ISSN 0265-8003

WHAT WILL TAKE THE CON OUT OF ECONOMETRICS?

Michael McAleer
Adrian R Pagan
Paul A Volker

Discussion Paper No. 39
January 1985

Centre for Economic Policy Research
6 Duke of York Street
London SW1Y 6LA

Tel: 01 930 2963

CEPR Discussion Paper No. 39
January 1985

What Will Take the Con out of Econometrics?*

ABSTRACT

The paper begins with the question of whether Leamer's Extreme Bounds Analysis (EBA) really does "Take the Con Out of Econometrics"! By analytically demonstrating that the extreme bounds are simply functions of the F-statistic for the deletion of variables from a regression, we conclude that the information provided by EBA represents no advance over that available from traditional methods. Furthermore, there is a degree of arbitrariness in EBA which exactly parallels the selective reporting of regressions it was designed to supplant. The last part of the paper attempts a positive response to its title. By following a well defined series of modelling steps, we maintain that Cooley and Le Roy's EBA-derived conclusions concerning the interest elasticity of money demand owe more to a faulty methodology than to the data.

JEL classification: 211 212

Keywords: methodology, econometrics, extreme bounds analysis, Bayesian statistics

Michael McAleer and Adrian R Pagan
Department of Statistics
The Faculties
Australian National University
Canberra ACT 2600
Australia

Paul A Volker
Bureau of Labour Market Research
Canberra ACT
Australia

* An earlier version of this paper appeared under the title "Straw-Man Econometrics?" as Working Paper in Economics and Econometrics No. 097, Australian National University. It is available on request. We would like to thank all those who took the trouble to comment on that version and to correct our misunderstandings. We believe that those comments sharpened our arguments considerably. Special thanks go to Ed Leamer, Tom Cooley, Trevor Breusch, David Hendry, Allan Gregory, Hashem Pesaran, Peter Schmidt and Pravin Trivedi.

NON-TECHNICAL SUMMARY

For many years economists have debated what should be the proper methodology for economics. By contrast, econometricians have not tended to do so, perhaps preferring to get on with the job. Generally, what debate there was merely tended to reflect the 'classical' versus 'Bayes' arguments in statistics. In some ways this lack of interest was unfortunate, as there are at least three important benefits one might anticipate as flowing from adoption of a methodology. First, it would provide a set of principles to guide applied work. Second, codifying this body of knowledge would greatly facilitate teaching. Finally, a style of reporting that is both informative and succinct should emerge.

In the last five years interest in the methodology of econometrics has increased, and a number of distinctive styles of doing econometrics has emerged. Four major ones might be distinguished. The Bayesian stance in econometrics has long been advocated by Arnold Zellner. Time-series approaches have been exemplified in the work of Christopher Sims. An 'LSE' methodology can be found primarily in the work of Denis Sargan and David Hendry. Finally, Edward Leamer has set out a collection of methods that Trivedi has dubbed 'vestigial Bayesianism'.

The first three sections of our paper are concerned with this last methodology and how effective it is in the trinitarian role of guide, communicator and teacher. Our main claim is that Leamer's methodology indeed performs very poorly in the first two roles. To substantiate that claim we examine one of the major techniques used by those advocating this methodology: Extreme Bounds Analysis (EBA). Essentially this technique aims to provide extreme bounds for the parameter estimates of a general model under a range of possible simplifications of it. We demonstrate analytically that these extreme values are in fact related to conventional tests of hypotheses that particular simplifications of the model are compatible with the data. This methodology therefore presents in a different format the same

information as does traditional regression analysis such as is subsumed in the 'LSE' approach. But what is most disturbing is that the technique provides <u>less</u> information about a wide range of other relevant characteristics of the model. Thus we argue that 'vestigial Bayesianism' is inadequate both as a guide and communicator.

The final part of the paper looks at the best known application of the EBA technique - a demand for money study by Thomas Cooley and Steven Le Roy. Our aim is to show that the misgivings expressed about the theory of 'vestigial Bayesianism' also apply to its practice. Cooley and Le Roy found that the impact of interest rates on money demand is 'ill-determined'. We argue that Cooley and Le Roy arrived at such a conclusion only as a result of their adoption of this inadequate methodology.

What Will Take the Con Out of Econometrics?

Michael McAleer, Adrian R. Pagan and Paul A. Volker*

More than twenty years ago Carl Christ recounted a story about a new typist who rendered "econometrics" as "economic tricks". No doubt this tale was greeted with some amusement at that time; equally without doubt, today it would probably only occasion wry and knowing smiles. Charting the course of this transition, and accounting for its direction, has been the concern of a number of recent articles. Perhaps the most perceptive of these has been Edward Leamer's (1983). His contribution is of special interest, as it seeks not only to be descriptive but prescriptive; methods are outlined, the use of which Leamer sees as essential to the restoration of confidence in econometric research. Such techniques have now been promulgated and applied in a number of contexts. Thomas Cooley, for example, looks at the impact of industry concentration upon profits; Louis Dicks-Mireaux and Mervyn King consider the effect of pensions on savings; Thomas Cooley and Stephen Le Roy are concerned with money demand. These constitute just three of the more prominent applications.

Although the number of applications of the methods is growing, and approving references are being made to them, surprisingly few have queried the basis of the contention that the procedures really do allay some of the suspicion greeting econometric results; a singular exception being Phoebus Dhrymes. Yet the claims being made for this methodology are such as to demand a close investigation. As witnesses to these claims we quote Edward Leamer and Herman Leonard.

"We propose that researchers be given the task of identifying interesting families of alternative models and be expected to summarize the range of inferences which are implied by each of the families. When a range of inferences is small enough to be useful and when the corresponding family of models is broad enough to be believable, we may conclude that these data yield useful information. When the range of inferences is too wide to be useful, and when the corresponding family of models is so narrow that it cannot credibly be reduced, then we must conclude that inferences from these data are too fragile to be useful. ...

...The proper test of our proposals is whether they are useful in practice. We believe that researchers will find them to be efficient tools for discovering the information in data sets and for communicating findings to the consuming public". (p.306)

The aim of our paper is to consider possible answers to the question in the title. Because of its position as one proposed answer, and its strong advocacy by a number of authors, e.g. Cooley and Le Roy (p.827), we pay particular attention to Leamer's Extreme Bounds Analysis (EBA). In that inquiry, contained in Sections I, II and III, the discussion is structured along the lines of the three themes in the statement above: The effect of looking at different families of models, the determinants of a fragile inference, and the nature of the conclusions that may be drawn from the information provided by EBA. Based on the arguments of those sections, we conclude that EBA does not go very far in removing the con from econometrics. Furthermore, in most instances, it can actively distract a researcher from asking important questions about an econometric model.

But just because the promise and the performance of EBA diverge, does not obviate the need for a methodology aiming to dispel doubts arising over conventional research presentation and analysis. Accordingly, section IV sets out our own prescription, the basic ingredients of which are the necessity for a clear and full disclosure of the process whereby a preferred model was selected, and the requirement that a thorough evaluation has been made of the properties of such a specification. Such an orientation is scarcely original, reflecting in its concerns an oral tradition that owes much to Denis Sargan's (1964) influential paper on wages and prices. Nevertheless, it is worth explicitly stating these principles, as our experience convinces us that, consistently applied, they can go a long way towards the "de-conning" of econometrics. As an example of this approach, and to contrast our prescription with EBA, section IV re-examines the conclusions drawn by Cooley and Le Roy from their demand for money study.

I. The Problems in Families

Trying to define "the family" nowadays is enough to give a sociologist a nervous breakdown. To keep things simple one is inclined to assign a few individuals to its core and then to generate a whole range of alternatives by adding on children, grandparents, aunts, uncles, and other "relatives". Such a homely analogy captures rather nicely the essence of the "family of models" mentioned by Leamer and Leonard. At their core are variables classified as *important*. Added on are variables termed *doubtful*. What demarcates them is that *only the latter can be combined in an arbitrary linear fashion*. We emphasize this definition, since much of the discussion and use of EBA tends to proceed as if the division were based upon whether the associated coefficients are likely to be zero or not.[1] Because this is not so, the decision to assign variables to their respective classifications is not a trivial one, and it is explored in detail in section III.

To complete the elements of EBA, the concept of a *focus* variable is needed. This derives from the assumption that the magnitude of one of the model coefficients is of special interest. By itself, this feature tells us nothing about the nature of such a variable; it may be free or doubtful. Examples of both are given by Leamer. His "bleeding heart liberal" regards the impact of execution probability upon murders as doubtful, while his "right winger" treats the same variable as free.

Proponents of EBA work with the maximum and minimum point estimates of the focus coefficient as the set of restrictions upon the doubtful variables is changed. If the gap between these values is wide, readers are generally informed that no reliable inference can be drawn about the influence of the focus variable. Thus Cooley and Le Roy express the belief that almost nothing can be said about the value of the interest elasticity of the demand for money. Within one of their families of models this elasticity could lie anywhere between -6.27 to 2.24.

Now it is a rare family that does not have a member with problems at some stage or other. Families of models also share this characteristic, but this is rarely mentioned by EBA advocates. Notwithstanding that, it has to be the case that a consumer of the conclusion drawn from an application of the methodology must take some notice of the nature of the model that generated the extreme bounds. When this is done, there are at least two situations in which inferences drawn from EBA would have to be heavily discounted.

First of all, the restrictions that are being imposed upon the doubtful variables may be entirely unacceptable. Suppose that γ_1 and γ_2 are the parameters associated with the income and lagged dependent variable terms in a money demand function, and both variables are treated as doubtful. Then a restriction of the form $\gamma_2 - \theta\gamma_1 = 0$, with θ negative, would offend against theoretical conceptions. An extreme bound generated with $\theta < 0$ in a money demand example would be of little interest and, yet, there is nothing to safeguard against such a possibility. While Leamer himself is aware of this problem - see Leamer (1978, p.199) - there have been few attempts at cautioning users of EBA about it. Cooley and Le Roy do not mention it at all, despite the fact that income and wealth elasticities associated with one of the extreme bounds of the 90-day Treasury bill rate are actually negative.

Attempts have been made to limit such conflicts. Leamer (1982) restricts the feasible parameter space by requiring an investigator to put upper and lower limits on prior variances. It is hard to know what to make of this "solution", as the choice of such limits is extremely difficult and essentially arbitrary. One person's view of what constitutes a reasonable bound is unlikely to coincide with another's, and there is always the residual suspicion that prior variances have been chosen to yield narrow or tight bounds. As a satisfactory alternative to current practice it leaves much to be desired. It re-introduces the very element of whimsy that EBA was supposed to ameliorate.

A second alternative is to ensure that the extreme bounds do not disagree too greatly with the sample. Cooley does this by invoking the constraint that estimates should lie within the α% confidence ellipsoid associated with the least squares estimates of the complete model. A range of sample-modified bounds can then be generated by varying α. When $\alpha = 100$ the ordinary extreme bounds are found.

Once we introduce the sample evidence to constrain the alternative models, we are implicitly being asked to accept a number of conventions underlying EBA (at least as presented in the literature). Namely, that the errors in models be normally distributed, non-autocorrelated and homoscedastic; that the regressors be predetermined; and finally, that sample sizes are large enough for "confidence intervals" to be known with accuracy. No doubt these conventions may be appealing, but Leamer (1983, p.38) himself has pointed out the problem with their use: "Though the use of conventions does control the whimsy, it can do so at the cost of relevance". That principle is certainly applicable here, as the breakdown of any of these conventions means that the "α% confidence intervals" are anything but, and exactly what constraint is being applied becomes increasingly hazy.

These considerations emphasize the absolute necessity to know the point estimates of all coefficients in the model generating the bounds. But such knowledge is still not sufficient to decide if we have just come across a problem child or not. It is perfectly possible for all point estimates to appear reasonable, but for the model to be rejected on other grounds, such as when it exhibits substantial serial correlation. An extreme value generated from such a model would not be of great interest, since an investigator would not regard it as a suitable candidate for conveying information about the focus coefficient. Without knowing the *full set of characteristics* of the models generating the extremes, it is impossible to know what weight should be placed on the latter. Mere provision of the bounds, as in Cooley and Le Roy for example, is not enough. Much more information is needed to assess whether these bounds are meaningful.

II When is an Inference Fragile?

In what has transpired so far we have been somewhat vague about exactly how the bounds are used to conclude that an inference is fragile. If left that way, EBA becomes a "black box", and no understanding of the factors leading to an inference being fragile would be available. For this reason, we have gleaned two interpretations of fragility from the literature applying EBA, each of which is sufficiently precise to enable analytical results to be established.

The first of these, henceforth referred to as Type A fragility, is given by Leamer and Leonard as follows:

"These extreme values, $\hat{\beta}_{min}$ and $\hat{\beta}_{max}$, delineate the ambiguity in the inferences about β induced by the ambiguity in choice of model. If the interval $[\hat{\beta}_{min}, \hat{\beta}_{max}]$ is short in comparison to the sampling uncertainty, the ambiguity in the model may be considered irrelevant since all models lead to essentially the same inferences." (p.307)

With the sampling uncertainty measured as k times the estimated standard deviation of the focus coefficient, k being a predetermined constant, such a definition has been adopted by Leamer and Leonard, Cooley and Cooley and Le Roy. The first provide no guidance about k. Cooley selects a value of $k = 4$ while the last opt for $k = 2$. To investigate the consequences of adopting this definition of fragility, we provide Proposition 1 (proof available on request).

<u>Proposition 1:</u>

(a) *When the focus variable is doubtful, the necessary and sufficient condition for Type A fragility to exist is that the chi-square statistic for the doubtful variable coefficients to equal their prior means* (χ_D^2) *exceeds* k^2.

(b) *When the focus variable is free, the necessary condition for Type A fragility is that* $\chi_D^2 > k^2$.

Proposition 1 is quite striking, as it shows that whether an inference is to be Type A - fragile or not depends upon two quantities: namely, the significance of the doubtful variables in the model and the value chosen for k. Regarding the first, its magnitude will depend crucially upon the prior means assumed for the doubtful variables. If the prior means are taken to be zero, whereas the OLS estimates lie a long way from that point, a large value of χ_D^2 is likely. The closer the means are to the OLS values, the smaller will be χ_D^2, and the less the evidence of fragility. Everything therefore depends upon the whimsy of the reporter in the choice of prior means for doubtful variables! Hardly a good method for getting rid of the con artists. Instead it gives them enormous scope for generating almost any result they like. In the examples of EBA usage available, only Fiebig attempts to spell out this sensitivity of bounds to prior mean specification.

Proposition 1 moreover tells us something else of importance: That inferences will only be fragile if doubtful variables are informative. Assuming for convenience that prior means are zero, a large value of χ_D^2 signals to a researcher that these variables should appear in any model from which inferences are to be drawn. From this perspective, EBA is just an inefficient (and incomplete) way of communicating to readers the fact that the doubtful variables are needed to explain the data; a better solution would be to just present estimates of the general model along with an associated χ_D^2 statistic, letting consumers of research judge whether any further simplification of the model is justified.[2]

The analytic results presented in Proposition 1 can also rationalise the findings of a number of different investigations in which EBA has been employed. Leamer and Leonard's nuclear reactor example treats as doubtful those variables with t-values all below 1.03, leading to a lack of Type A fragility. In contrast, Cooley's profits regressions exhibit four variables with t-statistics greater than 3.5, and three of the four *always* appear as doubtful variables. Is it any wonder then that he concludes that Type A fragility exists for a concentration/profits relationship?

Perhaps the ambiguities raised above could be dissipated by an alternative definition of fragility. Leamer and Leonard provide just that, and we will designate it as Type B fragility in what follows. They say: "An alternative definition of shortness derives from a decision problem based on $\hat{\beta}$: the interval is short if all values in the interval lead to essentially the same decision". (p.307, fn. 1)

When implemented by Leamer (1983), Type B fragility occurs if there is a sign change implicit in the bounds. Ignoring, as Leamer does, the fact that these bounds themselves have standard errors, we proceed to analyse the nature of this type of fragility using Proposition 2 (proof available on request).

<u>Proposition 2</u>

(a) When the focus variable is doubtful the necessary and sufficient condition for Type B fragility to exist is that $\chi^2_D > \chi^2_{FO}$, *where* χ^2_{FO} *is the* χ^2 *statistic for testing if the focus coefficient is zero.*

(b) When the focus variable is free the necessary condition for Type B fragility is $\chi^2_D > \chi^2_{FO}$.

The movement from Type A to Type B fragility only changes the benchmark against which the significance of the doubtful variables is checked. It is no longer set by the reporter but determined by the data (χ^2_{FO}). Furthermore, when the focus variable is doubtful it is always the case that χ_D^2 exceeds χ_{FO}^2 (ignoring singularities in the design matrix), and so Type B fragility is in evidence. Whilst such a result is solely a consequence of the fact that zero is an admissible value for that doubtful variable coefficient, it serves to emphasize how Type B fragility may eventuate purely by a classification of variables. An example of this is provided by Leamer (1983) in his discussion of the impact of execution on murders. After placing the execution variable in the doubtful class, thereby producing an opposite sign to that from unrestricted least squares, he concludes: "I come away ... with the feeling that any inference from these data about the deterrent effect of capital punishment is too fragile to be believed". (p.42)

Since the sign change did not depend in any way upon the data, we find such a conclusion a trifle hard to defend.[3]

III When is a Variable Doubtful?

Propositions 1 and 2 strongly suggest that the conclusions on fragility drawn from EBA are intimately bound up with the classification of variables as doubtful and free. The polar case where the focus variable is orthogonal to all other regressors gives a striking demonstration of that fact. When treated as free, the gap between the bounds is zero, as the point estimate of the focus coefficient is entirely insensitive to combinations of other variables. But, when treated as doubtful, the width of the bounds varies directly with χ_D^2; the more significant the focus variable the greater the degree of fragility inferred.

A concrete example may serve to highlight just how important this choice can be to the outcome. Accordingly we consider the model of murder rates set out in the April 1983 SEARCH manual (it resembles that in Leamer (1982)). Table 1 gives the extreme bounds, range (the absolute value of the difference between the bounds) and ratio of range to standard errors for the impact of execution on murders under different variable designations.

As the definition of Type A fragility reflected the relative magnitudes of the range and standard deviation, the last column of Table 1 contains the information that would be used to assess whether inferences about the impact of executions upon murders are fragile. Clearly the decision about which variables are doubtful can have enormous consequences for any conclusions. Such variation naturally poses the question of how we are to know which one of the four options is to be adopted? Or, when is a doubtful variable doubtful? The answer must be that there is no answer. A decision to nominate a variable as doubtful is a personalised one, resting very much upon the opinions and values of the nominator. Consensus is no more likely over this choice than in the traditional selection of regressors problem.

Having elicited this point, the most serious defect in EBA becomes transparent: Unless extreme bounds are presented for *all* possible classifications of variables as doubtful and free, an observer cannot be certain that the selection does not constitute a "con job". *Selectivity in regression reporting therefore has as an exact analogue in EBA the different classifications of variables as doubtful and free.* EBA users report results for only particular variable categories and so are as arbitrary and selective in their *modus operandi* as the practices they criticize and claim to be improving on.

We can see this effect in Table 1. With nine variables in the regression there are 181,440 possible doubtful/free splits. Hence, inevitably some selection from this huge number will be made. Someone intent on demonstrating that executions deter murders would undoubtedly quote the final classification (or an augmented version), while those wishing to denigrate such a position would opt for the first two doubtful variable choices. There seems no reason to suppose that all of the classifications in Table 1 would be given by either protagonist, any more than one would anticipate each individual presenting the equivalent set of regressions composed of the different types of free variables. Thus there is little reason to believe that EBA provides a reporting style that is any better than that currently practised.

Sections I to III can now be drawn together to highlight the fact that EBA is not a satisfactory solution to the question posed in the title of this paper.[4] Section I argued that the extreme bounds themselves are not enough to enable conclusions to be drawn regarding fragility; we need to know the characteristics of the models generating such bounds. Sections II and III demonstrated that EBA is as capable of manipulation as the traditional presentation it aims to replace; perhaps more so in one respect in that an additional arbitrary choice of prior mean must be made. Consequently, if one feels unhappy with the information provided by *selective regressions*, one should not be any more satisfied with extreme bounds obtained by *selective variable partitions*. A con-man in one mode would have no fear of becoming deskilled in the other.

IV Cooley and Le Roy's Demand for Money Function: Contrasting the Methodologies

Given our belief that EBA cannot de-con econometrics, is there anything that might? Not generally, as there are almost certainly instances in econometrics, just as in science, of outright fraud. Nothing will detect such deception, except a vigorous critical tradition and a requirement that utilized data be either available or easily replicable. But our perception of the scepticism greeting many econometric studies is that it does not arise from a high incidence of such a phenomenon. Rather it stems from a feeling that the sins are venial rather than mortal; something has been left undone that should have been done.

Now EBA clearly addressed itself to this problem by indicating, for *a given variable partition and universe of variables*, the worst outcomes if everything conceivable were done. What it leaves undetermined is both the process by which the partition it is conditional upon was arrived at and the operating characteristics of models generating the extremes. Three points therefore always need to be considered in assessing an EBA. In turn, these three elements also occur in the traditional line of research and are, we believe, the source of much of the dissatisfaction with it. Because they are pivotal to the methodology advanced in this section, we list them below:

(A) Selection of a general model;

(B) How and why any general model was simplified to the preferred one(s);

(C) Quality control of the preferred model(s).

Selection of a general model is a problem with all research methodologies (including EBA) and we can do no better than concur with Leamer and Leonard when they say: "But it is up to readers of research to decide if the reported family of models is credibly inclusive. If the researcher, for whatever reason, selects an incredibly narrow family of models, readers will properly ignore the results". (p.307)

Even if we largely agree that the choice of variables considered in an investigation was commendably large, it is frequently the case that little discussion is provided of the strategy employed to obtain a more parsimonious representation of the data. Where a systematic reduction is possible it should be followed; where it is not, detail should be sufficient to enable a consumer of the research to determine exactly the criterion adopted in performing the simplification. At a very minimum this forces the presentation of an estimated general model and some analysis of how the preferred model relates to it.

Our final category focusses upon the quality control exercised on the models presented. Frequently, this is little short of abysmal and, as James Ramsey comments "... it is amazing that so little is done to evaluate the model and the results" (p.242). Yet ultimately quality control is as important for the econometrics profession as it is for automobile manufacturers. A gradual realisation of this point has in fact stimulated the development of criteria for the formal evaluation of models. For later reference it is useful to summarise the outcome of that research by classifying derived methods into five major categories:

1. Consistency with theory.
2. Significance, both statistical and economic.
3. Indexes of inadequacy.
4. Fragility or sensitivity.
5. Whether a model can encompass or reconcile previous research.

These five categories can be viewed as a re-grouping of the criteria suggested in David Hendry and Jean-François Richard for settling upon a "tentatively adequate" model. Categories 1 and 2 have tended to dominate in past evaluative analysis and even now constitute the corpus of most applied econometrics courses and texts. Increasing attention has, however, been paid to the necessity of item 3, with David Hendry (1980) giving a general perspective and Hendry (1983) a detailed application. Robert Engle (1982b) and Adrian Pagan and Anthony Hall (1983) provide an account of much of the technology, emphasizing that these methods aim to extend the horizon in directions where errors might be anticipated. Some indexes, such as the Durbin-Watson statistic, have been routinely used in applied work. But, as the articles referenced above demonstrate, the set of indexes *conventionally* reported is much too small to be completely effective.

Item 4 encompasses considerations raised by EBA. However, in contrast to the emphasis placed by EBA upon sensitivity of point estimates to a change in the menu of included variables, there is an older tradition of assessing the fragility of models by reference to new data. This is done either through predictive failure, recursive estimation, or to interaction with other parts of a model as in simulation analysis. Fragility as an important criterion for model evaluation is therefore not a novel idea. Rather it is the emphasis EBA places upon variation in point estimates of a particular coefficient under model re-specification which is novel. In fact, an EBA would seem to constitute an important part of the evaluative process. It must be a rare instance in which some arbitrariness does not creep into the simplification process, particularly when working with cross section data. The extreme bounds then provide useful information upon the effects of such arbitrary decisions, at least in respect of the focus coefficient. Such is the way we employ EBA in the following case study.

The final category distinguished above challenges a model to encompass or explain alternative models, particularly those originating from past endeavours. Lack of reconciliation between studies is a glaring defect in much current applied research, and this requirement, whether interpreted formally as in Grayham Mizon and Jean-François Richard, or rather more informally as in James Davidson et al., must become an essential cornerstone for applied econometric research. Only if it is met can one be truly satisfied that progress has been made in understanding an empirical phenomenon.

In order to contrast the methodology outlined above with the approach of those viewing EBA as the cornerstone of econometric work, we will look at the money demand function inquiry presented in Cooley and Le Roy. This paper has been cited approvingly by a number of authors, both for what it said about econometric practice and for its claim about the likely magnitudes of interest elasticities. In doing our comparison we have presumed that the study was meant to be a serious application of the EBA methodology, rather than just illustrative. Certainly, there is support for this hypothesis in the stress Cooley and Le Roy laid upon the conclusions drawn from their analysis.

One fact that should by now be apparent from our assignation of EBA to a group of methods for model evaluation, is our belief that exclusive attention to the results from it can lead to quite erroneous conclusions about the robustness of parametric inferences. Such tunnel vision tends to distract researchers from the other vital questions needing to be asked. A primary example would be whether the model upon which EBA is being practiced is comprehensive enough. Later it is argued that, in Cooley and Le Roy's case, there is ample evidence of it not being so.

For the moment we accept their formulation of the problem, turning instead to one of the items in the list assembled earlier as bedevilling EBA; namely, the way in which conclusions on fragility are attendant upon the assumed doubtful/free division. Table 2 below shows how important such selections were for Cooley and Le Roy's conclusions concerning their second specification (Table 2, p.836).

The extreme bounds shrink dramatically when the intercept is made a free variable (note that Cooley and Le Roy (Table 1, p.835) do not indicate it as doubtful but the bounds of -6.27 and 2.24 from their Table 2 only occur when it is so treated). With a t-statistic of -3.96 such an outcome should not be surprising given Proposition 1 of section II. Building a case for the treatment of the intercept as doubtful rather than free would, to our minds, be quite difficult, but the most important lesson from Table 2 is how misleading it is to give the extreme bounds for a single doubtful/ free partition of the variables.[5]

Table 2 shows that any of the conclusions drawn by Cooley and Le Roy about the magnitude of interest elasticities must be treated with scepticism, even if the output of EBA is taken as the dominant source of information on these parameters. The wide bounds relied upon for their critique appear to have been manufactured solely by a particular variable classification. But the inadequacies in their work are even more serious than that. No attention was paid by them at all to the quality of the model used for EBA, and it is therefore appropriate that we briefly review it.

In Cooley and Le Roy's specification the demand for real money (M1) is held to be a function of two interest rate variables, the savings and loan passbook rate (RSL) and the ninety-day Treasury bill rate (RTB), real GNP (nominal GNP divided by the GNP deflator, P), the current inflation rate (INF), the real value of credit card transactions (VCC) and real wealth (W). They use seasonally adjusted quarterly data for the period 1952(2) to

1978(4), and present (p.834) estimates for a log-linear specification. Our estimates of their model are in Table 3.[6]

To evaluate their estimated equation it is sufficient to note that the most basic index of inadequacy, the Durbin-Watson statistic, is 0.063. This is an example of the situation condemned by Granger and Newbold in which the Durbin-Watson statistic is markedly exceeded by the R^2 and in which arises the danger of the "spurious regression" phenomenon. It is clearly not sensible to investigate fragility with such an inadequate model.

From the above discussion one is entitled to be dubious of the validity of Cooley and Le Roy's claim that the interest elasticity of the demand for money cannot be known with much precision. Nevertheless, it could be correct. Moreover, in view of the prominence of the topic in the literature, and the particular stand taken by Cooley and Le Roy on the issue, it is interesting to see what type of model would have evenutated, *given only the data series used by Cooley and Le Roy as input*, if a proper modelling strategy had been followed.[7] That strategy involves the three stages described at the beginning of this section.

A. Selection of the General Model

Under the restriction on the universe of available variables, the main direction in which generalization of Cooley and Le Roy's model can take place is in the order of dynamics. Given the wide use of distributed lags in modelling money demand it seems extraordinary that the authors chose to ignore Sims' maxim that "a time series regression model arising in econometric research ought in nearly every case to be regarded as a distributed lag model until proven otherwise" (p.289). The omission of dynamics is even stranger in the light of the authors' own comments (p.840): "Such lagged endogenous variables as the lagged money stock ... cannot plausibly be excluded from the demand side either explicitly as observable explanatory variables for the demand for money or implicitly through the time dependence of the error".

Our general model therefore has the same variables as Cooley and Le Roy but with four lags on all variables (including the dependent). This lag structure seems reasonable considering the data used are quarterly. The period selected for study was, however, shorter than that used by Cooley and Le Roy. John Judd and John Scadding have recently noted that a large number of studies have experienced difficulty in estimating conventional money demand functions for the post-1973 period. Not only do these models predict poorly, but in a large number of cases such models are dynamically unstable. Various reasons for the poor performances of the models are canvassed by Judd and Scadding. Among them, "the most likely cause of the observed instability in the demand for money after 1973 is innovation in financial arrangements" (p.1014), which originated from the rapid rise in inflation during the period. In accordance with this view, we restricted ourselves to the sub-sample 1952(2) to 1973(4), with the first four observations used for constructing up to four lags on all variables.[8]

B. Simplification of the General Model

Our first step in simplification of the general model represents an attempt to determine the order of dynamics on each of the variables through a sequence of nested tests. The procedure we use was proposed by Sargan (1980), and has been termed the COMFAC algorithm, due to the fact that it seeks to determine common factors in the distributed lag polynomials associated with each variable. Briefly the logic of the method is as follows.

Suppose the general model had the form

$$(1) \qquad y_t = b_1 y_{t-1} + .. + b_4 y_{t-4} + c_o x_t + .. + c_4 x_{t-4} + e_t \, .$$

With the aid of lag operators (1) can be re-written as

$$(2) \qquad b(L)y_t = c(L)x_t + e_t$$

where $b(L) = 1 - b_1L \ldots - b_4L^4$ and $c(L) = c_o + c_1L + \ldots + c_4L^4$ are polynomials in the lag operator L. If the term $(1 - \rho_1 L)$ is a common root of both polynomials, (2) can be re-expressed as

$$(3) \qquad b^*(L)y_t = c^*(L)x_t + u_t$$

with $b(L) = (1 - \rho_1 L)\, b^*(L)$, $c(L) = (1 - \rho_1 L)\, c^*(L)$ and $(1 - \rho_1 L)u_t = e_t$.

An examination of (3) shows that the presence of a common factor has created a new model with maximum lag of three in y_t and x_t and first order serial correlation (AR(1)) in the errors. As there were nine parameters in (2) and only eight in (3), a restriction has been imposed, whose validity may be tested. If the restriction is accepted, the model is capable of being simplified. Moreover, if ρ_1 turns out to be zero, the original model must have had both the orders in y_t and x_t overstated.

In our general model there are six regressors apart from the intercept. Hence, in the analogous move from (2) to (3), six restrictions are being imposed in the first attempt at simplification. If the first common factor is accepted, imposing the second leads to a further six restrictions, with the equation error now given as AR(2), $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t$. In this way, each additional common factor restriction leads to six fewer estimated coefficients. Since we have a sequence of nested tests, we set the level of significance of each test at one percent so as to have an overall level of significance of approximately four percent.[9]

There are two difficulties that can arise in using the F test to test the common factors. First, there will generally be multiple minima for the sums of squares (see Sargan, 1980) and, second, there is no guarantee that the common roots in the polynomials attached to the variables are real. In order to guard against complex roots, we test for two common factors initially, and then test for one common factor only if two are rejected. Thus, in testing for the first two common factors in Table 4, the calculated F statistic is 1.7555, the unrestricted (restricted) lag length is 4(2) and 12 restrictions are being tested (6 associated with each common factor).

Compared with the critical value of F(12, 48, .01) = 2.59, the calculated statistic is not significant. At this stage, then, the lag length has been reduced from 4 to 2 and the equation error can be expressed as AR(2). Testing third and fourth common factors gives a value of 3.904 for the F statistic which is significant at one percent. Therefore, the third and fourth common factors are rejected. Returning to the second-order lag and testing for the third common factor only gives a value of 0.426 for the F statistic. Since the calculated value is significantly less than 3.12, three common factors are accepted. The model can now be expressed as one lag on all variables, with an equation error given as AR(3). The resulting model is referred to as the "simplified" one in Table 3.

The following observations are relevant to the simplified dynamic specification given in Table 3. Of the four interest rates, only current RSL is significant, and, apart from the lagged dependent variable, the only significant lagged variable is real wealth. Moreover, current and lagged wealth have coefficients which add to zero exactly. Neither the current nor lagged real value of credit card transactions exerts a significant affect on real balances. Finally, the third common factor (ρ_3) is not significantly different from zero, thereby reducing the implicit lag length of the specification by one.

It is fairly clear that the model is still overparameterized. Accordingly, we imposed a further eight restrictions, namely zero coefficients for VCC and the lagged values of RTB, RSL, INF, GNP and VCC, a zero sum for the wealth coefficients and a zero value for the third common factor. The calculated F statistic of 1.001 is significantly less than F(8, 66, 0.01) $\simeq$ 2.8, leading to acceptance of the restrictions. Our preferred model is therefore the last one listed in Table 3.

C. Quality Control: Is the Model a Lemon?

How does the estimated model in Table 5 stand up to the five criteria for quality control listed at the beginning of this section? With the exception of the term $\Delta \ln W_t$ it constitutes a very traditional specification of money demand. The presence of the change in, rather than the level of, wealth is however consistent with theoretical considerations. If transactions requirements are held constant, ie. GNP_t is fixed, the fact that money (M1) is an asset dominated for portfolio purposes by interest-bearing deposits of near equal liquidity suggests that the long-term wealth effect should be zero. In the short run though it has been frequently noted that changes in wealth are initially held as demand deposits before re-allocation, and the combination of $\Delta \ln W_t$ and the lagged dependent variable describes such a process, the implied lag distribution being .178, -.029, -.025 etc. Perhaps the only difficulty with such an interpretation is that the portfolio re-allocation process is not faster.

Table 5 investigates whether there are any obvious "inferential monsters lurking beyond the horizon", by augmenting the moments of the preferred model with a number of variables designed to capture inadequacy. No striking deficiencies are in evidence. A number of other experiments were conducted to determine whether it was possible to reject the chosen model by the addition of particular variables. These included a number of lags in real GDP, RSL etc., time trends, seasonal dummies and estimation with up to seventh order serial correlation pattern. None of these augmentations was found to contribute anything of significance. A final point worth mentioning is that t-statistics, made robust to heteroscedasticity as suggested by Halbert White, were about 10 percent higher than those in Table 3. The only exception to this rule - that for inflation - was only slightly smaller.

A check on parameter constancy is available by examining the size of prediction errors made when an equation is estimated over a particular sample and then used to forecast out of sample.[10] In this vein, the preferred model was estimated to 1970(4) and one-step prediction errors were generated from 1971(1) to 1973(4) by augmenting the preferred equation with the Type B constructed variables in Adrian Pagan and Desmond Nicholls. The "F test" that the coefficients on the twelve constructed variables were jointly zero was 1.58, well below the critical F(12, 62, .01) value of 2.49. Although an examination of the individual errors does reveal one large error, namely that for 1972(1), where the t-value was 2.58, the prediction errors for 1971-1973 were much the same as the sample errors, with an average absolute value of .4 percent.[11]

D. Are Interest Elasticities of Money Demand Zero?

As our model was of satisfactory quality to 1973(4), it is reasonable to utilize it to shed light on the question of whether data is uninformative about interest elasticities, as alleged by Cooley and Le Roy.[12] Conditional upon the structure of the final model being valid we can say that all variables in the estimated relationship (including both interest rates) are highly significant, and to adopt their hypothesis of a zero interest rate effect as an acceptable interpretation of the data would be totally inappropriate. To be sure, this final specification was arrived at after a decision in which an arbitrary group of variables was dropped because of insignificance. To assure readers that the well defined interest elasticities found in our preferred model were not dependent upon this action, and to illustrate what we believe is the place of EBA, we computed the extreme bounds for the two long-run interest elasticities. This was done by making the coefficients on either RSL or RTB the focus, treating all excluded variables as doubtful, and using the estimates of parameters on $\ln M_{t-1}$, $\ln RTB_t$ (or $\ln RSL_t$), $\ln RTB_{t-1}$ (or $\ln RSL_{t-1}$) associated with the bounds to obtain long-run responses. To be consistent with Cooley and Le Roy we concentrate upon the long-run elasticities as they summed lagged coefficients when dynamics were admitted. These bounds were extremely narrow, being -.053 to -.068 (RTB) and -.400 to -.441 (RSL), indicating that the effect of interest rates upon money demand was not sensitive to our decision to exclude certain variables.

V Conclusions

That applied econometrics is not currently in the most robust of health is hard to deny, and it would be difficult to find as entertaining or as perceptive an analysis of its ills as that found in Leamer's various articles. What concerns us is that the prescriptions made in those articles are inappropriate, in part because of faulty diagnosis. Extreme bounds analysis (EBA) is most emphatically *not* the medicine to cure an ailing patient.

Section I argued that extreme bounds are generated by the imposition of highly arbitrary, and generally unknown, restrictions between the parameters of a model. Exactly why such bounds should be of interest therefore becomes something of a mystery. Furthermore, as shown in sections II and III, the methodology is flawed on other grounds. EBA demands a general, adequate model from which the bounds may be derived, and a consensus over which variables are critical to a relationship. These are highly questionable conventions and we demonstrated, both theoretically and empirically, that deviations from them almost completely negate the utility of EBA.

After largely rejecting EBA, section IV of the paper moved on to our own diagnosis and prescription. Both are founded on the belief that many of the difficulties applied econometrics currently faces originate in the very poor attempts currently made to accurately describe the process whereby a model was selected, and to ascertain its adequacy. Acceptance of this proposition leads to the necessity for the establishment and promulgation of standards with which to conduct applied research. Many other disciplines have faced and taken steps to solve this problem, and movement in this direction is long overdue in econometrics. With these considerations in mind we proposed a three stage approach to modelling, involving the selection and subsequent simplification of a general model and a rigorous evaluation of any preferred model. Under the latter heading five ways of performing such an evaluation were distinguished. It may not be too fanciful to think of such criteria as a "check-list" to be applied when reviewing or performing applied work. Only if a model passes most items on the list should it be seriously considered as augmenting our knowledge.

Having set up some yardsticks with which to evaluate models, section IV applied them to the money demand example in Cooley and Le Roy. Their specification was found to fail even the simplest of these criteria, making any conclusions drawn from it highly suspect. In sharp contrast to this failure, the application of a modeling strategy beginning with a general model and progressively constraining the parameter space led to a representation which passed all items of the "check-list". This example highlighted the benefit of a systematic approach to modelling and model evaluation.

In closing, a confession. We are only too aware that what has been described are the necessary rather than sufficient conditions for taking the con out of econometrics. As any users of corporate accounts will be aware, there are many ways around standards. But that is not to deny their value. It serves only to highlight the need.

* The first two authors are at the Department of Statistics, The Faculties, Australian National University, Canberra, A.C.T. 2600, Australia. Volker is with the Bureau of Labour Market Research, Canberra, A.C.T., Australia. An earlier version of this paper appeared under the title "Straw-Man Econometrics?" as Working Paper in Economics and Econometrics No. 097, Australian National University. It is available on request. We would like to thank all those who took the trouble to comment on that version and to correct our mis-understandings. We believe that those comments sharpened our arguments considerably. Special thanks go to Ed Leamer, Tom Cooley, Trevor Breusch, David Hendry, Allan Gregory, Hashem Pesaran, Peter Schmidt and Pravin Trivedi.

1 Unfortunately, the terminology of "important" and "doubtful" tends to bolster this impression. For this reason, we substituted "free" for "important" as that captures the nature of these variables much more closely. Ideally, a similar change would have been desirable for "doubtful".

2 It is of interest to specialize k to 2. When only a single doubtful variable is present Type A fragility occurs when the t-statistic of the doubtful variable exceeds 2, which is a conventional rule of thumb for selection of regressors. As the number of doubtful variables grows, however, a constant value of $k = 2$ means that the comparison of χ^2 with 4 corresponds to larger and larger levels of significance. Most researchers would presumably find this implicit assumption in EBA a little odd.

3 Note that a sign change also occurred when the execution variable was free under the "eye for eye" specification. With *eleven* doubtful variables, and a t-statistic of less than two for the execution coefficient, an application of Proposition 2(b) should leave us in little doubt over why that was so.

4 It is important to emphasize that an answer to this question is our central concern. We do not quibble with the contention that EBA displays the impact of prior information on posterior means. To do so would be inconsistent with our Proposition 1. Nor do we argue that, *for a given variable partition*, EBA might not be useful. In section IV we do, in fact, exploit it in exactly such a context.

5 In fact Cooley and Le Roy present a broader range of bounds than those in Table 2, invoking the extra constraint that coefficient estimates must lie in a specified confidence ellipsoid. Those in Table 2 correspond to the 100% ellipsoid, and represent wider bounds than most contained in their Table 2.

6 Tom Cooley kindly made their data available to us. We were able to reproduce their results with the exception that the real wealth elasticity should be -0.107 rather than +0.107, and the inflation rate should not be in logarithms since negative rates were observed over the sample period.

7 The restriction seems necessary to avoid the situation where differences in any conclusions we reach to those of Cooley and Le Roy are simply a consequence of our using information not available to them.

8 The counterpart to Cooley and Le Roy's model over this shorter period gives parameter estimates -2.90, -.021, -.382, -.009, -.616, -.017 and -.052 with standard error of estimate .0083.

9 To test the restrictions, we used the standard F-test given by $F = [(\tilde{e}'\tilde{e}-\hat{e}'\hat{e})/\hat{e}'\hat{e}].\ [(T-\hat{k})/r]$, where $\tilde{e}'\tilde{e}$ is the sum of squared residuals from the restricted model, $\tilde{e}'\tilde{e} \rightarrow \hat{e}'\hat{e}$ in its unrestricted counterpart, $(T-\hat{k})$ is the degrees of freedom of the unrestricted model and r is the number of restrictions to be tested. In this way some allowance is made for the number of parameters estimated in the unrestricted model.

10 A more detailed analysis is available in our "Straw-man Econometrics?".

11 Although our model gave satisfactory performance up to 1973, just like automobiles age finally caught up with it, and after that date its predictive performance declined dramatically. For the twelve quarters after 1973(4), the F-test that prediction errors were zero was 5.69, with only the errors for 1974 not being significantly different from zero individually. The absolute error was 1.7% over this three-year period. Thus Stephen Goldfeld's puzzle of the "missing money" is certainly not resolved by working with Cooley and Le Roy's data alone.

12 Encompassing tests were also advocated to assess model quality. These are not really possible here given the restriction placed upon the data set, although it is clear that our model dominates those which exclude either of RSL or RTB, the inflation rate, wealth or explicit dynamics. In Hendry and Richard's terminology our model strongly variance - encompasses Cooley and Le Roy's as is apparent from the standard errors of estimate in footnote 8 and Table 3.

Table 1

Extreme Bound Information for Execution Coefficient (PX)

Free Variables	Min	Max	Range	Range/Std. Dev.
None	-2.87	2.72	5.59	115.0
PX	- .45	1.35	1.8	37.0
PX, intercept	- .40	.10	.5	10.3
PX, intercept, other variables with t > 3(S, PC, PCTPOOR)	- .22	- .01	.21	4.3

Note: PC = probability of conviction, PX = probability of execution, S = sentence, PCTPOOR = percent poor, standard deviation of focus coefficient = .0486.

Table 2

Extreme Bounds for Long-Term Interest Elasticity (RTB)

Free Variables	Min	Max
None	-12.14	12.15
RTB	- 6.27	2.24
RTB, intercept	- .375	.019

Table 3

Alternative Estimates of the Money Demand Function[a]

	M1	RTB	RSL	INF	GNP	VCC	W[b]
Cooley and							
Le Roy		.010	-.175	-.036	.372	-.009	-.107
		(.011)	(.069)	(.167)	(.081)	(.055)	(.096)
	SEE = .028 DW = .063						
Simplified							
Model							
Lag 0		-.003	-.111	-.156	.048	-.009	.178
		(.003)	(.040)	(.029)	(.051)	(.018)	(.045)
Lag 1	.866	-.005	.053	-.012	.062	.007	-.178
	(.054)	(.003)	(.043)	(.026)	(.058)	(.017)	(.048)
	SEE = .0031 DW = 1.938 ρ_1 = .306 (.133) ρ_2 = -.219 (.129) ρ_3 = .194 (.131)						
Preferred							
Model	.835[c]	-.009	-.074	-.146	.126		.178
	(.047)	(.002)	(.021)	(.026)	(.027)		(.041)
	SEE = .0031 DW = 2.024 ρ_1 = .391 (.118) ρ_2 = -.301 (.112)						

Notes: (a) All regressors except the inflation rate are in logs. Constant term is not shown. Standard errors are in parentheses, SEE = standard deviation of residuals, DW = Durbin-Watson statistic.

(b) For the preferred regression this column is $\Delta \ln W$.

(c) Coefficient of lagged real money (in logs).

Table 4

Tests of Common Factors (83 observations)

Common factor	Unrestricted lag length	Restricted lag length	F statistic	D.F.	Critical F(0.01)
1,2	4	2	1.755	(12,48)	2.59
3,4	2	0	3.904	(12,60)	2.50
3	2	1	0.426	(6,60)	3.12

Table 5

Indexes of Adequacy for the Preferred Model

Statistic Type	Statistic Value	Critical Value
RESET[a]	3.27	$F(2,74,.01) \simeq 4.9$
Diff. Test[b]	2.59	$\chi^2(8,.01) = 20.09$
Normality Test[c]	1.74	$\chi^2(2,.01) = 9.21$
Hetero. Test[d]	1.82	$\chi^2(1,.01) = 6.63$
A.C.F. of squared residuals[e]	(1) -.68 (3) .51 (2) .49 (4) .36	S.N.D. (.01) = 2.33
A.C.F. of residuals[f]	(1) 1.22 (3) 0.73 (5) 1.62 (7) 0.22 (2) 1.17 (4) 0.53 (6) 0.28 (8) 1.12	

Notes: (a) The "F test" that the coefficients of the predictions squared and cubed in the regression of the residuals against these and the derivatives are zero. Computation was done via partitioned inversion to avoid serious numerical inaccuracy.

(b) The differencing test of Charles Plosser et al. One iteration of Sargan's (1959) AIV estimator upon the differenced model was performed from the estimates in Table 3. Instruments for the derivatives with respect to the coefficients of M_{t-1} and u_{t-1} were constructed as in Plosser et al. (footnote 7).

(c) The joint normality test of Bowman and Shenton or Anil Bera and Carlos Jarque.

(d) The LM test that $\gamma = 0$ in $\sigma^2 = \sigma^2(E(y_t))^\gamma$ where y_t is the dependent variable of a regression. Pagan et al. (1981) derive this LM test but it was proposed originally as a test for heteroscedasticity by Anscombe.

(e) The "t-statistics" were formed by regressing the squared residuals against their lagged values. This approach was used by Clive Granger and Allan Anderson for the detection of non-linear models but can also be used to check for Engle's (1982a) ARCH effects or as a general specification error test.

(f) Writing the model as a non-linear regression $y = f(X;\theta) + \varepsilon$, the "t-statistics" that the coefficient of the lagged residuals $\hat{\varepsilon}_{t-j}$ are zero in the regressions of $\hat{\varepsilon}_t$ against $\hat{\varepsilon}_{t-j}$ and

$$\partial f_t/\partial\theta \quad \text{for} \quad j = 1,..,8.$$

References

Anscombe, F.J., "Examination of Residuals," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 4 : 1961, 1-36.

Bera, Anil K. and Jarque, Carlos M., "An Efficient Large Sample Test for Normality of Observations and Regression Residuals," *Australian National University Working Papers in Economics and Econometrics*, No. 040, 1981.

Bowman, K.O. and Shenton, L.R., "Omnibus Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2," *Biometrika*, No. 2, 1975, 62, 243-250.

Christ, Carl F., "Econometrics in Economics : Some Achievements and Challenges," *Australian Economic Papers*, December 1967, 6, 155-170.

Cooley, Thomas F., "Specification Analysis with Discriminating Priors: An Application to the Concentration Profits Debate," *Econometric Reviews*, No. 1, 1982, 1, 97-128.

Cooley, Thomas F. and Le Roy, Stephen F., "Identification and Estimation of Money Demand," *American Economic Review*, December 1981, 71, 825-844.

Davidson, James E.H., Hendry, David F., Srba, Frank and Yeo, Stephen, "Econometric Modelling of the Aggregate Time-series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal*, December 1978, 88, 661-692.

Dhrymes, Phoebus J., "Comment," *Econometric Reviews*, No. 1, 1982, 1, 129-132.

Dicks-Mireaux, Louis and King, Mervyn," Pension Wealth and Household Savings : Tests of Robustness," *Journal of Public Economics*, 1984, forthcoming.

Engle, Robert F., (1982a) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, July 1982, 50, 987-1007.

——————— , (1982b) "A General Approach to Lagrange Multiplier Model Diagnostics," *Journal of Econometrics*, October 1982, 20, 83-104.

Fiebig, D.G., "A Bayesian Analysis of Inventory Investment," *Empirical Economics*, 1981, 6, 229-237.

Goldfeld, Stephen M., "The Case of the Missing Money," *Brookings Papers on Economic Activity*, 1976, 3, 683-730.

Granger, C.W.J. and Andersen, A., *An Introduction to Bilinear Time Series Models*, Gottingen : Vandenhoeck and Ruprecht, 1978.

Granger, C.W.J. and Newbold, P., "Spurious Regressions in Econometrics," *Journal of Econometrics*, July 1974, 2, 111-120.

Hendry, David F., "Econometrics : Alchemy or Science?," *Economica*, November 1980, 47, 387-406.

——————— , "Econometric Modelling : The Consumption Function in Retrospect," *Scottish Journal of Political Economy*, 1983, 30, 193-220.

Hendry, David F. and Richard, Jean-François, "On the Formulation of Empirical Models in Dynamic Econometrics," *Journal of Econometrics*, October 1982, 20, 3-33.

Judd, John P. and Scadding, John L., "The Search for a Stable Money Demand Function : A Survey of the Post-1973 Literature," *Journal of Economic Literature*, September 1982, 20, 993-1023.

Leamer, Edward E., *Specification Searches : Ad Hoc Inference with Non-experimental Data*, New York : Wiley, 1978.

―――――――― , "SEARCH, A Linear Regression Computer Package," mimeo., University of California - Los Angeles, 1981.

―――――――― "Sets of Posterior Means with Bounded Variance Priors," *Econometrica*, May 1982, 50, 725-736.

―――――――― , "Let's Take the Con Out of Econometrics," *American Economic Review*, March 1983, 73, 31-43.

Leamer, Edward E. and Leonard, Herman, "Reporting the Fragility of Regression Estimates," *Review of Economics and Statistics*, May 1983, 65, 306-317.

Mizon, Grayham E. and Richard, Jean-François, "The Encompassing Principle and Its Application to Non-nested Hypotheses," paper presented to the European meeting of the Econometric Society, Dublin 1982.

Pagan, A.R. and Hall, A.D., "Diagnostic Tests as Residual Analysis," *Econometric Reviews*, No. 2, 1983, 2, 159-218.

Pagan, A.R., Hall, A.D. and Trivedi, P.K., "Assessing the Variability of Inflation," *Australian National University Working Papers in Economics and Econometrics*, No. 049, 1981.

Pagan, A.R. and Nicholls, D.F., "Estimating Predictions, Prediction Errors and their Standard Deviations Using Constructed Variables," *Journal of Econometrics*, March 1984, 24, 293-310.

Plosser, Charles I., G. William Schwert and White, Halbert, "Differencing as a Test of Specification," *International Economic Review*, October 1982, 23, 535-552.

Ramsey, James B., "Perspective and Comment," *Econometric Reviews*, No. 2, 1983, 2, 241-248.

Sargan, J.D., "The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society, Ser. B*, No. 1, 1959, 21, 91-105.

———————, "Wages and Prices in the United Kingdom : A Study in Econometric Methodology," in Hart, P.E., G. Mills and J.K. Whitaker, eds., *Econometric Analysis for National Economic Planning*, London : Butterworths, 1964, 25-63.

———————, "Some Tests of Dynamic Specification for a Single Equation," *Econometrica*, May 1980, 48, 879-897.

Sims, Christopher A., "Distributed Lags," in Intriligator, M.D. and Kendrick, D.A., eds., *Frontiers of Quantitative Economics II*, Amsterdam : North-Holland, 1974, 289-338.

White, Halbert, "A Heteroskedasticity - consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, May 1980, 48, 817-838.