COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 893

REFLECTIONS ON ECONOMETRIC METHODOLOGY

by

Peter C. B. Phillips

December 1988

REFLECTIONS ON ECONOMETRIC METHODOLOGY

by

P. C. B. Phillips*

Cowles Foundation for Research in Economics
Yale University

First Draft November 1988
Second Draft December 1988

*Helpful comments were made by Adrian Pagan on a preliminary version of this paper. My thanks also go to Glena Ames for her skill and effort in keyboarding the manuscript of this paper and to the NSF for research support under Grant No. SES 8519595.

0. ABSTRACT

General issues about the methodology of empirical econometric research are discussed. It is argued that the most successful paradigms for applied work are the ones that have a capacity to survive and to evolve into more useful forms as these are needed. Paradigms that embrace progressive modeling principles, such as those espoused by David Hendry, seem most amenable to this criterion. It is also argued that econometric theory has a large role to play in helping us to understand the strengths and the weaknesses of a methodology and to codify what its prescriptions entail. The time series methodology of David Hendry is considered in some detail. It is shown that the Hendry methodology comes remarkably close to achieving an optimal inference procedure for long run structural relationships even though it is conducted on a single equation basis. The findings indicate that the methodology may be improved further to achieve results that are equivalent to optimal estimation.

Detailed responses are provided to the panel discussions on econometric methodology by Dennis Aigner, Clive Granger, Edward Leamer and Hashem Pesaran that were presented at the 1988 Australasian Meetings of the Econometric Society in Canberra. Some personal reflections are offered on the many issues that arise from this panel discussion, including the merits of Bayesian and classical approaches, asymptotic theory, experimental and non experimental data, model evaluation, diagnostic testing, sharp prior hypotheses, the use of graphics, and the role of economic theory in empirical modeling.

1. SYMPTOMS AND PARADIGMS

It is a sign of the maturity of our profession that discourse on methodology has advanced from the armchair and the faculty bar, where it has been a common companion for many years, to colloquia at international conferences. Public debate about econometric methodology has now become a regular activity at conferences. But in the transition from private to public companion it can all too easily lose its intellectual vigour and excitement.

To a certain extent the transition is inevitable as each new generation of econometrician gains the confidence to indulge in public self evaluation. But we must be careful to ensure that this indulgence is not carried to excess, lest we become too far removed from the research work itself. Given the enormous growth in econometric research in recent years, I think there is little danger of this happening to our community at large. But there are other risks: to some of becoming enslaved by a particular methodology; and to others of turning away from empirical research altogether, disillusioned by the multitude of conflicting advice.

Empirical econometrics is a difficult, time consuming and little rewarded task. On the research frontier problems confront us from all sides. In empirical work this means problems of data, problems of economic theory and problems of econometric method. In the face of so many difficulties who is to say, *a priori*, what is the best way to proceed? And if there is no best way to do empirical econometric research what value can there be in public methodological debate?

These rhetorical questions will constitute a theme for these reflections. To me the history of econometrics helps to provide the answer. Put simply, the most successful paradigms are the ones that survive and multiply so that they are ready to evolve into new forms as they are needed. Our conferences and our journals already give good audience to our leading empiricists and over time their paradigms are assimilated or discarded in a natural process of evolution. Methodological debate can assist in this process by sharpening our understanding of the merits of different approaches and by early demonstration of their failures. Herein lie the real lessons and guidelines.

In evaluating the paradigms that are put before us we must examine their capacity to stay alive in a continuing confrontation with data. The concerns of macroeconomics, for example, mean that good methodologies of time series econometrics need to be attuned to the temporal evolution of the economic infrastructure. Such evolution necessarily undermines the rigid classifications of variables as exogenous or endogenous and the fixed parametric structures that characterise most macroeconometric models. One recent example of major significance is the emergence of the USA as a large open economy during the course of the 1980's. Another is the switch in the Federal Reserve Open Market Committee's operating target from the Federal funds rate to unborrowed reserves in October 1979. In the face of such changes in the big economic picture, methodologies must permit a flexibility of specification that accommodates new sources of economic feedback and endogeneity. Paradigms that embrace progressive modeling principles (see Lakatos (1978)) seem most amenable to these ideas. Prominent among these is the methodology of time series econometrics advocated by David Hendry (1980, 1983a, 1985, 1987), about which I shall have more to say later.

Related issues arise also in microeconometrics. Here it is the problem of heterogeneity that is endemic. As our data sets expand to include more economic agents or to track the behaviour of agents over space and time the new observations bring with them new sources of variability that themselves call for explanation. Yet our explanatory power in regressions with huge microeconometric data sets is frequently very low. Sometimes it is so low in the companion regressions that generate instruments as to question the identifiability of the underlying relations. Heterogeneity also causes problems in the use of nonparametric regression methods, which are of rising popularity in this field. In this case heterogeneity can make the regression results highly sensitive to bandwidth choice in finite samples. This introduces a new source of misspecification whereby the effect of heterogeneity is simply a distorted conditional regression function which is overworked in tracking the data through the choice of too small a bandwidth.

Choices of econometric method, model specification and data sets all come into play in the methodology of empirical research. For a paradigm to be useful it must guide the investigator in these choices, accommodate his various objectives, alert him to the type of pitfalls that occur and show him how to deal with them when they arise. Example is usually cited as the most powerful teacher in such matters. But setting a good example in econometric methodology is not enough. The most successful and enduring paradigms now seem likely to be those that offer complete implementation packages via econometric software that provides the range of computational and graphical facilities that are required for the conduct of the methodological prescriptions. Naturally there is danger in reducing a methodology to a formula by which input, output and diagnostic buttons are pushed by a series of rote commands. But software that offers a rich menu of choices with commentaries on usages and suggestive graphics is much more likely to elicit intelligent responses from an investigator who is seriously committed to learning what the data has to say about the hypotheses that interest him.

Finally, we must never underestimate the role that theory and simulation have to play in understanding methodological prescriptions. The role of simulation in this context has been creatively studied by David Hendry for many years (see Mizon and Hendry (1980), for a good example and Hendry (1983b) for a full discussion). Computer simulations give us enormous scope in learning how effective different modeling strategies can be in hunting out plausible data generating mechanisms. Likewise, good theoretical analysis helps us to understand what it is that certain methodological prescriptions achieve and others do not. In this way, theory can provide the key to generalisations that help the research paradigms to evolve. Without such metamorphosis empirical research methodologies cannot survive in the face of new economic theory, new data sets and an evolving economic infrastructure.

2. RESPONSES TO THE PANELISTS

The organisers of the 1988 Australasian Meetings of the Econometric Society convened a panel of econometricians to give their personal perspectives on the subject of methodology. Those perspectives should be seen in the context of the wider ongoing activity that I have referred to earlier. Since none of the panelists mention this activity let me provide a few references for readers who wish to pursue the subject further. Foremost among these are the contributions of Hendry (1987), Leamer (1987) and Sims (1987) to the methodology symposium at the Fifth World Congress of the Econometric Society in 1985. Pagan (1987) provides a helpful and pragmatic review of the three methodologies presented there. More recently, Poirier (1988) has advanced a Bayesian perspective on the subject of empirical model building. Poirier's paper and its attendant discussion point to some important pitfalls in both Bayesian and classical methodology. Readers are alerted to the limitations of both prescriptions and the debate steers empiricists away from an unthinking and mechanical imitation of either paradigm. But, like any abstract discussion, it does not and cannot provide a map for unknown territory. The good empiricist builds his own model and analyses his own data. As Hendry (1987) so aptly points out, the latter includes actually looking at the data. Good software packages provide the means for doing this in many different ways.

Having said this, let me take the views expressed by the panelists in the order of their presentation. I shall preface each discussion with my own précis of the panelists' main points. This will help to make the present paper self contained and to focus the arguments in my own discussion.

2.1. Dennis J. Aigner

Dennis Aigner appeals for better non experimental data, more use of experimental data, clearer logic in the search for causal laws and a healthy skepticism about the results of empirical research. It is hard to argue with his plea for better data. However, as I have indicated earlier, more data (even experimental data) does not necessarily lead to the type of improvements he has in mind. An example from astronomy, which is largely an observational science like economics, may help to elucidate this point. In understanding the structure of the heavens outside our immediate solar system it might be thought that it would be an enormous advantage to have in one's possession a small telescope rather than simply the naked eye. Certainly the amount of data increases by at least an order of magnitude (from about 5,000 stars to more than 100,000). Yet even this enormous increase in data is unequal to the task of disproving certain key beliefs of ancient astronomers e.g. that the stars are fixed on the inside of a celestial sphere or that the constellations are groups of connected stars. For to do so we must measure stellar distances and this requires (at least for nearby stars) triangulation using the earth's orbit as a base line and painstakingly accurate measurements of angles that are taken six months apart. Thus, to challenge the ancient theory of the constellations we don't need to see more stars we just need very precise measurements about a few of them. These precise measurements had to await the technological improvements in instrumentation of the nineteenth century.

In some fields of economics like finance, of course, there is no shortage of high quality data. With ongoing advances in computerised banking and financial activity it seems clear that there will be continuing improvements in the quality and quantity of much economic data. Sooner or later we will have seismographic-type monitoring of economic activity with continuous or near continuous time records. However, while such a wealth of data will help to tell us where the economy is and has been, by itself it will not help to untangle the complex of evolving interdependencies that characterise industrialised economies. For this purpose we will continue to need good theories, stylised models, statistical methods and computer simulations.

Before leaving the topic of data, let me mention that an interesting new type of experimental data is now with us in econometrics. This is experimental stock market data especially created for forecasting future events. Forsyth *et al.* (1988) have recently developed such a stock market for the presidential elections in the USA in 1988. Their market, which was open 24 hours a day for several months prior to the elections, had hundreds of participants whose trading decisions reflected informational updating on a continuous basis. The market prices in the Forsyth model may be regarded as efficiently embodying huge information sets about the individuals in the presidential race. They should therefore represent optimal predictors of the outcome given that information. Such predictors would seem to satisfy the pragmatic requirements of the Feigl definition of causality in terms of predictability that is discussed by Aigner.

Let me turn to the diagram (Figure 1) that is used in Aigner's discussion to encapsulate the process by which one gets "to the bottom line" in producing causal laws. Schematics like this are designed to chart the circumstances and flow of ideas and activities that give rise eventually to confirmation or disconfirmation of a theory by empirical evidence. Such charts are heavily used in engineering and in practical trade manuals where they are intended to nurse operatives to a level of competence in working with and maintaining a piece of machinery. In automobile repair they are called troubleshooting flow charts. Many of them are much more complicated than Figure 1. All of them suffer from the same serious deficiency of linear thinking. Attempts are sometimes made to compensate for this deficiency by having parallel flows from one level mix as they enter the next node. But linearity is pervasive and it seems naive, for example, to represent the process of model creation from data, theory and experience by a bubble and two arrows as Figure 1 does. We can take an alternative lesson in creative thinking from Jim Durbin, who remarked in a recent interview in *Econometric*

Theory (see Phillips (1988a)):

> Think about how your intuition works when you are solving problems. Anyone who thinks he's going to do it by writing down one logical step after another is usually a very dull chap. The clever people think all the way around a problem. They have a hierarchy of thoughts, they get insights and then only at a later stage try to convert it into a logical stream.

2.2. Clive W. J. Granger

The construction and use of empirical models is certainly one of the goals of econometric research. Clive Granger's comments focus on the problems of model transmission from the builders to the users of such models. He argues that this process of transference, if it is to be effective, involves not only the model itself but congeries of peripheral exercises which serve as a testimonial to the model's properties and capabilities. Good diagnostics, robustness and encompassing characteristics, demonstrated forecasting ability and acceptable internal properties such as balance all serve as evidence in this accompanying testimonial. If one accepts Granger's axiom, then the stronger the evidence in this testimonial then the more persuasive the empirical model will be to its potential users.

I suspect that many researchers would not accept Granger's distinction between builders and users. In fact, the people who have made the greatest contributions to empirical econometric modeling have been those who feel comfortable in dual roles. In microeconometric modeling one thinks naturally of people like Hausmann, Heckman, Griliches and McFadden. In macroeconometrics the list is much larger and includes such people as Bergstrom, Hendry, Klein, Pagan and Sims. There are others like Deaton and Goldberger who have established expertise in both areas. In many respects it is this group of real achievers in empirical econometric modeling who are the most qualified to speak out on issues of model transference and methodology. Fortunately, many have already done so. One issue that regularly comes up (as it does in Granger's discussion) is that of comparisons between competing models. In a recent interview in *Econometric Theory* (see Phillips (1988b)) Rex Bergstrom suggested that we compare models on the basis of their respective Gaussian likelihoods--as in some non nested testing procedures. Other suggestions include general variable addition methods (Pagan (1984)), encompassing tests (Hendry (1985), Mizon (1984)) and, of course, forecasting performance.

In comparing models it is useful to recognise each as an approximate mechanism. Like most investigators, Granger sees all models as approximations to a true data generating mechanism. Some theoretical work in time series such as that embodied in the recent book by Hannan and Deistler (1988) explicitly recognises this principle also. In other work researchers have gone further with this idea and questioned the discoverability of laws and the existence of "true data generating mechanisms" (see, for instance, Bewley (1988)). Such fundamentalist questioning rightly throws doubt on the value of many textbook modeling paradigms. But hard (as opposed to soft) statistical modeling will, in my view, retain its great strengths even in the face of such criticism. This is because of the huge body of knowledge that has been accumulated on the performance of statistical methods from real and simulated data sets. These findings put us in a good position to evaluate the merits of different procedures even if the underlying assumptions are only approximately satisfied.

2.3. Edward E. Leamer

Of all the panelists it is Leamer whose paper will most polarise opinion. In spite of its title, this paper raises concerns that are anything but superficial. Leamer questions what we purport to do in empirical research, how we do it and whether our objectives are noble in the sense that they concern genuine economic issues. The checklist of items that bother Leamer are worth looking at individually. I believe they are important enough to provide a good forum of discussion in any applied econometrics course. I shall look at them in the order in which they were given in his presentation.

(i) *Too few genuine issues*

If this is true then I think that we all bear some responsibility for it. As teachers, researchers, referees or editors part of our professional duty is to identify what we regard as the genuine issues. This does not mean that to be valuable every paper has to be a winner in the sense that it opens up big new issues. On the other hand, I strongly agree with Leamer that there should be some intellectual capital at risk. Staying on safe ground by protecting intellectual territory that has already been captured is a certain recipe for decline. Note that this criticism is as valid in theoretical areas as is it for empirical research. It is also true of disciplines outside of economics. The best way to counteract the inertia which keeps us on safe ground is to have healthy discussions at our conferences, journal outlets which promote informed commentaries and successive waves of good new people coming into the subject every year, all of which forces us to reevaluate established positions.

(ii) *Too many sharp hypotheses*

Bayesians frequently level this argument against classical statisticians. Yet I have never found it very convincing. For one thing Bayesian analysis itself makes extensive use of sharp hypotheses through the simple exclusion restrictions that are always implicit in setting up parametric models. Interestingly, classical *nonparametric* methods offer one way around such sharp specifications. These classical methods have undergone extensive recent development in statistics and econometrics. It is ironic in the face of Bayesian criticism about sharp hypotheses that there are no methods of comparable generality available in Bayesian methodology. This is because of the major problems that are involved in nonparametric Bayesian decision making. In effect, for high dimensional spaces the Bayesian approach swamps the data with prior information and in infinite dimensions it is generally inconsistent (see e.g. Diaconis and Freedman (1986)). These are serious objections and ones that should temper hasty criticism of classical statistical method.

(iii) *If the sample size is large you reject everything*

This is certainly not true. A major counterexample in time series is the unit root hypothesis. As I have earlier remarked, the accumulation of new data typically exposes new weaknesses in empirical models and as more data are brought to bear the model tends to become less relevant (unless it is repaired by respecification and reestimation to eliminate emerging flaws). In time series what this means is that the superposition of new shocks to the system inevitably leads to a stochastic drift away from the given model. This stochastic drift induces a unit root in the system. Thus, as the sample size increases we would naturally expect to see more evidence in favour of the unit root hypothesis. The very fact that constant term adjustments or structural shifts (such as those in Perron (1987)) are needed to eliminate this statistical evidence is itself support for the hypothesis because such adjustments themselves attach unit weight (and hence persistence) to certain innovations.

The unit root hypothesis is an example of a sharp prior that Bayesians often oppose (see the discussion under (ii) above). Yet to do so is to ignore the strong reasons given above for this hypothesis. Economic as well as statistical arguments apply in this case. In macroeconomics, for instance, it is possible to develop a theory according to which the aggregate economy can be regarded as an efficient market in the long run. What this means is that the engine of economic growth itself stems from the superposition of technological and demographic shocks over time. These shocks demonstrate temporal persistence-like the shocks that led to the discovery of the printing press, the renaissance, the industrial and scientific revolutions, refrigeration and computer technology. Macroeconomic models of this kind with persistent effects from innovations are the object of many recent studies, including those of Prescott (1986), for example. In financial economics, of course, efficient market theory has a long and well established tradition.

(iv) *Too many diagnostics*

This statement raises some interesting issues that have yet to be discussed in the literature. For one thing it is silly to indulge in diagnostic testing to the point where it becomes counterproductive in terms of data reduction. It is not all that unusual to see regressions reported in which the number of estimated coefficients, covariances and diagnostics come close to equaling the number of observations. Furthermore, it is very difficult for any investigator, no matter how experienced, to diagnose more than a few diagnostics simultaneously. One quickly reaches the point where it is more informative to use diagnostic graphs (see (v) below). Yet, on the other hand, it seems to me to be even more silly to ignore the useful evidence that good diagnostics can impart. Later on in his paper Leamer talks about empirical modeling as map reading and says:

> We can't sit in our offices and talk about what is a useful map. We need to get out there and use it.

Finding out what distortions occur in the map by detailed field work on the real territory is just what diagnostic testing is all about. How Leamer can repudiate it and at the same time embrace his map reading metaphor I do not know.

(v) *Too few graphs*

This is, in my opinion, a very good point. Well done graphics are an enormously powerful and expressive tool. Statisticians have been making similar points in recent years (see, for instance, Tufte (1983) and Cleveland (1984a)). This is certain to be an area where we shall see many enhancements in the near future. I, for one, would like to see much more use of graphics in residual diagnostic testing. A long time ago, Durbin (1969) argued for this when he suggested the cumulative periodogram in place of single test statistics for serial correlation.

On the other hand, graphics can be misused like any other statistical method. Cleveland (1984b) provides numerous examples from scientific publications. In econometrics one often sees graphics that impart very little information, like those that report the effects of innovation accounting from vector autoregressions. These must be used sparingly and in proportion to their real information content if they are to help empirical research.

(vi) *Too little confusion*

This point is easily misinterpreted. No researcher should be given a licence to confuse. If inferences are fragile then it is correct to say so. But by itself this is surely insufficient. All investigators bear a responsibility for presenting the clearest discussion and the fullmost interpretation of their empirical findings. It is simply unprofessional for an investigator to project all of his individual sources of confusion onto his audience. Mercifully, most journals inhibit the publication of uncontrolled spasms of confusion.

The form of confusion that Leamer has in mind is very specific. He characterises confusion as the inability to discriminate within an entire family of probability distributions, all of which adequately represent an investigator's prior views. The data may resolve this confusion by leading to sharp posteriors that are similar for all members of the prior family; or it may perpetuate or even accentuate the confusion, leading to excess sensitivity or fragility. This concept of Leamer-confusion, to give it a name, is related to the idea of Knightian uncertainty that has recently been explored by Bewley (1986, 1987, 1988). It might also be modeled through hierarchical priors. Either way, we end up with empirical measures of sensitivity. As Leamer argues, this is one of the strengths of the idea and I agree with him on this point. But I strongly disagree that other approaches fail in this respect and I shall explain why in my response to the next item on his list.

(vii) *Too much asymptotic theory*

It is a frequent refrain from Bayesians that what they do is exact in finite samples. Like Leamer they also frequently criticise the dependence of classical methods on asymptotics. Now I am surely one of the last people to be critical of the merits of finite sample theory (see, for example, my defence of the subject in Phillips (1982)). However, I also see great value in asymptotics and I believe Leamer's criticisms are misplaced. He argues as follows:

> In a small sample whatever method you use will work well sometimes and not work well with others and we need to know when. The asymptotic theory doesn't tell you when.

On the contrary, one of the major reasons for studying second order asymptotics is to glean this type of information. Remarkably, it is not even necessary for the second order asymptotics to work well in the sense of being good approximation for them to carry this information. A recent example of the use of second order asymptotics for this very purpose is Phillips and Park (1988). In effect, what Edgeworth corrections tell us is when the first order asymptotics are likely to do well (namely, when the second order corrections are small) and when they are likely to do poorly (when the corrections are large). The differences depend on parameters, including nuisance parameters.

Now it is interesting to observe the way in which different statistical procedures deal with nuisance parameters. In crude first order asymptotics, nuisance parameters are eliminated by the action of the asymptotics. For example, estimated covariance matrices converge to constants whose effects are usually eliminated in large samples by transformation (such as the use of a suitable metric in forming asymptotic chi-squared criteria). In

Bayesian methods nuisance parameters are flushed out of the system by integration. Monte Carlo numerical methods give Bayesians the tool they have needed to make their methods operational in this sense. Yet Bayesian methods are as constrained by this success as they are liberated by it. For the dependencies that are present in small samples are eliminated by Bayesian averaging, just as they are by the action of classical asymptotics. It is these very dependencies which produce the sensitivities that bother Leamer.

Indeed, model misspecifications are inevitably absorbed by the errors and show up in finite samples through nuisance parameter dependencies. Leamer proposes that we inject sensitivities back into the system through families or hierarchies of priors after Bayesian numerical integration methods have effectively already averaged them out. This is the height of irony. One of the great strengths of the classical paradigm is that it recognises many different sources of sensitivity and misspecification. The object of diagnostic testing, in particular, is to face up to this uncertainty over specification and to deal with the parameter dependencies that are induced by it--rather than flush them out of the system as in the Bayesian paradigm.

(viii) *The map making metaphor*

Leamer compares economic models to maps. He argues that models are tools for learning about economic activity just as maps are guides to the territory. Few would quarrel with this. Indeed, the map making metaphor is a common one at least to this point. But Leamer goes further and suggests that implicit in modeling is a usefulness function ($U(\varphi)$ in his notation) which measures how distortions of reality help the models to achieve their objectives. He gives the example of red lines on a map that signify roads and tells us that

> if I draw a red line on a map, the redness of that road can be falsified. Hence, that would be a map that would be rejected. So the traditional approach of statistical theory does not allow deliberate distortions.

Cartographers do use colour coding to help their maps carry more information and to make them easier to read. But I do not see this or similar artifices of cartography as the deliberate distortions Leamer has in mind. The map maker is just making the best use of his medium for the transmission of information. One may as well argue that you can falsify maps by noting that we do not drive on paper roads. To me this is no criticism of traditional statistical theory.

On the other hand, maps do simplify the territory. They omit information that the cartographers themselves do not know about or that they think will be of little use to people using the map. Simplifications also arise because of problems of scale: some information cannot be included without changing the scale; other information is selectively included or excluded according to the intended functions of the map. These innocent and deliberate omissions can lead to distortions and some of them will, indeed, figure in Leamer's usefulness function $U(\varphi)$. As a result, map reading requires some training and experience. Good readers study the legend, know the code and recognise that the territory itself changes. When a map doesn't fit their needs they look around for one that does.

Similar factors come into play in modeling. Ideally, one has a card file of models that, as Koopmans (1957) put it, "seek to express in simplified form different aspects of an always more complicated reality." In practice, of course, the prototype models are rarely enough. The hard work of empirical modeling lies as much in the crafting of a good model as it does in the skills of statistical inference. It is in this constructive process of crafting a good model that the decisions are made which determine Leamer's uncertainty parameter φ .

Leamer attacks traditional econometric modeling with the argument that maximum likelihood is "clearly not the right metric because it leaves out purposeful distortions." I do not accept this criticism as valid. For, in my view, neither the classical nor the Bayesian approaches are equal to the complete task of modeling. Good modelers (like Jim Durbin's problem solvers) think all the way around their subject before they start to build. They recognise the difference between uncertainty about the world and statistical variability. In the early stages of modeling the likelihood is not even a well defined concept. Neither the variable list, nor the data, nor the relevant economic theory are properly determined. Only as the research hypothesis and objectives emerge and become embodied in quantitative form do we begin to approach the outer limits where Bayesian and classical statistical paradigms come into service. It is in the antecedent thinking that the modeler shapes his purpose. Statistical paradigms are inevitably conditional on choices made in this antecedent stage. These choices rather than the statistical paradigms are the source of the distortions to which Leamer refers.

2.4. M. Hashem Pesaran

Pesaran emphasises the role of economic theory in applied econometrics. His message to us is that the good empiricist will study the background of economic theory that relates to the phenomena of interest then select and adapt theory so that it becomes more suited to the intended application. Out of this process comes a potentially powerful marriage of theory and econometric analysis that gives birth to the best applied research. This message is, of course, quite an old one. It is strongly evident in the writings and work of early researchers like Frisch (1932) and Stone (1951) and recently it has been forcefully restated elsewhere by Hendry and Wallis (1984) and by Pesaran (1986) himself.

It is easy to be sympathetic with the general idea behind this message. But differences soon start to appear at the point of practical implementation. Here an empiricist errs either by taking theory too seriously and putting in too many restrictions or by ignoring theory and following data oriented approaches. How much economic theory to use in applied work is a vexing question. In macroeconomics, theory usually provides little information about the process of short run adjustment. As a result the problem of dynamic specification is a ball that has for long been in the econometric end of the court out of default. Modern dynamic optimisation models offer an alternative that has been vigorously pursued by some. But the assumptions of representative agent behaviour that underlie these models are heroic and they shake the confidence of all but the most dedicated proponents of this approach. Moreover, problems of suitable dynamics for the error processes still surface in this approach and call for methodologies that can cope with the problems at a sufficient level of generality. In this respect it seems natural to use general dynamic models as a benchmark against which more specific theory driven models can be compared. Both the Hendry (1987) and Sims (1980) methodologies allow for this to a greater or lesser degree. In an illuminating personal commentary on this topic Pagan (1987) makes the observation:

> Over many years of looking at my own and my students' empirical studies I have found the rule of starting with a general model of fundamental importance for eventually drawing any conclusions about the nature of a relationship, and cannot imagine an econometric methodology that did not have this as its primary precept.

Pesaran has no advice to offer on these operationally vital issues of methodology. But he does express skepticism about both the atheoretical reduced form approach favoured by Sims and the error correction methodology advocated by Hendry. His view on the latter may seem surprising since the Hendry approach retains important elements of the "older style" of structural models whose strong theoretical content seems to accord with Pesaran's overall message. However, Pesaran argues that in the Hendry approach

> the economic theory is usually taken into account to the extent that it specifies some kind of equilibrium relationship. It is difficult to know where the equilibrium comes from...it is like a black box and it is difficult to identify the equilibrium conditions.

It is easy to give counterexamples that contradict Pesaran's stated view. Non trivial examples arise in steady state growth theory (where long run balance is an essential ingredient to the theory) in the permanent income theory of consumption, in present value models and in the purchasing power parity theory of international finance. None of these theories is a black box. Further examples occur in partial equilibrium theories of individual markets, such as those that underlie the formulations of the equations of the Bergstrom-Wymer (1976) model of the UK economy.

Pesaran seems to be calling for more extensive use of economic theory in empirical modeling. If long run equilibrium relationships are not enough for him then he must mean short run adjustment mechanisms as well. Yet the theories that give rise to these are subject to the difficulties mentioned earlier and in any event still need to be corroborated against more general dynamic specifications. It seems to me, therefore, that if one begins this line of argument it will lead almost inevitably to a methodology that most closely resembles that of Hendry. I shall offer some rather different evidence in favour of this conclusion in the following section.

3. THE ROLE OF ECONOMETRIC THEORY IN METHODOLOGY: SOME NEW SUPPORT FOR THE HENDRY APPROACH

There is one dimension in which the methodological debate of the last few years seems deficient to me. This is in terms of the use of econometric theory to understand the implications of the various methodological prescriptions. There are, of course, a few exceptions. Most prominent is the work of McAleer, Pagan and Volker (1986) and Pagan (1987) in explaining the extreme bounds analysis of Leamer (1978, 1983). Then there is the work of Sims, Stock and Watson (1986) and Park and Phillips (1988, 1989) in explaining asymptotics for vector autoregressions with integrated and cointegrated regressors. And perhaps my own work in Phillips (1986) explaining spurious regressions might also be included in this list. To complement this work I want now to put forward some ideas from theory which may help to understand the Hendry methodology.

Suppose we start with a cointegrated system that has no special trimmings (see my paper Phillips (1988c) for a full development) and which we write as

$$(1) \qquad y_{1t} = \beta' y_{2t} + u_{1t}$$

$$(2) \qquad \Delta y_{2t} = u_{2t}$$

where $y_t = (y_{1t}, y'_{2t})'$ is an (n+1)-vector I(1) process and $u_t = (u_{1t}, u'_{2t})'$ is stationary. This system has an error correction model representation of the form

$$(3) \qquad \Delta y_t = \gamma\alpha' y_{t-1} + v_t$$

with

$$\gamma' = (-1, 0) \ , \quad \alpha' = (1, -\beta')$$

and

$$v_t = \begin{bmatrix} 1 & \beta' \\ 0 & I \end{bmatrix} u_t \ .$$

Now (3) is a triangular system with contemporaneously and serially correlated errors. As shown in Phillips (1988c, 1988d) optimal inference of the (long run) cointegrating vector α in (3) can be achieved by maximum likelihood or by an asymptotically equivalent spectral regression. In both cases this involves systems estimation and would therefore appear to be quite a different prescription from the single equation error correction methodology espoused in Hendry's research and explained in detail in Hendry and Richard (1982, 1983). However, it turns out that the two prescriptions are intimately related. This is because the triangular structure of (3) means that systems estimation gives, in effect, a time series version of generalised least squares and can be shown to have a single equation analogue that is identical in form to a typical regression in the Hendry methodology. As we shall see, in some cases this ensures that the Hendry methodology will indeed lead to asymptotically optimal inferences. To explain this conclusion it is helpful to proceed by example.

EXAMPLE 1. We shall consider first the case where v_t is iid $N(0,\Sigma)$ and Σ is partitioned conformably with u_t as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22} \end{bmatrix} \quad \text{and} \quad v_t = \begin{bmatrix} v_{1t} \\ u_{2t} \end{bmatrix} .$$

Then maximum likelihood on (3) is just single equation least squares on (see Remark (h) of Phillips (1988c))

$$\Delta y_{1t} = \gamma_1 \alpha' y_{t-1} + \sigma_{21}' \Sigma_{22}^{-1} \Delta y_{2t} + v_{1\cdot 2t} \tag{4}$$

where

$$v_{1\cdot 2t} = v_{1t} - \sigma_{21}' \Sigma_{22}^{-1} u_{2t}$$

$$\gamma_1 = -1 .$$

Observe that (4) has both levels and differences on the right hand side and is a highly simplified version of a typical empirical equation that arises in the Hendry methodology. Indeed, it is identical in form to the stylised example used by Pagan (1987) (see his equation (2b)) in his description of the Hendry approach. However, as I shall demonstrate, the link between the Hendry approach and optimal estimation theory runs much deeper than this example. □

In the time series case u_t and, hence, v_t in (3) are quite general stationary processes. They may be regarded as embodying all of the short run dynamics of the adjustment mechanism. To model these processes explicitly calls for some model selection procedure. In the Hendry methodology this involves working back from a general unrestricted dynamic specification towards a parsimoniously reparameterised model whose regressors are temporal transformations that are interpretable in some economic sense and nearly orthogonal. This is, of course, done on a single equation basis. Again, Pagan (1987) provides a good discussion of the process. It will be useful to us to bear in mind that the objective of the Hendry approach is to seek out a single equation model that is a tentatively adequate, conditional data characterization. Such a model satisfies the following criteria which we shall call the Hendry-Richard prescriptions, based on the recommendations in Hendry and Richard (1982, p. 21; 1983, p. 140):

(a) The model is coherent with the data (i.e. fits the data up to an innovation that is white noise and, further, a martingale difference sequence relative to the selected data base);

(b) It validly conditions on variables that are weakly exogenous with respect to the parameters of interest;

(c) It encompasses rival models;

(d) Its formulation is consistent with economic theory;

(e) It has parsimoniously chosen and orthogonal decision variables.

In the time series context my own optimal inference approach can also be reduced to a single equation method. This has been done in Phillips and Hansen (1988). The essential idea is to use a semiparametric correction which fully modifies conventional least squares and its attendant standard errors for the effects of simultaneity and serial correlation. We have designed the corrections so that they may be employed directly in a regression in levels (rather than differences) such as the cointegrating equation (1). But the effects are the same if we work in error correction format. I shall do the comparison below in levels since this is easier to follow.

Working from the single equation (1) Phillips and Hansen (1988) give a fully modified least squares estimator of β that takes the form

$$\beta^{+} = (Y_2'Y_2)^{-1}(Y_2'y_1^{+} - \hat{\delta}^{+}) \tag{5}$$

where Y_2 $(T \times n)$ is the observation matrix for y_{2t}, y_1^{+} $(T \times 1)$ is the observation vector for

$$y_{1t}^{+} = y_{1t} - \hat{\omega}_{12}'\hat{\Omega}_{22}^{-1}\Delta y_{2t} \tag{6}$$

and

$$\hat{\delta}^{+} = \hat{\Delta}\begin{bmatrix} 1 \\ -\hat{\Omega}_{22}^{-1}\hat{\omega}_{21} \end{bmatrix} \tag{7}$$

is a bias correction term with $\hat{\Delta}$ a consistent estimate of

$$\Delta = \Sigma_{k=0}^{\infty}E(u_{20}u_k') \ . \tag{8}$$

In (6) and (7), $\hat{\omega}_{21}$ and $\hat{\Omega}_{22}$ are consistent estimates of the corresponding elements in the long run covariance matrix

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix} = 2\pi f_{uu}(0)$$

where $f_{uu}(\lambda)$ is the spectral density matrix of u_t . Under rather general conditions

(9) $$T(\beta^{+} - \beta) \Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1} \int_0^1 B_2 dB_{1\cdot 2} = \int_{G>0} N(0, \omega_{11\cdot 2} G^{-1}) dP(G)$$

where

$$\begin{bmatrix} B_{1\cdot 2}(r) \\ B_2(r) \end{bmatrix} = BM(\Omega) \ , \quad \Omega = \begin{bmatrix} \omega_{11\cdot 2} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix}$$

(i.e. Brownian motion with covariance matrix Ω) and

$$\omega_{11\cdot 2} = \omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21} \ .$$

The limit distribution given by (9) is a covariance matrix mixture of normals. The mixing variate is $G = \int_0^1 B_2 B_2'$, which is a quadratic functional of the vector $BM(\Omega_{22})$.

Fully modified standard errors for individual elements $\hat{\beta}_i$ of $\hat{\beta}$ are given by s_i^{+} where

$$s_i^{+2} = \hat{\omega}_{11\cdot 2}[(Y_2'Y_2)^{-1}]_{ii}$$

and

(10) $$\hat{\omega}_{11\cdot 2} = \hat{\omega}_{11} - \hat{\omega}_{21}' \hat{\Omega}_{22}^{-1} \hat{\omega}_{21} \ .$$

With these standard errors, we get t ratios for which conventional asymptotic theory applies. In particular, as the sample size $T \to \infty$

(11) $$t_i^{+} = (\beta_i^{+} - \beta_i)/s_i^{+} \Rightarrow N(0,1) \ .$$

As discussed in Phillips and Hansen (1988) this approach is asymptotically equivalent to optimal systems procedures like full maximum likelihood. It involves corrections for endogeneity (through the use of y^+_{1t} in place of y_{1t}) and serial correlation (through the bias correction term δ^+ that occurs in the numerator of (5)). These modifications correct the conditional mean of u_{1t} in (1) for the long run simultaneity in the system that is due to cointegration; and they also correct the estimates for the serial covariance properties of the error u_t , particularly the temporal covariances between u_{1t} and u_{2t} . Note also that when these corrections are made it becomes necessary to use the conditional variance estimate $\hat{\omega}_{11\cdot 2}$ as in (10) and (11) for inferential purposes. Thus, ordinary least squares and its attendant standard errors can be modified with these corrections to account for the long run effects of endogeneity and serial dependence. The resulting regression equation that is based on $(\beta^+_i, s^+_i, t^+_i)$ is ready for inference with conventional (asymptotic) procedures.

It turns out some of these corrections are also implicitly performed in the Hendry methodology. In this methodology the starting point is a general unrestricted regression of the form

$$(12) \qquad y_{1t} = \hat{\beta} y_{2t} + \hat{\gamma} x_t + \hat{w}_t \ .$$

Again we shall work in levels for convenience, although the same analysis could be performed for error correction format. In (12) x_t is an autoregressive distributed lag vector whose components are lagged values of the regressors and the dependent variable and possibly other explanatory variables. Although it is not explicit in the Hendry-Richard prescriptions we shall assume that k , the number of variables in x_t, grows with T but in such a way that $k/T \to 0$ as $T \to \infty$. It can also be assumed that (12) satisfies the data coherency criterion (a). In the second stage of the methodology, x_t is parsimoniously orthogonalised by temporal transformation (criterion (e)). This leads to a representation in which the components of the new vector x_t are stationary and no longer asymptotically collinear (or cointegrated). Typically, x_t involves differences, higher order differences and lagged differences of the variables in the system. The parsimonious reduction process is not fully explained and perhaps for this reason poses real difficulties for novices in the methodology. However, it seems reasonable to suppose that only those variables whose coefficients are significant according to conventional asymptotic tests are retained.

For our purposes we can treat x_t as stationary and assume that it contains lagged values of Δy_{1t} and present and lagged values of Δy_{2t}. We then have in conventional regression notation

$$(13) \qquad \hat{\beta} = (Y_2'Q_xY_2)^{-1}(Y_2'Q_xy_1)$$

and thus

$$(14) \qquad T(\hat{\beta}-\beta) = (T^{-2}Y_2'Q_xY_2)(T^{-1}Y_2'Q_xu_1) \ .$$

Since x_t is stationary

$$(15) \qquad T^{-2}Y_2'Q_xY_2 - T^{-2}Y_2'Y_2 \underset{p}{\to} 0$$

so that we have the asymptotic equivalence

$$T(\hat{\beta}-\beta) = (T^{-2}Y_2'Y_2)^{-1}(T^{-1}Y_2'Q_x u_1) .$$

$Q_x u_1$ projects u_{1t} onto the orthogonal complement of the span of $\{x_t : t = 1, \ldots, T\}$. Asymptotically, this projection works like a conditional expectation adjusting the mean of u_{1t} for the effects of the past history of u_{1t} and the present and past history of u_{2t} . This leads to the process

$$\eta_t = u_{1t} - E(u_{1t}|F_{t-1})$$

where $F_{t-1} = \sigma(u_{1t-1}, u_{1t-2}, \ldots; u_{2t}, u_{2t-1}, \ldots)$ of the Hendry-Richard prescriptions. Note that η_t is a martingale difference sequence and therefore satisfies criterion (a) of the Hendry-Richard prescriptions.

In particular, if u_t is the linear process

$$(16) \qquad u_t = \Sigma_{j=0}^{\infty} A_j \epsilon_{t-j} \ ; \quad A_0 = I \ , \quad \Sigma_{j=0}^{\infty} \|A_j\| < \infty \ , \quad (\epsilon_t) \equiv \text{iid } N(0,\Sigma)$$

then

$$(17) \qquad \eta_t = \epsilon_{1t} - E(\epsilon_{1t}|\epsilon_{2t}) = \epsilon_{1t} - \sigma_{21}'\Sigma_{22}^{-1}\epsilon_{2t}$$

where Σ is partitioned conformably with Ω . The variance of η_t is $\sigma_{11\cdot 2} = \sigma_{11} - \sigma_{21}'\Sigma_{22}^{-1}\sigma_{21}$ and η_t is orthogonal to ϵ_{2t} as well as the entire past history $(\epsilon_{t-1}, \epsilon_{t-2}, \ldots)$.

The Hendry approach, which starts with an unrestricted regression like (12), is asymptotically equivalent to maximising the conditional likelihood of $(\epsilon_{11}, \ldots, \epsilon_{1T})$ given $\{\epsilon_{2t} : t = 1, \ldots, T\}$ i.e.

$$(18) \qquad -(T/2)\ell n\ \sigma_{11\cdot 2} - (1/2\ \sigma_{11\cdot 2})\Sigma_1^T\left(\epsilon_{1t} - \sigma_{21}'\Sigma_{22}^{-1}\epsilon_{2t}\right)^2 .$$

Assume that (16) can be inverted and written in autoregressive form as

$$(19) \qquad B(L)u_t = \epsilon_t , \quad B(L) = \Sigma_{j=0}^{\infty}B_jL^j , \quad B_0 = I .$$

Partitioning B conformably, we see that maximising (18) is equivalent to minimising

$$\Sigma_1^T\left\{(b_{11}(L),\ b_{12}(L))u_t - \sigma_{21}'\Sigma_{22}^{-1}(b_{21}(L),\ B_{22}(L))u_t\right\}^2 .$$

This is the same as running least squares on the equation

$$(20) \qquad y_{1t} = \beta_{2t}' + d_1(L)(y_{1t} - \beta' y_{2t}) + d_2(L)\Delta y_{2t} + \eta_t$$

where $d_i(L)$ are lag polynomials of infinite order, in general $\left(d_1(L) = \Sigma_{j=1}^{\infty}d_{1j}L^j,\ d_2(L) = \Sigma_{j=1}^{\infty}d_{2j}L^j\right)$. The error on (20) is iid $N(0,\ \sigma_{11\cdot 2})$ and is independent of the regressors. As we have seen least squares on (20) is also equivalent to conditional maximum likelihood. These considerations suggest that equation (20) should form a basis for optimal inference. In other words it would appear that (20) achieves by direct inclusion of regressors what the Phillips-Hansen procedure obtains through its semi-parametric corrections. If this were true precisely then it would provide very powerful arguments in favour of the Hendry methodology. In fact, single equation estimation of (20) falls short of attaining optimal inference in general.

We shall not develop a fully general theory here, but it can be shown that if $\hat{\beta}$ is the least squares coefficient from a regression on (20) then

$$(21) \qquad T(\hat{\beta}-\beta) \Rightarrow \left[\int_0^1 B_2 B_2'\right]^{-1} \left(\int_0^1 B_2 dB_\eta\right) .$$

This is close to (9) but not equivalent to it. The difference lies in the Brownian motion $B_\eta(r)$ (based on the residual from (20)) which can, in general, be correlated with the Brownian motion $B_2(r)$ (based on u_{2t}). This is because, although η_t is orthogonal to y_{2t} and the past history of u_{2t} , the process u_{2t} is not necessarily orthogonal to the past history of u_{1t} and, hence η_t . In other words, there is a failure of weak exogeneity or valid conditioning in (20). The approach therefore falls short of attaining objective (b) in the Hendry-Richard prescriptions. We shall give two examples to illustrate what can happen. In Example 2 u_{2t} and η_t are *strictly* exogenous. In Example 3 they are not.

EXAMPLE 2. Let (19) take the special form:

$$u_{1t} = \epsilon_{1t}$$

$$u_{2t} = Bu_{2t-1} + \epsilon_{2t}$$

Then

$$\eta_t = u_{1t} - E(u_{1t}|F_{t-1}) = u_{1t} - \{\sigma_{21}'\Sigma_{22}^{-1}u_{2t} - \sigma_{21}'\Sigma_{22}^{-1}Bu_{2t-1}\}$$

$$= \epsilon_{1t} - \sigma_{21}'\Sigma_{22}^{-1}\epsilon_{2t}$$

and the spectral density matrix of $\zeta_t = (\eta_t, u_{2t}')'$ is

$$f_{\zeta\zeta}(\lambda) = \frac{1}{2\pi}\begin{bmatrix} 1 & -\sigma_{21}'\Sigma_{22}^{-1} \\ 0 & \left(I - Be^{i\lambda}\right)^{-1} \end{bmatrix} \Sigma \begin{bmatrix} 1 & 0 \\ -\Sigma_{22}^{-1}\sigma_{21} & \left(I - B'e^{i\lambda}\right)^{-1} \end{bmatrix}$$

$$= (1/2\pi)\begin{bmatrix} \sigma_{11\cdot 2} & \\ 0 & \left(I - Be^{i\lambda}\right)^{-1}\Sigma_{22}\left(I - B'e^{-i\lambda}\right)^{-1} \end{bmatrix} .$$

It follows that η_t and u_{2t} are incoherent (i.e. uncorrelated at all frequencies) and the limit Brownian motions are

$$\begin{bmatrix} B_\eta(r) \\ B_2(r) \end{bmatrix} = BM(2\pi f_{\zeta\zeta}(0))$$

where B_η and B_2 are independent. In fact, $B_\eta = B_{1\cdot 2}$ and the limit distribution (2) is identical to that of the Phillips-Hansen estimator given by (9).

In this example we can write equation (20) directly as

$$y_{1t} = \beta' y_{2t} + \sigma_{21}'\Sigma_{22}^{-1}\Delta y_{2t} - \sigma_{21}'\Sigma_{22}^{-1}B\Delta y_{2t-1} + \eta_t .$$

In error correction format this equation takes the form

$$\Delta y_{1t} = \gamma_1(y_{1t-1} - \beta' y_{2t-1}) + \sigma_{21}'\Sigma_{22}^{-1}\Delta y_{2t} - \sigma_{21}'\Sigma_{22}^{-1}B\Delta y_{2t-1} + \eta_t \tag{22}$$

thereby extending (4) to the case where u_{2t} is a vector autoregressive error.

It is worth examining the form of the Phillips-Hansen estimator β^+ in this case. First observe that

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{21}'(I-B')^{-1} \\ (I-B)^{-1}\sigma_{21} & (I-B)^{-1}\Sigma_{22}(I-B')^{-1} \end{bmatrix}$$

and thus

$$y_{1t}^{+} = y_{1t} - \sigma_{21}'\Sigma_{22}^{-1}(I-B)\Delta y_{2t} \ . \tag{23}$$

Moreover

$$\Delta = \Sigma_{k=0}^{\infty} u_{20}u_k' = [\sigma_{21},\ \Sigma_{22}(I-B')^{-1}]$$

so that

$$\delta^{+} = \Delta\begin{bmatrix} 1 \\ -(I-B')\Sigma_{22}^{-1}\sigma_{21} \end{bmatrix} = 0$$

and the true bias correction is zero in this case.

Note that the correction in (23) involves one term--the long run endogeneity correction

$$\omega_{21}'\Omega_{22}^{-1}\Delta y_{2t} = \sigma_{21}'\Sigma_{22}^{-1}(I-B)\Delta y_{2t} \ . \tag{24}$$

This correction is achieved in (22) through the presence of the two regressors Δy_{2t} and Δy_{2t-1}. The additional economy in (24) is obtained because a consistent semi-parametric estimate of the coefficient $\omega_{21}'\Omega_{22}^{-1}$ is employed. In (22) this is estimated componentwise through the coefficients of the two regressors. □

EXAMPLE 3. Let (16) take the special MA(1) form

$$u_t = \varepsilon_t + \Theta\varepsilon_{t-1}$$

and partition Θ conformably as

$$\Theta = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \Theta_{22} \end{bmatrix} .$$

Then

$$\eta_t = u_{1t} - E(u_{1t}|F_{t-1}) = \varepsilon_{1t} - \sigma_{21}'\Sigma_{22}^{-1}\varepsilon_{2t} \tag{25}$$

and

$$u_{2t} = \varepsilon_{2t} + \theta_{21}\varepsilon_{1t-1} + \Theta_{22}\varepsilon_{2t-1} . \tag{26}$$

The process $\zeta_t = (\eta_t, u_{2t}')'$ now has spectral density matrix

$$f_{\zeta\zeta}(\lambda) = \frac{1}{2\pi}\begin{bmatrix} 1 & -\sigma_{21}'\Sigma_{22}^{-1} \\ \theta_{21}e^{i\lambda} & I + \Theta_{22}e^{i\lambda} \end{bmatrix}\Sigma\begin{bmatrix} 1 & \Theta_{21}'e^{-i\lambda} \\ -\Sigma_{22}^{-1}\sigma_{21} & I + \Theta_{22}'e^{-i\lambda} \end{bmatrix} .$$

Observe that

$$2\pi f_{\zeta\zeta}(0) = \begin{bmatrix} \sigma_{11\cdot2} & \sigma_{11\cdot2}\theta_{21}' \\ \sigma_{11\cdot2}\theta_{21} & H \end{bmatrix} .$$

where

$$H = \sigma_{11}\theta_{21}\theta_{21}' + (I + \Theta_{22})\sigma_{21}\theta_{21}' + \theta_{21}\sigma_{21}'(I + \Theta_{22}')$$
$$+ (I + \Theta_{22})\Sigma_{22}(I + \Theta_{22}') .$$

The Brownian motions B_η and B_2 are therefore correlated unless $\theta_{21} = 0$. In fact, from Lemma 3.1 of Phillips (1989) we may decompose B_η as

$$B_\eta(r) = \sigma_{11\cdot 2}\theta_{21}'H^{-1}B_2(r) + B_{\eta\cdot 2}(r)$$

where

$$B_{\eta\cdot 2}(r) = BM(\sigma_{11\cdot 2} - \sigma_{11\cdot 2}^2\theta_{21}'H^{-1}\theta_{21})$$

and is independent of $B_2(r)$. With this decomposition we obtain

$$T(\hat{\beta}-\beta) = \left(\int_0^1 B_2B_2'\right)^{-1}\int_0^1 B_2dB_2'H^{-1}\theta_{21}\sigma_{11\cdot 2} + \left(\int_0^1 B_2B_2'\right)^{-1}\int_0^1 B_2dB_{\eta\cdot 2} \ . \tag{27}$$

The first term in (27) is a matrix form of unit root distribution. It imports a bias and an asymmetry to the limit distribution when $\theta_{21} \neq 0$. The second term is a mixture of normals and is equivalent to (9) when $\theta_{21} = 0$.

It is interesting to examine why the error correction methodology fails to produce an asymptotically optimal inferential procedure in this case. The reason is that the error vector u_{2t} depends on the past history of the innovations ϵ_{1t} that determine $u_{1t} = y_{1t} - \beta' y_{2t}$. This obvious but important fact has been noted before, of course, in moving average models with stationary time series (e.g. Hall and Pagan (1981), Hansen and Hodrick (1980)). In the present example, the term $\theta_{21}\epsilon_{1t-1}$ occurs in (26). Optimal inference of β requires that the generating mechanism for u_{2t} be estimated jointly with the single equation error correction model. In effect, the innovation sequence $\{\epsilon_{1t}, \epsilon_{1t-1}, \ldots\}$ jointly determines both

u_{1t} and u_{2t}. The error correction methodology successfully reduces the stationary errors on the cointegrating equation (1) to the orthogonal sequence η_t. And η_t is a martingale difference sequence with respect to the event field generated by $(\epsilon_{1t-1}, \epsilon_{1t-2}, \ldots; \epsilon_{2t}, \epsilon_{2t-1}, \ldots)$. In *stationary* regression models this would be sufficient to ensure asymptotically median unbiased inferences. However, there is an information loss in this approach to the extent that the generating mechanism for u_{2t} is not fully estimated. Note that β is the coefficient of y_{2t} in (1) and $y_{2t} = \Sigma_1^t u_{2j}$ accumulates the errors u_{2t}. When there is an information loss in the estimation of the data generating mechanism, this loss therefore has an accumulative effect which results in the asymptotic correlation between the Brownian motions $B_2(r)$ (arising from u_{2t}) and $B_\eta(r)$ (arising from η_t). Only in cases where the full information content of the errors is purged from η_t will the resulting process B_η be independent of B_2. Systems estimation as in Phillips (1988c) or fully modified least squares estimates as in Phillips and Hansen (1988) achieve this. Single equation error correction methodology generally does not. Note that the dependence between B_2 and B_η that is induced by the (unmodified) single equation approach results not only in inefficiency (as would naturally be expected in stationary regression models) but also an asymptotic bias--see Phillips (1988c) for a full discussion. □

It should be clear from the above examples that the Hendry methodology comes remarkably close to achieving an optimal inference procedure. In some cases it actually does so and in the other cases it can be further modified to achieve it, as the Phillips-Hansen corrections indicate. These findings can be taken to provide strong support for the Hendry approach. They also suggest ways in which it may be developed to achieve improved results. This is surely one of the strengths of analytic research.

4. PROGNOSES

In an essay to his grandchildren Keynes (1931) suggested that economists may eventually play a role in society like that of dentists. Aigner in his presentation tells us that if we cannot truck it as econometricians we can become plumbers. He cites plumbing, whose history I may add is a good deal longer than economics, as a trade where troubleshooting flow charts actually work. Leamer in his presentation tells us "to give me parameters or give me death. Death means to do econometric theory but not analyse data sets." Pesaran advises us "to become a better economist than a better philosopher." Comparisons such as these are not uncommon in public self evaluations where scientists seek to speak out about their subject in general terms. But, frankly, I do not find them to be very helpful.

Keynes' prognosis for economics has not yet come to pass. This is in large part due to the huge credibility gap that exists between economic theory, empirical evidence and policy prescriptions. The fact is that government, industry and the general public have more faith in dentists and plumbers than they do in economists and econometricians. Professional respect must be earnt through long and meritorious service in which policy prescriptions are seen to work. In this we are still paying our dues.

Leamer is right to stress that we must analyse data. But he is wrong to malign econometric theory. Sometimes, I think that a lot of what we do in applied econometrics is really just descriptive statistics that we dignify by the name of inference. Yet, I believe we must be very forgiving of empirical research and give it active professional encouragement. Not to do so increases the risks of certain methodological prescriptions emerging as dogma, thereby inhibiting their capacity to survive and evolve into more useful forms.

We must also recognise the role that econometric theory has to play in guiding the hand of empirical research. As I have attempted to show in Section 3, theory can help us to understand the strengths of a methodology and to codify what its prescriptions entail. It can also point the direction for further development. As such, theory and applied research can be powerful allies in the right hands. Nevertheless, even the most learned among us have a lot more to master if econometrics is ever to be as successful as plumbing or dentistry. If this is a humbling thought then it should also be said that econometrics presents a more complex and engrossing challenge. Perhaps it is silly even to contemplate these comparisons. The simplest prognosis is then the most fitting. Econometrics will become what we turn the subject into, just as at the moment it reflects no more and no less than what we are doing today.

REFERENCES

Bergstrom, A. R. and C. R. Wymer (1976), "A model of disequilibrium neoclassical growth and its application to the United Kingdom," in A. R. Bergstrom (ed.), *Statistical Inference in Continuous Time Economic Models*. Amsterdam: North Holland.

Bewley, T. F. (1986), "Knightian decision theory: Part 1," Cowles Foundation Discussion Paper #807, Yale University.

_________ (1987), "Knightian decision theory: Part II," Cowles Foundation Discussion Paper #835, Yale University.

_________ (1988), "Knightian decision theory and econometric inference," Cowles Foundation Discussion Paper #868, Yale University.

Cleveland, W. S. (1984a), "Graphical methods for data presentation: Full scale breaks, dot charts and multibased logging," *The American Statistician*, 38, 270-280.

_________ (1984b), "Graphs in scientific publications," *The American Statistician*, 38, 261-269.

Diaconis, P. and D. Freedman (1986), "On the consistency of Bayes estimates," *Annals of Statistics*, 14, 1-26.

Durbin, J. D. (1969), "Tests for serial correlation in regression analysis based on the periodogram of least squares residuals," *Biometrika*, 56, 1-15.

Forsyth, R., F. Nelson, G. Neumann and J. Wright (1988), "The Iowa presidential stock market: A field experiment" (mimeo) Iowa University, forthcoming in *Research in Experimental Economics*, 4 (1989).

Frisch, R. (1932), "Editorial," *Econometrica*, 1, 1-4.

Hall, A. D. and A. R. Pagan (1981), "The LIML and related estimators of an equation with moving average disturbances," *International Economic Review*, 22, 719-730.

Hannan, E. J. and M. Deistler (1988), *The Statistical Theory of Linear Systems*. New York: Wiley.

Hansen, L. P. and R. J. Hodrick (1980), "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis," *Journal of Political Economy*, 8, 829-853.

Hendry, D. F. (1980), "Econometrics--alchemy or science," *Economica*, 47, 387-406.

_________ (1983), "Econometric modeling: The consumption function in retrospect," *The Scottish Journal of Political Economy*, 30, 193-220.

_________ (1984), "Monte Carlo experimentation in econometrics," in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2. Amsterdam: North Holland.

_________ (1985), "Monetary economic myth and econometric reality," *Oxford Review of Economic Policy*, 1, 72-84.

_________ (1987), "Econometric methodology: A personal perspective," in T Bewley (ed.), *Advances in Econometrics Vol. 2*. Cambridge: Cambridge University Press.

Hendry, D. F. and J-F. Richard (1982), "On the formulation of empirical models in dynamic econometrics," *Journal of Econometrics*, 20, 3-33.

_________ (1983), "The econometric analysis of economic time series" (with discussion), *International Statistical Review*, 51, 111-163.

Hendry, D. F. and K. F. Wallis (1984), "Editors' introduction" to *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell.

Keynes, J. M. (1931), "Economic possibilities for our grandchildren," in *Essays and Persuasions*. London: MacMillan.

Koopmans, T. C. (1957), *Three Essays on the State of Economic Science*. New York: McGraw Hill.

Lakatos, I. (1978), *The Methodology of Scientific Research Programmes*, Vol. 1, J. Warval and G. Currie (eds.). Cambridge: Cambridge University Press.

Leamer, E. E. (1978), *Specification Searches*. New York: Wiley.

_________ (1983), "Let's take the con out of econometrics," *American Economic Review*, 73, 31-44.

_________ (1987), "Econometric metaphors," in T. Bewley (ed.), *Advances in Econometrics*, Vol. 2. Cambridge: Cambridge University Press.

McAleer, M., A. R. Pagan and P. A. Volker (1985), "What will take the con out of econometrics?", *American Economic Review*, 75, 293-307.

Mizon, G. E. (1984), "The encompassing aprpoach in econometrics," in D. F. Hendry and K. F. Wallis (eds.), *Econometrics and Quantitative Economics*. Oxford: Blackwell.

Mizon, G. E. and D. F. Hendry (1980), "An empirical application and Monte Carlo analysis of tests of dynamic specification," *Review of Economic Studies*, 47, 21-45.

Pagan, A. R. (1984), "Model evaluation by variable addition," in D. F. Hendry and K. F. Wallis (eds.), *Econometrics and Quantitative Economics*. Oxford: Blackwell.

_________ (1987), "Three econometric methodologies: A critical appraisal," *Journal of Economic Surveys*, 1, 3-24.

Park, J. Y. and P. C. B. Phillips (1988), "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4, 468-497.

_________ (1989), "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory* (forthcoming).

Perron, P. (1987), "The great crash, the oil price shock and the unit root hypothesis," Cahier #8749, Université de Montréal.

Pesaran, M. H. (1986), "Editorial statement," *Journal of Applied Econometrics*, 1, 1-4.

Phillips, P. C. B. (1983), "Exact small sample theory in the simultaneous equations model," in M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics Vol. 1*. Amsterdam: North Holland.

_________ (1986), "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311-340.

_________ (1988a), "An interview with Professor J. Durbin," *Econometric Theory*, 4, 125-158.

_________ (1988b), "An interview with Professor A. R. Bergstrom," *Econometric Theory*, 4, 301-328.

_________ (1988c), "Optimal inference in cointegrated systems," Cowles Foundation Discussion Paper No. 866, Yale University.

_________ (1988d), "Spectral regression for cointegrated time series," Cowles Foundation Discussion Paper No. 872, Yale University.

_________ (1989), "Partially identified econometric models," *Econometric Theory* (forthcoming).

Phillips, P. C. B. and J. Y. Park (1988), "On the formulation of Wald tests of non-linear restrictions," *Econometrica*, 56, 1065-1084.

Poirier, D. J. (1988), "Frequentist and subjectivist perspectives on the problems of model building in economics" (with discussion), *Journal of Economic Perspectives*, 2, 121-144.

Prescott, E. C. (1986), "Theory ahead of business cycle measurement," *Carnegie-Rochester Conference Series on Public Policy*, 25, 11-44.

Sims, C. A. (1980), "Macroeconomics and reality," *Econometrica*, 48, 1-47.

________ (1987), "Making economics credible," in T. Bewley (ed.), *Advances in Econometrics*, Vol. 2. Cambridge: Cambridge University Press.

Sims, C. A., J. H. Stock and M. W. Watson (1986), "Inference in linear time series models with some unit roots," mimeographed, University of Minnesota.

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

Tufte, E. R. (1983), *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.