Econometrics-Alchemy or Science?
Author(s): David F. Hendry

Source: *Economica*, New Series, Vol. 47, No. 188 (Nov., 1980), pp. 387-406
Published by: Blackwell Publishing on behalf of The London School of Economics and Political Science and The Suntory and Toyota International Centres for Economics and Related Disciplines
Stable URL: http://www.jstor.org/stable/2553385
Accessed: 21/09/2011 03:47

Economica, **47**, 387–406

Econometrics—Alchemy or Science?

By David F. Hendry

The London School of Economics

I. Alchemy and Science

While there are many distinguished precedents for public lectures at the School being discourses about subjects on which the speaker is an evident amateur, I do not intend to discuss at length either "Scientific Method" or the general relationship between "Alchemy" and "Science". No doubt my colleagues in the Philosophy and Scientific Method Department will be greatly relieved. Nevertheless, some background will be useful, especially to distinguish connotative from denotative aspects of "alchemy" and "science".

Alchemy denotes the putative art of transmuting base metals into noble ones, a possibility implicit in Greek *theories* of matter; as such, alchemical *experiments* helped focus chemical effort and could be interpreted as embryonic systematic chemistry. In this sense, my question is simply a matter of timing—after all, the title does not juxtapose astrology and science!

The familiar connotations of alchemy are less happy, and are well represented by Ben Jonson's erudite comedy *The Alchemist* (1612) with its bogus and obscurantist "puffer" (so-called from the phrenetic use of bellows in transmutation attempts) called Subtle. That the pejorative sense is now dominant may derive partly from the mystical associations of the quest for the "Philosophers' Stone" and partly from "recipes" for simulating gold using alloys of base metals; intended to deceive the public, such recipes may well have deceived many alchemists themselves. The relevance of these comments to the current state of econometrics will be apparent shortly.

Precisely what "science" denotes is remarkably unclear, but the present mental associations of objectivity and progress ensure that simply using this prestigious epithet confers an air of authority; to wit, the London School of Economics and Political Science—would anyone attend the London School of Economics and Political Alchemy? Parenthetically, the implication of authority is rather odd given that the fifteenth-century revival of science in western Europe was a reaction against argument by authority. In any case, the high reputation of the physical sciences may decline in the next decade should public expectations on environmental control remain unfulfilled; if there are many more nuclear accidents, we may yet be glad to be called "political economists" rather than "economic scientists".

What is this thing called Science? (See the excellent text by Chalmers, 1976.) During an address under the shadows of Sir Karl Popper and the late Imre Lakatos, whose distinguished contributions have revolutionized our understanding of "Science", there is a distinct risk of yielding several hostages to fortune by trespassing on a debate that has flourished since Francis Bacon (see Popper, 1968, 1969; and Lakatos, 1974). This danger notwithstanding, an adequate if condensed view is as follows.

Science is a public process. It uses systems of concepts called theories to help interpret and unify observation statements called data; in turn the data are used to check or "test" the theories. Theory creation may be inductive, but demonstration and testing are deductive, although in inexact subjects, testing will involve statistical inference. Theories that are at once simple, general and coherent are valued as they aid productive and precise scientific practice. In particular, restrictiveness increases the hazards of possible rejection and hence augments "plausibility" if disconfirmation does not occur. Although objectivity and potential falsifiability against data are crucial to science, in practice observations are theory-dependent, rejections can be rationalized (often leading to degenerate research programmes), and even when evidence is highly unfavourable and reasonable alternative theories exist, views are usually changed only slowly: after all, we are discussing a *human* endeavour! As Baron Turgot expressed the matter in 1749: "Suppositions which are arrived at on the basis of a small number of poorly understood facts yield to suppositions which are less absurd, although no more true" (Meek, 1973, p. 45). The history of natural science (for example, Mason, 1962) provides many instances of ideas derided at conception which are taken as axiomatic later, and Kuhn (1962) has argued that science actually progresses through "revolutionary" changes in basic theoretical frameworks brought about by cumulative failures to solve problems. Note that in this characterization experimentation may be a useful, but is not an essential, attribute.

Alchemy could well have remained "scientific"—perhaps as a degenerate research programme or a rejected theory—but instead it seems to have turned to mysticism and away from objectivity. Stanislas de Rola (1973) argues that the unfortunate connotations of alchemy are undeserved since "immature science" is a false interpretation of alchemy and "true" alchemy is actually a secret art striving for the "absolute". Feel free to choose the intended meaning of "Alchemy" in my title!

II. Econometrics

Unfortunately, I must now try to explain what "econometrics" comprises. Do not confuse the word with "econo-mystics" or with "economic-tricks", nor yet with "icon-ometrics". While we may indulge in all of these activities, they are not central to the discipline. Nor are econometricians primarily engaged in measuring the heights of economists.

A more accurate idea of the subject is provided in the constitution of the Econometric Society, founded in 1930, which defined its main objectives as "the advancement of economic theory in its relation to statistics and mathematics" (*Econometrica*, 1933, p. 1). In this broad sense, econometrics commences an analysis of the relationships between economic variables (such as quantities and prices, incomes and expenditures, etc.) by abstracting the main phenomena of interest and stating theories thereof in mathematical form. The empirical usefulness of the resulting "models" is evaluated using statistical information of supposed relevance, and econometrics in the narrow sense (used hereafter) concerns the interpretation and analysis of such data in the context of "established" economic theory.

Thus, econometric theory is the study of the properties of data generation processes, of techniques for analysing economic data, methods of estimating numerical magnitudes of parameters with unknown values, and procedures for testing economic hypotheses; it plays an analogous role in primarily non-experimental disciplines to that of statistical theory in inexact experimental sciences (for example Blalock, 1961). As expressed by Wold (1969), "Econometrics is seen as a vehicle for fundamental innovations in scientific method, above all, in the development of operative forecasting procedures in non-experimental situations." In Wold's view, econometrics needs to overcome both a lack of experimentation (which precludes reproducible knowledge) and the passivity of forecasts based on extrapolative methods.

Applied and empirical econometrics are sometimes regarded as separate "engineering" branches of the subject, literally involving the mere application of standard statistical methods to economic data. Since, to quote Frisch (1933), "the mutual penetration of quantitative economic theory and statistical observation is the essence of econometrics", the greatest loss from our inability to experiment may be the artificial divisions it promotes between data collectors, data users, econometric theorists and mathematical economists.

The need for quantitative empirical knowledge to answer questions involving changes in economic variables has been adequately promulgated by Schumpeter (1933) and Phillips (1956) (the former argued that economics is really the most quantitative of *all* the sciences since economic quantities are made numerical by life itself whereas other subjects had to *invent* their measurement processes). For predicting the consequences of changes, forecasting likely future outcomes and controlling variables to attain objectives, econometric models play a central role in modern economics. Substantial resources have been devoted to empirical macroeconometric models which comprise hundreds or even thousands of statistically calibrated equations, each purporting to represent some autonomous facet of the behaviour of economic agents such as consumers and producers, the whole intended to describe accurately the overall evolution of the economy.

Despite its obvious potential, econometrics has not had an easy time from many who have made major contributions to the development of economics, beginning from Keynes' famous review in 1939 of Tinbergen's book, *Statistical Testing of Business-Cycle Theories*. In an oft-quoted passage in his Comment (1940, p. 156) Keynes accepts that Tinbergen's approach is objective but continues:

> "No one could be more frank, more painstaking, more free from subjective bias or *parti pris* than Professor Tinbergen. There is no one, therefore, so far as human qualities go, whom it would be safer to trust with black magic. That there is anyone I would trust with it at the present stage, or that this brand of *statistical alchemy* is ripe to become a branch of science, I am not yet persuaded. But Newton, Boyle and Locke all played with Alchemy. So let him continue. [Keynes, 1940, p. 156; my italics].

It is interesting to record the following quotation from Geoffrey Keynes (1946): "Newton was *not* the first of the Age of Reason. He was the last of the magicians ... an unbridled addict [of alchemy] ... [during] the very years when he was composing the *Principia*." Oh that econometrics had such alchemists as Newton! Again the issue is one of timing since Keynes, despite his trenchant criticisms, does *not* liken econometrics to a theoryless reading of entrails as some seem to believe. (For a fuller discussion of Keynes' views on econometrics, see Patinkin, 1976.) Notwithstanding Keynes' comments, Tinbergen was later joint recipient of the first Nobel Prize in *economics*.

An echo of this debate recurs in the early 1970s. For example, following a sharp critique of mathematical economics as having "no links with concrete facts", Worswick (1972) suggests that some econometricians are not "engaged in forging tools to arrange and measure actual facts, so much as making a marvellous array of *pretend-tools*" (my italics). In the same issue of the *Economic Journal*, Phelps Brown (1972) also concludes against econometrics, commenting that "running regressions between time series is only likely to deceive". Added to these innuendoes of "alchemical" practices, Leontief (1971) has characterized econometrics as "an attempt to compensate for the glaring weakness of the data base available to us by the widest possible use of more and more sophisticated statistical techniques". To quote Hicks, "the relevance of these methods [i.e. econometrics] to economics should not be taken for granted; ... [Keynes] would not have been surprised to find that ... econometrics is now in some disarray" (1979, p. xi). With the manifest breakdown in the early 1970s of the large empirical macro-econometric systems, outside scepticism does not bear mention.

Rather than abandon the study of econometrics or reply to those criticisms by quoting equally eminent authorities who hold more favourable views (for example, Koopmans, 1957, 1979; and Stone, 1951), I should like instead to demonstrate the scientific status of econometrics by first showing alchemy at work empirically. This will enable us to understand the sense in which the quoted criticisms are valid, and by explaining why various apparently alchemical results are obtained *en route* my approach will suggest constructive strategies for enhancing the role of scientific method in econometrics.

So let us practise alchemy!

III. Econometrics as Alchemy

Econometricians have found their Philosophers' Stone; it is called regression analysis and is used for transforming data into "significant" results! Deception is easily practised from false recipes intended to simulate useful findings, and these are derogatively referred to by the profession as "nonsense regressions" (although I could not find an equivalent of "puffer", regressor already having another meaning).

Figure 1 presents (seasonally adjusted) quarterly time-series data for the UK over the period 1964(II)–1975(II) relating to the age-old and seemingly unresolved controversy concerning the effect of money (here personal sector M3) on prices (here the consumer price index); the variables, denoted M and P, are plotted on a logarithmic scale. Advance warning that "alchemy" may be present could be gleaned from the letters to *The Times*, 4–6 April 1977, where Llewellyn and Witcomb establish a higher correlation between annual inflation and cases of dysentery in Scotland (one year earlier) than Mills obtained between inflation and the rate of change of excess money supply (two years before).

The plot of M against P in Figure 2 seems to confirm their close relationship (the correlation is over 0·99). Regression estimates of the explanation of P by M yield the results in Figure 3; the fit is impressive as M "explains" 98 per cent of the variation of P and has a "significant" coefficient (the quantities in parentheses are estimated standard errors).[1] The residuals are systematic rather than random, but

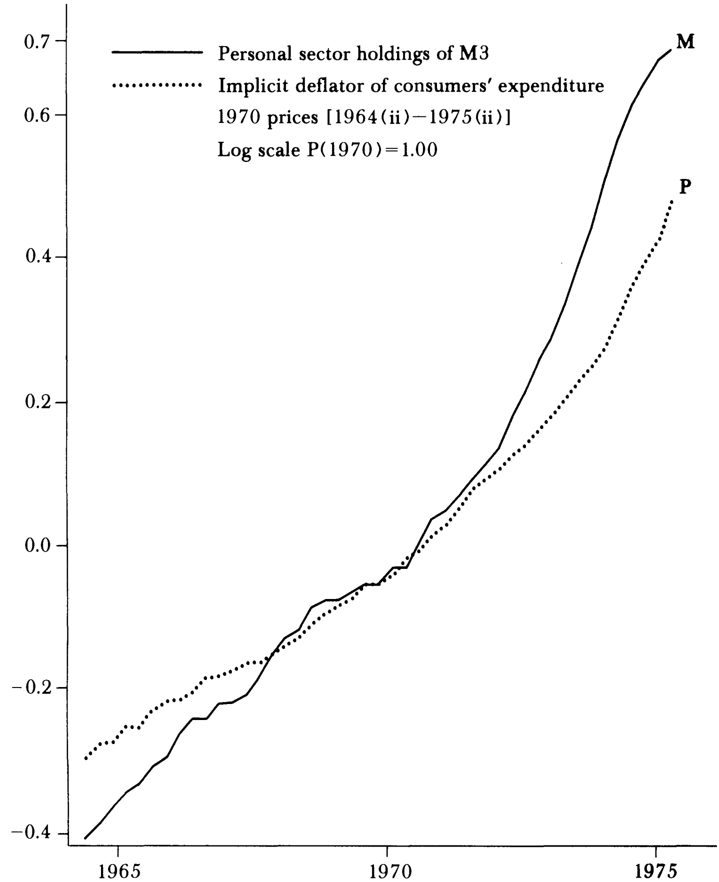

FIGURE 1

this so-called "nuisance" of autocorrelation (see Hendry and Mizon, 1978) can be "eliminated" by suitably transforming the equation to introduce lagged values of the variables (i.e. by the values of the variables in previous periods, denoted M_{t-1}, P_{t-1}): see Figure 4. The squared correlation is now 0·9985 but the money variables no longer significantly influence P and a prediction test rejects the constancy of the parameters of the equation. Evidently, we can make money matter or not by appropriate specification of the model, and hence "(Self?) deception" is easy by selecting whichever finding "corroborates one's theory".

A second example will clarify this issue. Hendry's theory of inflation is that a certain variable (of great interest in this country) is the "real cause" of rising prices. I am "certain" that the variable (denoted C) is exogenous, that causality is from C to P *only* and (so far as I am aware) that C is outside government control although data are readily available in government publications. Figure 5 shows the quarterly time series (seasonally *un*adjusted) and Figure 6 the cross-plot of P against C (again in logs). There is evidently a close but nonlinear relationship, and regression analysis assuming a quadratic equation yields the results in Figure 7. As earlier, there is a "good fit", the coefficients are "significant", but autocorrelation remains and the equation predicts badly. However, assuming a

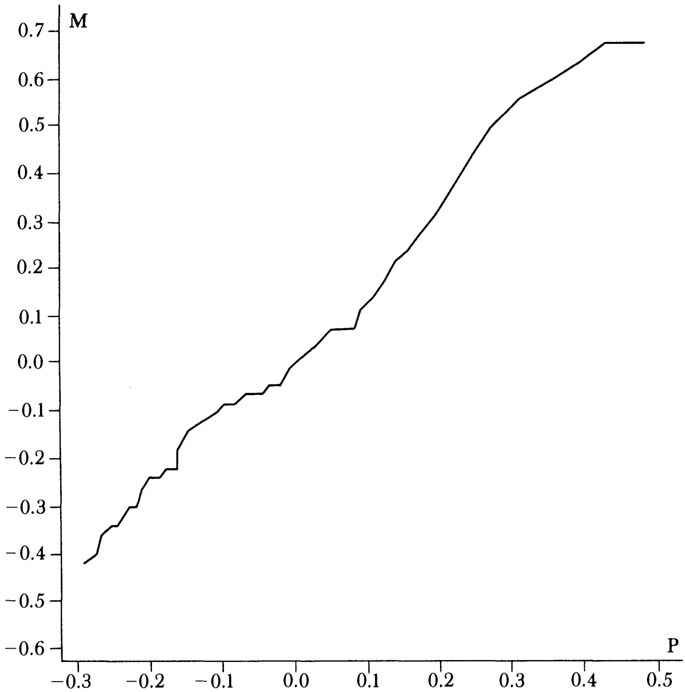

FIGURE 2

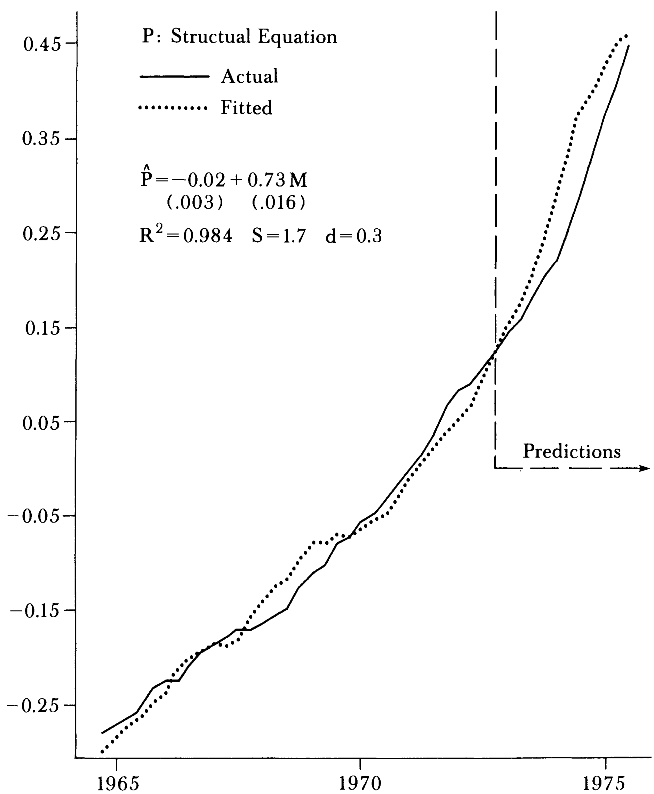

FIGURE 3

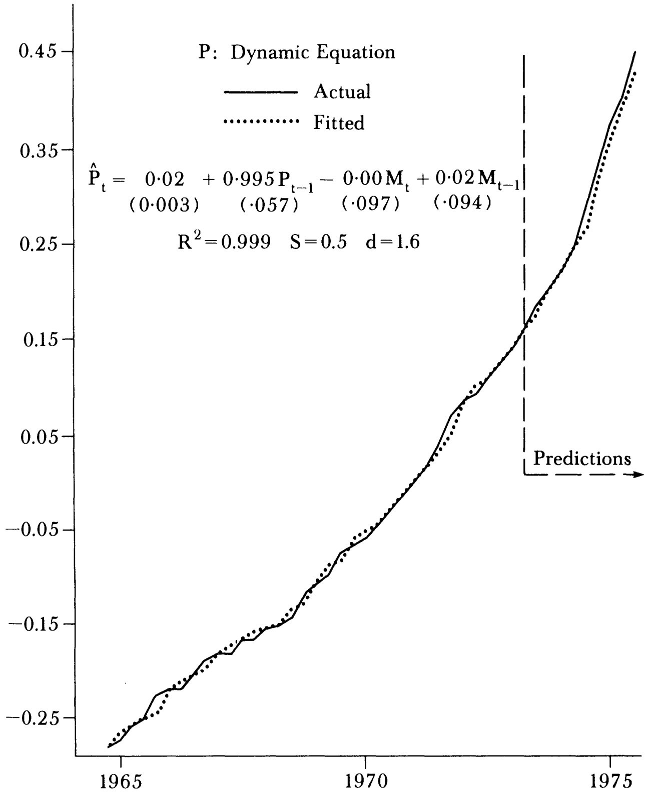

Figure 4

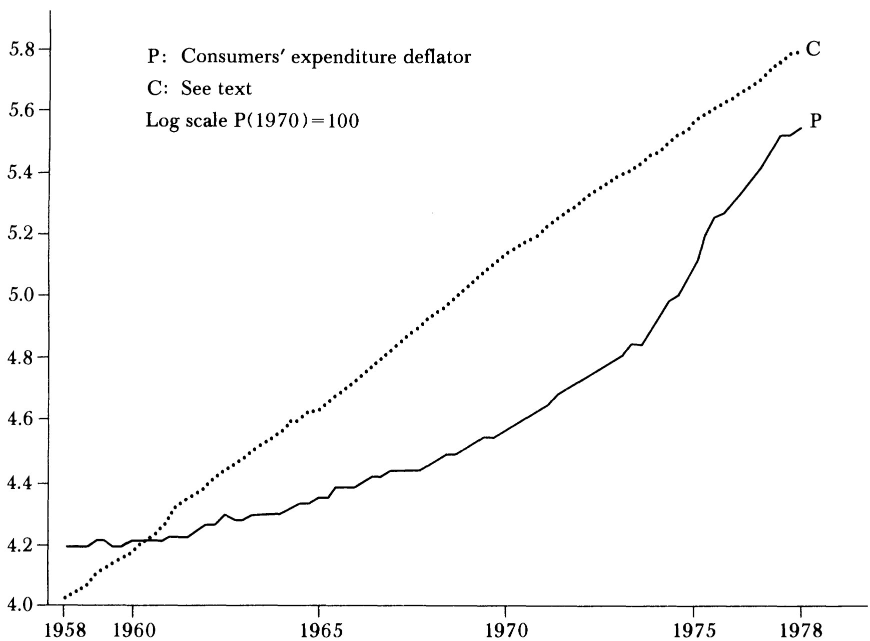

Figure 5

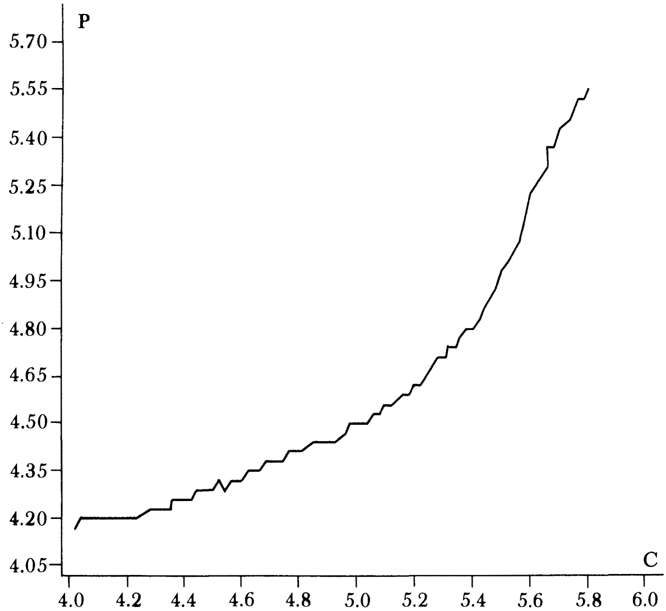

FIGURE 6

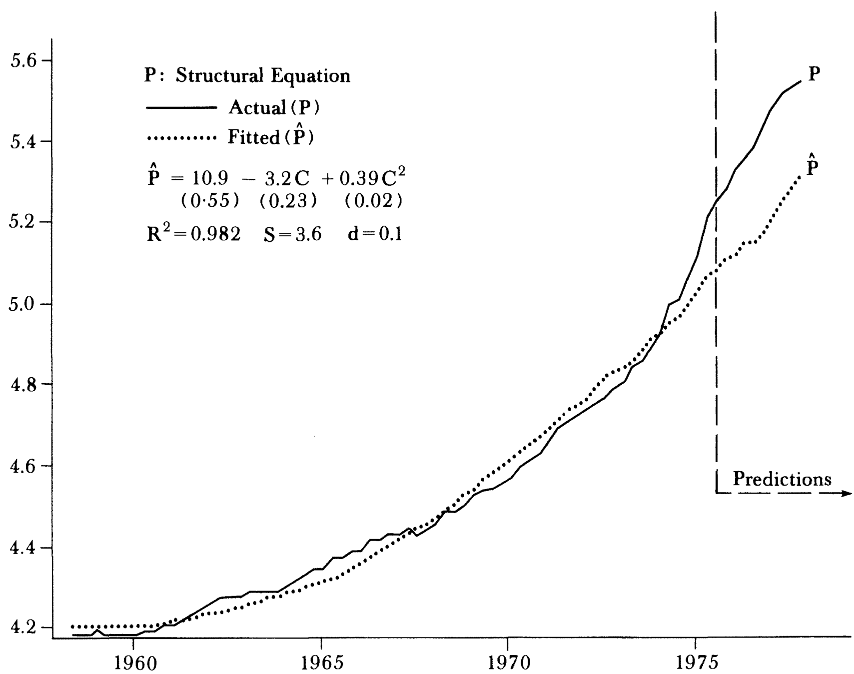

FIGURE 7

first-order autoregressive error process[2] at last produces the results I anticipated (see Figure 8); the fit is spectacular, the parameters are "highly significant", there is no obvious residual autocorrelation (on an "eyeball" test), and the predictive

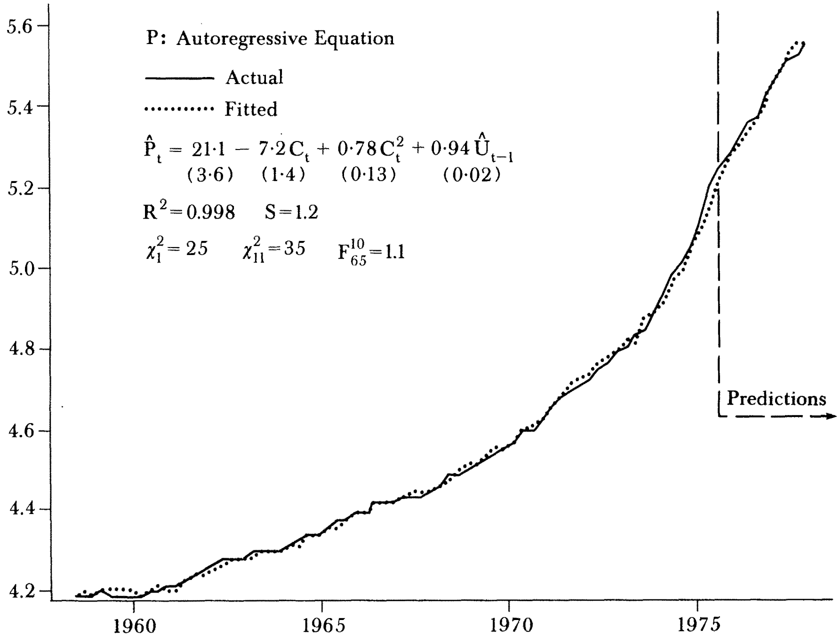

FIGURE 8

test does not reject the model. My theory performs decidedly better than the naive version of the monetary one, but, alas, the whole exercise is futile as well as deceitful since *C* is simply *cumulative rainfall* in the UK. It is meaningless to talk about "confirming" theories when spurious results are so easily obtained.

Since correlaton does not entail any direction of causation, perhaps the rapid inflation explains our wet weather? One must regret the omission of such an important theory from the otherwise excellent *History of the Theories of Rain* by Middleton (1965).

Doubtless, some equations extant in econometric folklore are little less spurious than those I have presented. Before you despair at this hopeless subject, the statistical problem just illustrated was analysed in one of its manifestations by Yule in 1926 and has been re-emphasized many times since (see in particular Granger and Newbold, 1974). The crucial factor for my argument is that *before* doing these regressions the relevant theory enabled me to *deduce* what would occur and hence to construct the desired examples on my first try—what could be more scientific? We understand this problem and have many tests for the validity of empirical models (those just quoted duly fail two such tests[3]). We even have theories that reveal that prediction need not be a powerful test of a model since false models can manifest parameter constancy (Hendry, 1980).

Such understanding is well past the stage of alchemy even if some editors can be persuaded to publish on the basis of econometric fools-gold: *caveat emptor*, but do not denigrate the whole subject. That modern chemistry can explain alchemical results is a confirmation of its scientific status, not cast into doubt by any modern charlatans who might use chemical theory to simulate gold. The case for scientific econometrics rests instead on *best* practice empirical work such as Sargan (1964)—a precursor of many useful developments in recent econometrics. My discussion also highlights that an essential requirement of any useful model in a *non-experimental* subject is that it can explain why previous *false* models provided their observed results (Davidson *et al.*, 1978).

To conclude this section, it must be stressed that none of the evidence presented lends support to, or casts doubt on, any theory of aggregate price determination, nor is it asserted that "nonsense" regressions of the type illustrated constitute the basis of the criticisms noted earlier, a point amplified below.

IV. Econometrics' Problems

To quote Patinkin (1976), "though not all of Keynes' criticisms were well taken ... I find it somewhat depressing to see how many of them are, in practice, still of relevance today". Forty years after Keynes wrote, his review should still be compulsory reading for all who seek to apply statistical methods to economic observations. Taken literally, Keynes comes close to asserting that no economic theory is ever testable, in which case, of course, economics itself ceases to be scientific—I doubt if Keynes intended this implication. However, his objections make an excellent list of what might be called "problems of the linear regression model", namely (in modern parlance): using an incomplete set of determining factors (omitted variables bias); building models with unobservable variables (such as expectations), estimated from badly measured data based on index numbers (Keynes calls this the "frightful inadequacy of most of the statistics"); obtaining "spurious" correlations from the use of "proxy" variables and simultaneity as well as (and I quote) the "mine [Mr Yule] sprang under the contraptions of optimistic statisticians"; being unable to separate the distinct effects of multicollinear variables; assuming linear functional forms not knowing the appropriate dimensions of the regressors; mis-specifying the dynamic reactions and lag lengths; incorrectly pre-filtering the data; invalidly inferring "causes" from correlations; predicting inaccurately (non-constant parameters); confusing statistical with economic "significance" of results and failing to relate economic theory to econometrics. (I cannot resist quoting Keynes again—"If the method cannot prove or disprove a qualitative theory and if it cannot give a quantitative guide to the future, is it worth while? For, assuredly, it is not a very lucid way of describing the past".) To Keynes' list of problems, I would add stochastic mis-specification, incorrect exogeneity assumptions (see Koopmans, 1950 and Engle *et al.*, 1979), inadequate sample sizes, aggregation, lack of structural identification and an inability to refer back uniquely from observed empirical results to any given initial theory.

That the subject is exceedingly complicated does not entail that it is hopeless. Considerable progress has been made on the technical aspects, such as studying the consequences of the various problems just listed, designing means of detecting these, developing methods that mitigate some of their ill effects or handle several complications at once, and analysing the properties of estimators when the sample size is small (see Sargan, 1976; and Phillips, 1977, *inter alia*). Much of this technical work is essential background to understanding and correctly interpreting empirical findings and, although some work may have turned out to be otiose in retrospect, the ever-increasing level of technique is not a symptom of alchemy. To borrow Worswick's phrase, whether or not "econometric escalation" is justifiable will depend on whether it facilitates clearer findings or camouflages tenuous evidence.

Empirical practice has tended to lag behind the theory "frontier" with unfortunate consequences. Well before the oil crisis, critics suggested that macroeconometric systems were seriously mis-specified and hence would manifest predictive failure if changes in the process generating the data merely altered the *correlation* structure of the variables (see, for example, the discussion in Hickman, 1972). Many of the specification mistakes were obvious and relatively easy to correct, and doing so might have helped to prevent the models failing so badly just when they were most needed. Even so, that cataclysm and similar government-induced events are one of the few ways in which false models can be rejected—econometrics may be the sole beneficiary from government manipulation of the economy. Without wishing to look this particular gift horse in the mouth, dare one suggest that controlled experiments could be more informative than inadvertent and uncontrolled ones?

At the micro-level, experimentation is occurring (for example on diurnal variation in energy consumption with changing tariff structures). Regrettably, experimental "control" is proving elusive, especially for relativities and dynamical and inertial patterns of behaviour. Despite such difficulties, experimentation in economics merits far greater resources than the meagre financial ration currently allocated by our political masters allows. This is *not* a criticism of the Social Science Research Council, which has played a major role in supporting econometric research in the UK from a very limited budget roughly equal to the interest on the annual grant to the Science Research Council. As Leontief (1971) expressed the matter, "the scientists have their machines while the economists are still waiting for their data". To mention one constructive step, the collection of panel data would be of very great assistance in testing economic theories at a disaggregated level.

Economic data are notoriously unreliable (see for example, Morgenstern, 1950) and in an important sense econometrics is little more than an attempted solution to our acute shortage of decent data. Yet accurate observation is vital. To take one important example, a variable like "real personal disposable income" is extremely difficult to "measure" accurately and a constant price series of after-tax "income" of the personal sector bears little relation to the economist's concept of "income" (as defined, for example, by Hicks, 1939, Chapter 14). Unfortunately, discrepancies in measuring income may have major policy implications. If income is measured using *real* rather than nominal interest rates to ensure that changes in real wealth are equal to real income less real expenditure, then the ratio of consumers' expenditure to adjusted income has not fallen particularly sharply, unlike the ratio of the unadjusted series (see Hendry and von Ungern-Sternberg, 1980). Thus, the savings ratio "properly measured" may not have risen at all. A non-obvious converse is that the government may not be in deficit. A recent Bank of England study (Taylor and Threadgold, 1979), has done the appropriate "inflation accounting" at the macro-level, with dramatic results: if the implicit tax created by inflation eroding the real value of those financial assets that are public sector debt is added to government revenue, and subtracted from personal sector savings, then the government has frequently been in real surplus and the private sector in real deficit (see their Table C).

One might anticipate that the massive nominal borrowing by the public sector, now apparently the main focus of government policy, has altered the "national debt", and this expectation is quite correct—in a most surprising way. Certainly, the nominal level of the debt has increased rapidly, but as Reid (1977) has shown, the ratio of national debt to national income—which seems a sensible measure of real public indebtedness—was in 1975 similar to the value prevailing at the end of the *last* century and hence probably close to its *lowest* value since the Napoleonic wars!

An implication of these two statistics (namely, the real government surplus and the falling real debt ratio) is that the state of *net* government indebtedness to the rest of the economy must have been changing. Hibbert (1979) has kindly provided the orders of magnitude that he has recently calculated, and even with all the usual *caveats* about definitions and data accuracy, the numbers are stunning. In 1957 the public sector was a net debtor to the tune of about 8 per cent of total net national wealth; by 1966 it had become a net creditor to a similar extent, and by 1975 the public sector owned 26 per cent of net national wealth. The statistics mesh consistently and reveal enormous and very rapid real changes behind the monetary veil. Such an outcome does not seem to have been an intended consequence of any postwar government policy. Yet a further implication of these data is that the recession manifest in the current high level of unemployment may be due in part to the implicit government surplus with the Public Sector Borrowing Requirement (PSBR) being a mere monetary epiphenomenon.

The facts in this last conjecture are fairly well established, although the interpretation and policy implications may not be unique. For my purposes, the conjecture need not even be correct since my point is that attempts to reduce the PSBR in the belief that it is a "cause" of inflation rather than a "consequence" of recession will impose major costs on society *if that belief is mistaken.* Would it not have been worthwhile to devote rather greater resources to researching the matter beforehand? Yet our government has reduced the SSRC's budget—and in its calculations of "average student costs" implicitly values the entire research output of the university sector at zero. However little the government might value our theories or empirical evidence, to base policy on hope or belief really is alchemy. Keynes, this time in his *General Theory* (1936, p. 383), provides the most apt quotation: "Practical men, who believe themselves to be quite exempt from any intellectual influences, are actually the slaves of some defunct economist." I hesitate to continue his quote—but he did then say "Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back." Hopefully, that will not be the fate of this lecture a few years hence.

Stretching somewhat my argument about the value of data, endless billions of dollars have been spent on space exploration by the United States government just to collect *a few observations* of some lumps of rock and gas (with incidental kudos, "technical spin-off" and tenuous "defence" advantages). What government anywhere has spent one-thousandth as much in deliberately observing (experimentally or non-experimentally) or trying to understand an economic system of at least equal importance to our lives?

V. A Structure for Econometrics

Econometricians are the natural critics of economists' empirical findings, and although that is an easy way to make enemies, the counter-criticisms of econometrics noted earlier are not simply the revenge of the aggrieved. However, their valid basis is *not* econometric alchemy but a misallocation of resources. (This is not a new theme; for an earlier debate, see Orcutt (1952) and the following discussion.)

What should have become a relatively minor aspect of the subject, namely deriving methods for estimating the parameters of known models, has been accorded the centre of the stage as casual perusal of any current econometrics textbook will confirm. The rapid development of computer speeds and storage capacity should by now have relegated most of estimation theory to footnotes about numerical approximations and refocused attention on all of the issues surrounding methodology, inference, model formulation and equation selection (see Griliches, 1974; Leamer, 1978; and Mizon, 1977). We have responded as quickly as Diplodocus used to move on a frosty morning and should remember that the Saurischia once dominant are now extinct.

The economic system is the outcome of centuries of adaptive human behaviour; agents seem to optimize their "state" given the environment, which adapts in response both socially and physically. Econometricians conceptualize this system as a complex nonlinear, interdependent, multivariate, disequilibrium dynamical process dependent on agents' expectations and their adjustments, subject to random shocks, and involving many phenomena that are unobservable; relevant time-series data are inaccurate, exist for only short periods and for a few major variables; economic theories are highly simplified abstractions usually of a comparative static form invoking many explicit *ceteris paribus* clauses (with yet others implicitly required), most of which are invalid in empirical applications—little wonder our macroeconometric representations are less than perfect.

This conceptualization is the real basis for Keynes' critique, but instead of construing the issue as one of "problems for the linear model", turn the matter on its head and begin with a characterization of the economy that does have the relevant properties. As elsewhere, it may pay to take an overview to be simplified if allowable rather than attempt to generalize a simple approach in many different directions simultaneously. A crude schematic structure for econometrics is as follows. To a first approximation, after suitably transforming the original variables (with all nonlinearities allocated to identities), many data generation processes in economics can be conceived of as (see, *inter alia*, Richard, 1980):

$$(1) \qquad \mathbf{y}_t/\mathbf{z}_t \sim N(\mathbf{\Pi z}_t, \mathbf{\Omega}) \qquad (t = 1, \ldots, T)$$

where $\mathbf{y}_t$ is a vector of endogenous variables, $\mathbf{z}_t$ is a vector of all relevant past and present information (so that $E(\mathbf{y}_t/\mathbf{z}_t) = \mathbf{\Pi z}_t$ where E denotes the expectations operator) and $\mathbf{x}_t \sim NI(\boldsymbol{\mu}, \mathbf{\Psi})$ denotes a variate that is normally and independently distributed, with a mean of $\boldsymbol{\mu}$ and a covariance matrix of $\mathbf{\Sigma}$. The parameter matrix $(\mathbf{\Pi}, \mathbf{\Omega}) = \mathbf{P}$ is taken as approximately constant by working in a sufficiently large (but assumed finite) dimensional parameter space. Normality is a convenient fiction which restricts attention to sample information in the first two moments of the data, and independence of successive observations is achieved by construction. For sufficiently large T, accurate data and knowledge of both the required data transformations and the composition of $\mathbf{z}_t$, the enormous number of parameters in $\mathbf{P}$ could be estimated directly using the fact that (1) defines the likelihood function:

$$(2) \qquad L(\mathbf{P}; \mathbf{y}_t/\mathbf{z}_t).$$

An "economic theory" corresponds to asserting that $\mathbf{P}$ depends on only a smaller number of parameters, denoted by the vector $\boldsymbol{\theta}$, and written as:[4]

$$\text{(3)} \qquad \mathbf{P} = \mathbf{f}(\boldsymbol{\theta}) \qquad \boldsymbol{\theta} \in \boldsymbol{\Theta}$$

where $\boldsymbol{\Theta}$ is the paramater space; if $\boldsymbol{\theta}$ is identifiable (i.e. uniquely entailed by $\mathbf{P}$) then all hypotheses like (3) can be tested using the principle due to Wald (1943).

In terms of my discussion of "science", estimation of $\mathbf{P}$ hardly qualifies and is far from providing a *simple* theory. A major role of equation (3) is to limit the number of variables that have to be considered (which is a crude application of Occam's Razor) but the real case against "measurement without theory" has been powerfully presented by Koopmans (1949) in his well known debate with Vining. Many of my present criticisms were noted by *both* parties to that debate. Accepting that we must work within the best available economic theory framework to contribute towards scientific knowledge, the econometric problem arises because the scale of the model and the paucity of the available observations preclude direct estimation of $\mathbf{P}$ (but see Sargent and Sims, 1977) and indeed of $\boldsymbol{\theta}$. Attention is thereby focused on submodels and hence on the weak exogeneity properties of the "regressor" variables in the submodels. If $L(.)$ can be factorized in terms of both data and parameters such that:

$$\text{(4)} \qquad L(.) = L_1(\boldsymbol{\theta}_1, \mathbf{y}_{1t}/\mathbf{y}_{2t}, \mathbf{z}_t) L_2(\boldsymbol{\theta}_2; \mathbf{y}_{2t}/\mathbf{z}_t) \qquad \text{where} \qquad (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta_1 \times \Theta_2$$

so that any changes in either $\boldsymbol{\theta}_i$ leaves the other *unaffected* (for a precise statement, see Engle *et al.*, 1979) and $\boldsymbol{\theta}_2$ are "nuisance parameters", then $L_1(.)$ can be analysed separately from $L_2(.)$ (Koopmans, 1950). In such a case, $\mathbf{y}_2$ is said to be weakly exogenous for $\boldsymbol{\theta}_1$ and $\mathbf{y}_{2t}$ can be taken as given when analysing the submodel that determines $\mathbf{y}_{1t}$. One interesting implication is that variables about which agents form "rational expectations" cannot be taken as weakly exogenous, since, by hypothesis, $\boldsymbol{\theta}_1$ depends on $\boldsymbol{\theta}_2$ in such models.

Even assuming that no mistakes have been made in formulating $L_1(.)$ and that the dimensionality is tractable, it is still unlikely that detailed analysis of the likelihood function will be feasible and some summarization will prove essential (Edwards, 1972). *Estimation* theory concerns alternative rules of attaching numbers to $\boldsymbol{\theta}_1$ given the data, and this can be done in (infinitely) many ways which can have very different properties. Nevertheless, the entire topic can be resolved by noting that (for $\mathrm{L} = \log_e L$):

$$\text{(5)} \qquad \frac{\partial \mathrm{L}_1}{\partial \boldsymbol{\theta}_1} = \mathbf{q}_1(\boldsymbol{\theta}_1)$$

is an *estimator-generating equation* in that other estimators can be interpreted as *approximations* to solving $\mathbf{q}_1(\boldsymbol{\theta}_1) = \mathbf{0}$ (see Hendry, 1976, based on ideas considered by Durbin, 1963). Since computers have greatly alleviated the need to choose approximations that minimize the computational burden, we may as well solve for the most likely value of $\boldsymbol{\theta}_1$, i.e. $\hat{\boldsymbol{\theta}}_1$ such that $\mathbf{q}_1(\hat{\boldsymbol{\theta}}_1) = \mathbf{0}$ and $(\partial \mathbf{q}_1 / \partial \boldsymbol{\theta}_1')|_{\hat{\boldsymbol{\theta}}_1}$ is negative definite (unless the likelihood function is such that the summarization in (5) will be misleading). Inference is also almost entirely dependent on $\mathbf{q}(.)$ (see for example, Rao, 1965, and Breusch and Pagan, 1980), so we can proceed to other matters.

Additional problems which are less easily solved are, first, that at present $\mathbf{f}(\boldsymbol{\theta})$ is based on an excessively idealized abstraction (which is more a guide to how the econometric model should look if the idealized state were to occur than a useful set of restrictions for imposing on data), and, second, that the structure and composition of $\mathbf{z}_t$ are unknown. Thus we have "econometric modelling", that activity of matching an incorrect version of (3) to an inadequate representation of (1), using insufficient and inaccurate data. The resulting compromise can be awkward, or it can be a useful approximation which encompasses previous results, throws light on economic theory and is sufficiently constant for prediction, forecasting and perhaps even policy. Simply writing down an "economic theory", manipulating it to a "condensed form" (see Desai, 1979) and "calibrating" the resulting parameters[5] using a pseudo-sophisticated estimator based on poor data which the model does not adequately describe constitutes a recipe for disaster, not for simulating gold! Its only link with alchemy is self-deception.

As an illustration consider the transactions demand for money. In an equilibrium world with constant transactions technology and static expectations, agents are assumed to keep a constant ratio between nominal (real) money and nominal (real) income:

$$(6) \qquad M/PY = K(.)$$

Between such worlds, $K(.)$ will be lower if interest rates (r) or inflation ($\dot{p}$) are higher, yielding, for example,

$$(7) \qquad M = K^* PYr^{\alpha}(1 + \dot{p})^{\beta} \qquad \alpha, \beta < 0.$$

In spite of the strong assumptions, (7) embodies a number of useful ideas (including independence from units of nominal variables) which it seems reasonable to require of an econometric model's *solved equilibrium form.* However, (7) is a *demand schedule, not a behavioural plan*, and it is not sensible to attempt *direct* estimation of α and β. Indeed, attempting to do so for $M1$ yields (see Hendry, 1980):

$$(8) \qquad \ln M_t = \underset{(2\cdot 9)}{7\cdot 6} + \underset{(0\cdot 30)}{0\cdot 18} \ln Y_t + \underset{(0\cdot 17)}{0\cdot 84} \ln P_t - \underset{(0\cdot 02)}{0\cdot 12} \ln r_t + \underset{(0\cdot 76)}{0\cdot 17} \Delta \ln P_t$$

$$T = 32 \quad R^2 = 0\cdot 75 \quad S = 0\cdot 019 \quad d = 0\cdot 9 \quad \chi^2(20) = 399,$$

where T is the sample size. Such results are uninterpretable since d indicates significant autocorrelation (so that the quoted standard errors are badly downward-biased) and the model is rejected by the χ^2 (20) test for parameter constancy. The results hardly "corroborate" the "theory", so we do not seem to find a relationship where one was anticipated on grounds of "common sense" as much as "economic theory". Restricting the coefficients of Y and P to be unity increases S to $0\cdot 067$ and lowers d to $0\cdot 45$, so that "solution" can be rejected. Even neglecting the possibility that (8) is just another "spurious regression", it is not possible to decide whether or not the "theory" has been rejected since the model obviously does not adequately describe the *disequilibrium* data. Yet the dynamic equation eventually chosen as a reasonable model of the same data series had $S = 0\cdot 13$ and yielded the "equilibrium" solution:

$$(9) \qquad \ln (M/PY) = \ln K^* - \underset{(0\cdot 12)}{0\cdot 38} \ln r - \underset{(1\cdot 98)}{3\cdot 67} \ln (1 + \dot{p})$$

which is *consistent with the hypothesized demand schedule.* Moreover, the long-run homogeneity postulates could not be rejected, nor could parameter constancy (which also tested the weak exogeneity assumptions concerning P, Y and r) despite the obvious failure of (8).

My approach is admittedly *ad hoc*, since although "optimization" is a sensible organizing principle for economic theory, derived models will be *empirically* useful only if the associated criteria functions adequately represent agents' decision problems (that is, their objectives, costs and constraints). Present formulations are not entirely satisfactory. Consequently, my own empirical "research programme" has been to investigate modelling based on *minimal* assumptions about the intelligence of agents and the information available to them, with maximal reliance on data using "economic theory" guidelines to restrict the class of model considered, as in the $M1$ example. Agents form contingent plans, but respond like servomechanisms to changes in weakly exogenous variables (see for example Phillips, 1954). The resulting feedback models mimic "rational" behaviour for disequilibrium states around an otherwise constant steady-state growth path, and highlight features that seem worth incorporating in empirical time-series equations based on tighter theoretical specifications. The approach is complementary to both pure time-series analysis and theory-based quantitative economics, and has as its next stage the introduction of expectational and adaptive behaviour so that agents can learn to react rationally in non-steady-state worlds. Fortunately, others are also successfully tackling modelling from an economic theory viewpoint (see Nerlove, 1972) and, in particular, Muellbauer (1979) has derived interesting empirical equations from explicitly dynamic theories.

VI. Is Econometrics Alchemy or Science?

The ease with which spurious results could be created suggested alchemy, but the scientific status of econometrics was illustrated by showing that such deceptions are testable. In our rapidly changing world, undetected fallacies quickly become positive instances of Goodhart's "Law" (1978) to the effect that all econometric models break down when used for policy.

It is difficult to provide a convincing case for the defence against Keynes' accusation almost 40 years ago that econometrics is *statistical alchemy* since many of his criticisms remain apposite. The characterization of science offered earlier did not exclude econometrics *a priori* simply because of its inability to conduct controlled experiments. But empirical substantiation of the claim to be a science does require the existence of credible evidence, namely findings that are acceptable independently of political beliefs or preconceptions about the structural form of the economy (for a related critique from a systems theorist, see Kalman, 1979). The turbulence of the 1970s has greatly facilitated the rejection of "false" models, and although we are a long way from producing "answers", striking progress has been achieved since Keynes wrote, albeit at the cost of making the subject highly technical and increasingly inaccessible to non-specialists (for an interesting exposition, see Bray, 1979).

The alternative claim has been made by Hicks (1979, p. xi) that "as economics pushes on beyond 'statics' it becomes less like science and more like history". While this correctly highlights both the importance of the historical context and the fact that there is only one realization of any economic time series, it does *not* rule out a scientific approach to dynamic economics.

Econometricians may well tend to look too much where the light is and too little where the key might be found. Nevertheless, they are a positive help in trying to dispel the poor public image of economics (quantitative or otherwise) as a subject in which empty boxes are opened by assuming the existence of can-openers to reveal contents which any 10 economists will interpret in 11 ways.

Whether or not econometrics will prove to be more analogous to alchemy than to science depends primarily on the spirit with which the subject is tackled. Obviously, I cannot speak for how others will choose to use econometrics, although I believe that at this School we have attempted to tackle the subject scientifically. Hopefully, my examples may persuade you that such is at least potentially feasible. Far more rapid progress could be made if all empirical studies would provide greatly improved test information to allow readers to correctly judge plausibility. The three golden rules of econometrics are test, test and test;[6] that all three rules are broken regularly in empirical applications is fortunately easily remedied. Rigorously tested models, which adequately described the available data, encompassed prevous findings and were derived from well based theories would greatly enhance any claim to be scientific.

The study of what little econometric light we have is far from being an easy option, especially as taught at this School; nevertheless, there can be few more exciting or intellectually rewarding subjects and I commend its study to you.

ACKNOWLEDGMENTS

I am indebted to many colleagues for their help and advice in preparing this inaugural lecture but should like to thank in particular Mary Morgan, John Muellbauer, Frank Srba and Raija Thomson. The research was financed in part by Grant Number HR6727/1 from the Social Science Research Council to the Study in the History of Econometric Thought at the London School of Economics.

NOTES

[1] S and d respectively denote the equation standard error and the Durbin–Watson statistic. An estimated coefficient is conventionally called significant if the interval of plus and minus two standard errors does not include zero; in that case one can reject with approximately 95 per cent confidence the hypothesis that the coefficient is zero. Since both coefficients and their standard errors are estimated, and the numbers obtained depend on the method of estimation and the choice of model, "significance" of coefficients can change radically with the equation specification, as indeed occurs below.

[2] That is where the residual in one period is proportional to the residual one period previously plus a random variable; i.e., $\hat{u}_t = \lambda \hat{u}_{t-1} + \hat{e}_t$, where $\hat{u}_t$ is the tth residual.

[3] The two chi-squared values in Figure 8 are a (likelihood ratio) test for a common factor (χ_1^2) and a "Box–Pierce" test for residual auto-correlation (χ_{11}^2) respectively—see Sargan (1975), Mizon and Hendry (1980), Pierce (1971) and Breusch and Pagan (1980)—both of which "reject" the model specification.

[4] It may be useful to have a "microeconomic foundation" for macroeconometrics but it is not essential and may be counter-productive: "If it were necessary in the equations of hydrodynamics to specify the motion of every water molecule, a theory of ocean wave would be far beyond the means of 20th century science" (Wilson, 1979).

[5] Or, to quote Hicks (1979, p. xi): "decorated with least squares and confidence intervals".

[6] Notwithstanding the difficulties involved in calculating and controlling type I and II errors.

REFERENCES

BLALOCK, H. M. JR (1961). *Causal Inferences in Nonexperimental Research.* Chapel Hill: University of North Carolina Press.

BRAY, J. (1979). New models of the future. *New Statesman*, 18 May 1979, 710–714.

BREUSCH, T. S. and PAGAN, A. R. (1980). The Lagrange multiplier test and its applications to model specification. *Review of Economic Studies*, **47**, 239–253.

CHALMERS, A. F. (1976). *What is this Thing Called Science?* Queensland: University of Queensland Press.

DAVIDSON, J. E. H., HENDRY, D. F., SRBA, F. and YEO, S. (1978). Econometric modelling of the time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692.

DESAI, M. J. (1979). Testing monetarism: an econometric analysis of Professor Stein's model of monetarism. Unpublished paper, London School of Economics.

DURBIN, J. (1963). Maximum likelihood estimation of the parameters of a system of simultaneous regression equations. Paper presented to the Copenhagen Meeting of the Econometric Society.

EDWARDS, A. W. F. (1972). *Likelihood*. Cambridge: University Press.

ENGLE, R. F., HENDRY, D. F. and RICHARD, J.-F. (1979). Exogeneity. Unpublished paper, London School of Economics.

FRISCH, R. (1933). Editorial. *Econometrica*, **1**, 1–4.

GOODHART, C. A. E. (1978). Problems of monetary management: the UK experience. In *Inflation, Depression and Economic Policy in the West: Lessons from the 1970's*. (A. S. Courakis, ed.). Oxford: Basil Blackwell.

GRANGER, C. W. J. and NEWBOLD, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, **2**, 111–120.

GRILICHES, Z. (1974). Errors in variables and other unobservables. *Econometrica*, **42**, 979–1002.

HENDRY, D. F. (1976). The structure of simultaneous equations estimators. *Journal of Econometrics*, **4**, 51–88.

—— (1980). Predictive failure and econometric modelling in macro-economics: the transactions demand for money. *Economic Modelling*, (P. Ormerod, ed.), Chapter 9. London: Heinemann Educational Books.

—— and MIZON, G. E. (1978). Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England. *Economic Journal*, **88**, 549–563.

—— and VON UNGERN-STERNBERG, T. (1980). Liquidity and inflation effects on consumer's expenditure. In *Essays in the Theory and Measurement of Demand* (A. S. Deaton, ed.). Cambridge University Press.

HIBBERT, J. (1979). National and sectoral balance sheets in the United Kingdom. Paper presented to the Austrian Meeting of the International Association for Research in Income and Wealth, August 1979.

HICKMAN, B. G. (1972). *Econometric Models of Cyclical Behaviour*. New York: Columbia University Press.

HICKS, J. (1939). *Value and Capital: An Enquiry into Some Fundamental Principles of Economic Theory* (2nd ed. 1950). Oxford: Clarendon Press.

—— (1979). *Causality in Economics*. Oxford: Basil Blackwell.

JONSON, BEN (1612). *The Alchemist*. London: Thomas Snodham.

KALMAN, R. E. (1979). System theoretic critique of dynamic economic models. Unpublished paper, University of Florida, Gainsville.

KEYNES, G. (1946). Newton, the man. Paper read at the Newton Tercentenary Celebrations at Trinity College, Cambridge; in KEYNES, J. M. (1951), *Essays in Biography*. London: Rupert Hart-Davies.

KEYNES, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan.

—— (1939). Professor Tinbergen's method. *Economic Journal*, **49**, 558–568.

—— (1940). Comment, *Economic Journal*, **50**, 154–156.

KOOPMANS, T. C. (1947). Measurement without theory. *Review of Economics and Statistics*, **29**, 161–179.

—— (1949). A reply. *Review of Economics and Statistics*, **31**, 86–91.

—— (1950). When is an equation system complete for statistical purposes? *Statistical Inference in Dynamic Economic Models* (T. C. Koopmans, ed.), Chapter 17. New York: John Wiley.

—— (1957). *Three Essays on the State of Economic Science*. New York: McGraw-Hill.

—— (1979). Economics among the sciences. *American Economic Review*, **69**, 1–13.

KUHN, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University Press (1970, ed.).

LAKATOS, I. (1974). Falsification and the methodology of scientific research programmes. In I. Lakatos and A. E. Musgrave, *Criticism and the Growth of Knowledge*, pp. 91–196. Cambridge. University Press.

LEAMER, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: John Wiley.

LEONTIEF, W. (1971). Theoretical assumptions and nonobserved facts. *American Economic Review*, **61**, 1–7.

MASON, S. F. (1962). *A History of the Sciences.* New York: Collier Books.

MEEK, R. L. (ed.) (1973). *Turgot on Progress, Sociology and Economics.* Cambridge: University Press.

MIDDLETON, K. W. E. (1965). *A History of the Theories of Rain (and other forms of Precipitation).* London: Oldbourne.

MIZON, G. E. (1977). Model selection procedures. In *Studies in Modern Economic Analysis* (M. J. Artis and A. R. Nobay, eds). Oxford: Basil Blackwell.

MIZON, G. E. and HENDRY, D. F. (1980). An empirical application and Monte Carlo analysis of tests of dynamic specification. *Review of Economic Studies*, **47**, 21–46.

MORGENSTERN, O. (1950). *On the Accuracy of Economic Observations.* Princeton: University Press.

MUELLBAUER, J. (1979). Are employment decisions based on rational expectations? Unpublished paper, Birkbeck College.

NERLOVE, M. (1972). On lags in economic behaviour. *Econometrica*, **40**, 221–252.

ORCUTT, G. H. (1952). Toward a partial redirection of econometrics. *Review of Economics and Statistics*, **34**, 195–213.

PATINKIN, D. (1976). Keynes and econometrics: on the interaction between macroeconomic revolutions of the interwar period. *Econometrica*, **44**, 1091–1123.

PHELPS BROWN, E. H. (1972). The underdevelopment of economics. *Economic Journal*, **82**, 1–10.

PHILLIPS, A. W. (1954). Stabilisation policy in a closed economy. *Economic Journal*, **64**, 290–323.

—— (1956). Some notes on the estimation of time-forms of reactions in interdependent dynamic systems. *Economica*, **23**, 99–113.

PHILLIPS, P. C. B. (1977). Approximations to some finite sample distributions associated with a first order stochastic difference equations. *Econometrica*, **45**, 463–485.

PIERCE, D. A. (1971). Distribution of residual autocorrelations in the regression model with autoregressive-moving average errors. *Journal of the Royal Statistical Society*, B, **33**, 140–146.

POPPER, K. R. (1968). *The Logic of Scientific Discovery.* London: Hutchinson.

—— (1969). *Conjectures and Refutations.* London: Routledge & Kegan Paul.

RAO, C. R. (1965). *Linear Statistical Inference and Its Applications.* New York: John Wiley.

REID, D. J. (1977). Public sector debt. *Economic Trends*, May 1977, 100–107.

RICHARD, J.-F. (1980). Models with several regimes and changes in exogeneity. *Review of Economic Studies*, **47**, 1–20.

DE ROLA, S. K. (1973). *Alchemy: The Secret Art.* London: Thames & Hudson.

SARGAN, J. D. (1964). Wages and prices in the United Kingom: a study in econometric methodology. In *Econometric Analysis for National Economic Planning.* (P. E. Hart, G. Mills and J. K. Whitaker, eds). London: Butterworths.

—— (1975). A suggested technique for computing approximations to Wald criteria with application to testing dynamic specifications. Discussion paper A2, London School of Economics.

—— (1976). Econometric estimators and the Edgeworth expansion. *Econometrica*, **44**, 421–428.

SARGENT, T. J. and SIMS, C. A. (1977). Business cycle modelling without pretending to have too much a priori economic theory. In *New Methods in Business Cycle Research* (C. A. Sims, ed.). Federal Reserve Bank of Minneapolis.

SCHUMPETER, J. (1933). The common sense of econometrics. *Econometrica*, **1**, 5–12.

STONE, R. (1951). *The Role of Measurement in Economics.* Cambridge: University Press.

TAYLOR, C. T. and THREADGOLD, A. R. (1979). 'Real' national savings and its sectoral composition. Bank of England discussion paper no. 6, 1979.

TINBERGEN, J. (1939). *A Method and its Application to Investment Activity* (Statistical Testing of Business-Cycle Theories I). Geneva: League of Nations.

VINING, R. (1949a). Methodological Issues in Quantitative Economics. *Review of Economics and Statistics*, **31**, 77–86.

—— (1949b). A rejoinder. *Review of Economics and Statistics*, **31**, 91–94.

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, **54,** 426–482.

WILSON, K. G. (1979). Problems in physics with many scales of length. *Scientific American*, **241** (August), 140–157.

WOLD, H. O. (1969). Econometrics as pioneering in non-experimental model building. *Econometrica*, **37,** 369–381.

WORSWICK, G. D. N. (1972). Is progress in economic science possible? *Economic Journal*, **82,** 73–86.

YULE, G. U. (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.