methods.

are reported. If standard errors of estimates are included, this number grows to 108. It grows further once the implicit (eyeball) test statistics conveyed by the figures are considered.

Two final remarks on statistical tests: first, there is always the problem of choosing the appropriate significance level (why does 5% deserves special attention?)--a problem not specific to Hendry's methodology; second, a statistical rejection may not be an economically meaningful rejection. If rational behaviour (in whatever sense) is statistically rejected (even if the context is literally one of repeated sampling) at some significance level, this does not mean that the objects of inference could increase their utility by changing behaviour. Money-metric test statistics would be more informative in econometric inference than statistical tests, but such tests are not discussed by Hendry or most econometricians (to the best of my knowledge, Hal Varian is a sole exception). This issue is of importance in many (proclaimed) tests of perfect markets currently presented in the finance literature.

6. Model design and inference

Hendry argues that all models are derived from the DGP and, therefore, also the properties of the models are derived entities. One does not need to invoke a DGP to arrive at this important insight of non-experimental data analysis (e.g., replace 'DGP' by 'data') - after all, what is this thing called 'DGP'? However, Hendry's emphasis on this insight deserves praise. Indeed,

> '... the consequence of given data and a given theory model is that *the error is the derived component*, and one cannot make "separate" assumptions about it properties' (p. 73).

One of the strengths of Hendry's methodological observations is his recognition of the importance of model design. I agree that models are econometricians' constructs, and the residuals are '*derived* rather than *autonomous*' and, hence, models can 'be designed to satisfy pre-selected criteria ' (p. 246). The distribution of the residuals cannot be stated *a priori*, as may be done in specific situations of experimental design (see Fisher, 1960). The 'axiom of correct specification' (Leamer, 1978) does not hold, which is why the econometrician has to investigate the characteristics of the derived residuals. Model design aims at constructing models that satisfy a set of desired statistical criteria. Models should be sufficient statistics.

However, this has strong implications for the interpretation of the statistical inferences that are made in econometric investigations. Most importantly, the tests are not straightforward instances of Neyman-Pearson or Fisherian tests.[32] Accommodating the data (description) is crucially different from inference (prediction), in particular if the model that accommodates the data is not supported by *a priori* considerations.[33]

In the case of model design, the model simultaneously 'designs' its hypothetical universe. Hence, the model, given the data, cannot be used for the purpose of inference about the universe as the universe is constructed during the modelling stage, and is unique to the particular set of data which were used to construct it. Fisher (1955, p. 71), criticizing repeated sampling in Neyman and Pearson's theory, argues:

> 'if we possess a unique sample in Student's sense on which significance tests are to be performed, there is always,a s Venn (1876) in particular has shown, a multiplicity of populations to each of which we can legitimately regard our sample as belonging: so that the phrase

[32] See Keuzenkamp and Magnus (1995) for further details about both approaches.

[33] See Howson (1988) for a useful view on accommodation and prediction, from a Bayesian perspective. Howson claims that the predictionist thesis is false. This thesis is that accommodation does not yield inductive support to a hypothesis, whereas independent prediction does. Howson's claim is valid (roughly speaking) if the hypothesis has independent *a priori* appeal. Dharmapala and McAleer (1995), who deal with (objective) truth values instead of degees of belief, share Howson's claim, but quite naturally ignore the condition of *a priori* support. In the kind of econometric model design, where *ad hoc* considerations play an important role, the predictionist thesis remains relevant once dealing with degrees of belief.

> ‘repeated sampling from the same population’ does not enable us to determine which population is to be used to define the probability level, for no one of them has objective reality, all being products of the statistician’s imagination.’

This does not imply that Fisher rejected statistical inference in cases of unique samples. However, he thought experimental design a crucial element for valid inference. Economic theory is not an alternative to experimental design, in particular in macroeconomics where many rival theories are available. Model design (accommodation) is not an alternative either, if the *ad hoc* element in modelling reduces the prior support of the model. The real test, therefore, remains the ability to predict out of sample and in different contexts. Friedman (1940, p. 659) made this point when he reviewed Tinbergen’s work for the League of Nations:

> ‘Tinbergen’s results cannot be judged by ordinary tests of statistical significance. The reason is that the variables with which he winds up ... have been selected after an extensive process of trial and error *because* they yield high coefficients of correlation’ (emphasis in original).

This conveys the same message as Friedman’s reply (in the postscript to Friedman and Schwartz, 1991) to Hendry and Ericsson (1991). Indeed, Hendry (pp. 425-6) also argues that a real test is possible if (genuinely) new data become available, whereas the test statistics obtained in the stage of model design ‘demonstrate the appropriateness (or otherwise) of the *design* exercise’.

The position of Friedman and Schwartz, who do not rely on modern econometric methods but are aware of the fundamental issues in statistics, might be related to Leamer’s (1978) emphasis on sensitivity analysis, in combination with a strong emphasis on the importance of new data. Econometrics may not be the only, or most useful, method to find, or test, interesting hypotheses. Theory is an alternative, or even a careful (non-statistical) study of informative historical events (like the stock market crash).[34] There are few occasions where statistical tests have changed the minds of mainstream economists (with liquidity constraints in consumption behaviour arguably being a rare exception).

The hypothesis that most human behaviour follows an error-correcting pattern has not been accepted in economic textbooks, whereas Hall's consumption model and the more recent related literature is a standard topic in such books. Why is this so? Perhaps it is because the adequacy criteria in model design do not necessarily correspond to the requirements for an increase in knowledge of economic behaviour. Even the ability to 'encompass' Hall's model does not necessarily contribute to progress in economic knowledge. Encompassing, Hendry (p. 440) argues,

> 'seems to correspond to a "progressive research strategy" [...] in that encompassing models act like "sufficient statistics" to summarize the pre-existing state of knowledge.'[35]

As noted above, this is not what Lakatos meant by progress. Hendry is able to provide novel interpretations of existing data or models, but this is not equivalent to predicting novel facts. However, encompassing might provide a viable alternative to the Popperian concept of verisimilitude (closeness to the truth). Popper's formal definition of verisimilitude has collapsed in view of the problem of false theories. Encompassing does not necessarily suffer from this problem due to the availability of pseudo maximum likelihood estimators (where the models do not have to be correctly specified). Here is a potential subject where econometrics may contribute to philosophy of science.

---

[34] See also Summers (1991).

[35] Between the square brackets, among others a reference to Lakatos is given. Hendry and Ericsson (1991, p. 22) make nearly the same claim, using a very subtle change of words. Encompassing is 'consistent with the concept of a progressive research strategy [...] since an encompassing model is a "sufficient representative" of previous empirical findings.' One wonders whether the move from 'statistic' to 'representative' has any deeper meaning (as in the distinction between a 'test' and a 'check').

7. Conclusion: alchemy or science?

The quest for the Holy Grail resulted in a shattering of the brotherhood of the Round Table. Still, some of the knights gained worship. A quest for the DGP in econometrics is more likely to gain econometricians disrespect. Like making gold for the alchemists, searching for the DGP is a wrong aim for econometrics. Relying on the trinity of tests may be respectable, but does not deliver the status of science. The Popperian straitjacket does not fit econometrics (or any other science). Model design makes econometrics as scientific as fashion design. Modelling is an art, and one in which Hendry excels.

Occasionally, Hendry interprets models as ‘useful approximations’ (*e.g.* p. 276). Karl Pearson, a founder of statistics and an early positivist, argued along those lines, and so did other great statisticians in the tradition of British empiricism (in particular, Fisher and Jeffreys). Of course, the issue is: approximations to what? If the answer is the Holy Grail, known as the DGP, then econometrics is destined to be a branch of alchemy or, worse, metaphysics. Popperian cosmetics will not render econometrics a scientific status. If, on the other hand, the answer is approximations to the data, helping to classify the facts of economics, econometrics may join the positivist tradition which, in a number of cases, has yielded respectable pieces of knowledge. Criticism is an important ingredient of this positivist tradition, but not a methodological dogma.

References.

Darnell, Adrian C. and J. Lynne Evans (1990), *The Limits to Econometrics*, Aldershot, Edward Elgar.

Dharmapala, Dhammika and Michael McAleer (1995), Prediction and accommodation in econometric modelling, *Environmetrics*, forthcoming.

Fisher, Ronald A. (1955), Statistical methods and scientific induction, *Journal of the Royal Statistical Society* B **17**, 69-78.

Fisher, Ronald A. (1960), *The design of Experiments*, seventh edition, first published 1935, London, Oliver and Boyd.

Friedman, Milton (1940), Review of Jan Tinbergen, *Statistical Testing of Business Cycle Theories, II: Business Cycles in the United States of America*, *American Economic Review* **30**, 657-61.

Friedman, Milton and Anna J. Schwartz (1991), Alternative approaches to analyzing economic data, *American Economic Review* **81**, 39-49.

Gilbert, Chris (1986), Professor Hendry's econometric methodology, *Oxford Bulletin of Economics and Statistics* **48**, 283-307.

Godfrey, L.G. (1989), Misspecification tests in econometrics, Cambridge, Cambridge University Press.

Goldberger, Arthur S. (1964), *Econometric Theory*, New York, Wiley.

Granger, Clive W.J. and P. Newbold (1974), Spurious regressions in econometrics, *Journal of Econometrics* **2**, 111-20.

Hacking, Ian (1983), *Representing and intervening*, Cambridge, Cambridge University Press.

Hendry, David F. (1987), Econometric methodology: a personal perspective, In Truman F. Bewley (ed.), *Advances in Econometrics*, Cambridge, Cambridge University Press.

Hendry, David F. (1993), *Econometrics: Alchemy or Science; Essays in Econometric Methodology*, Oxford, Blackwell.

Hendry, David F. and Neil R. Ericsson (1991), An econometric analysis of U.K. money demand in *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna J. Schwartz, *American Economic Review* **81**, 8-38.

Hendry, David F., Edward E. Leamer and Dale J. Poirier (1990), The ET dialogue: a conversation on econometric methodology, *Econometric Theory* **6**, 171-261.

Howson, Colin (1988), Accommodation, prediction and Bayesian confirmation theory, *PSA* **2**, 381-92.

Keuzenkamp, Hugo A. and Anton P. Barten (1995), Rejection without falsification, on the history of testing the homogeneity condition in the theory of consumer demand, *Journal of Econometrics*, **67**.

Keuzenkamp, Hugo A. and Jan Magnus (1995), On tests and significance in econometrics, *Journal of Econometrics*, **67**.

Keuzenkamp, Hugo A. and Michael McAleer (1995), Simplicity, scientific inference and econometric modelling, *Economic Journal*, **105**, January.

Keynes, John Maynard (1939), Professor Tinbergen's method, *Economic Journal* **49**, 558-68.

Lakatos, Imre (1970), Falsification and the methodology of scientific research programmes, in: Lakatos and Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press, 91-196.

Morgan, Mary S. (1990), *The History of Econometric Ideas*, Cambridge, Cambridge University Press.

Pagan, Adrian R. (1987), Three econometric methodologies: a critical appraisal, *Journal of Economic Surveys*, **1**, 3-24.

Pagan, Adrian R. (1994), Three econometric methodologies: an update, to appear in Les Oxley *et al.* (eds.), *Surveys in Econometrics*, Oxford, Basil Blackwell.

Pollock, Stephen S. G. (1992), Lagged dependent variables, distributed lags and autoregressive residuals, *Annales d'Économie et de Statistique* **28**, 143-64.

Spanos, Aris (1986), *Statistical Foundations of Econometric Modelling*, Cambridge, Cambridge University Press.

Summers, Lawrence (1991), The scientific illusion in empirical macroeconometrics, *Scandinavian Journal of Economics* **93**, 129-48.

Yule, George Udny (1926), Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series, *Journal of the Royal Statistical Society* **89**, 1-64.