1. Introduction

This paper gives some miscellaneous comments from the outside viewpoint of a statistician on the challenges associated with quantitative methods in econometrics. There are several matters to be considered and the relative importance of these must depend on the particular context. Emphasis can be placed on

1. the provision of better data and of data more specific to the research questions of concern, including detailed discussion of measurement issues and data definition and especially data quality;
2. more intensive analysis by already developed methods of currently available data;
3. more critical interpretation of analyses made by currently available methods;
4. the development and deployment of more elaborate methods of analysis;
5. the deeper incorporation into analysis of specific subject-matter concepts and theory.

All these are surely important and some are complementary rather than contradictory. In particular the last two issues are quite strongly connected; more incorporation of specific quantitative considerations is likely to lead to new types of model in turn needing special methods for their analysis. While the development of new methods is the aspect of most immediate interest to the theoretical statistician, there is some danger of overweighting this aspect! In particular, issues of data quality and relevance, while underemphasized in the theoretical statistical and econometric literature, are certainly of great concern in much statistical work.

Of central importance in much statistical thinking is the explicit recognition that conclusions are uncertain and that this uncertainty, or at least that part of it that arises from haphazard variation in the data, should be measured. There is moreover often an implicit tradition, stemming perhaps from R.A. Fisher's work largely in biometry, that each investigation should be self-contained, providing its own estimate of error. At the same time it is clear that interpretation cannot proceed in isolation and the merging of the results of
different investigations, and especially investigations of different types, is crucial. This underlies the fifth point above and some of the criteria to be discussed in the next section.

While every individual field of application has its own specific difficulties many of the issues are of very broad relevance in the observational and experimental sciences and their associated technologies.

This paper in particular sketches some of the areas of statistical research that are especially active at the moment; in most but not all there is parallel work in the econometric literature.

2. Criteria for Statistical Models

The analysis of data via tables and graphs is very important especially, but not only, for the presentation of conclusions. Nevertheless we concentrate here on methods with an explicit base in a probabilistic model. Much formal statistical discussion takes a family of models as given and develops the general concepts required for its analysis. Yet it seems clear that the initial choice of a family of models as a basis for interpretation is critical and this is harder to discuss in general terms. Much more is typically required than that the model provides a set of probability distributions yielding an adequate fit to the data under analysis.

Some desiderata for consideration are as follows:
1. the model should incorporate information from previous similar studies, allowing testing of consistency with these;
2. more specifically the model should use qualitative and quantitative ideas from subject-matter background and theory and allow testing consistency with these;
3. the model should be suggestive of a possible data generation process;
4. the model should contain parameters of interest that individually have clear substantive interpretation;
5. the "error structure" should allow realistic assessment of precision of the primary conclusions;
6. the "error structure" should take suitable account of any peculiarities of the data collection process;
7. the model should be consistent with the data in key respects.

3. Specificity of Objectives

An important general issue arises when the objective is a specific one of relatively short-term forecasting one or more quantities. Is it better to use some simple quite empirical technique, disregarding some of the more ambitious desiderata listed in Section 2, or should a deeper analysis be attempted, based in some sense on understanding the system under study? Empirical evidence on forecasting performance, for example from the so-called forecasting competitions (Fildes and Makridakis, 1995) is ambivalent. Presumably the
longer the time horizon the greater the preference for the second kind of approach.

At a more technical statistical level it is common to distinguish between estimation of unknown parameters and prediction of as yet unobserved random variables, the latter being relatively understudied. In Bayesian formulations the distinction at one level disappears, all unknowns being treated as random variables. At another level, in terms of the kind of model formulation to be adopted explicitly or implicitly, the broader issue remains.

4. Modes of Inference

Over the last 100 years and more there has been much discussion of the nature of probability and the role of Bayes's theorem in statistical inference. Bayesian approaches are of three very different kinds, the entirely uncontroversial use of empirical Bayes methods to represent a large number of essentially similar parameters via their frequency distribution, secondly by the use of standardized reference priors and thirdly the explicitly personalistic approach.

Implementation of these requires integration often in a large number of dimensions. A major technical advance recently has been the implementation of Markov Chain Monte Carlo (MC-squared) methods introduced originally in physics. For some applications in the context of financial time series, see Shephard (1996). The methods are applicable also to various non-Bayesian likelihood based problems.

The use of reference priors goes back at least to Laplace and more recently to the systematic account by Jeffreys (1961), whose work derives to some extent from Keynes's doctoral thesis. For recent developments, see Bernardo (1979) and Berger and Bernardo (1992). For many relatively standard estimation procedures the differences from neo-Fisherian likelihood-based inference are fairly minor. Recent research in this has concentrated on higher order asymptotic theory and the study of modified likelihood functions (Barndorff-Nielsen and Cox, 1994). There seems to be scope for fruitful investigation of the relation between higher order asymptotics, Markov Chain Monte Carlo, and the so-called bootstrap and on the implications of all these for more complex econometric techniques such as cointegration. The issues can be critical whenever the number of nuisance parameters is relatively large.

The explicit introduction of additional information via a personalistic prior density raises much more challenging conceptual considerations and cannot be discussed adequately here.

One general comment is that issues about mode of inference within a given formulation may often be less critical than the choice of model itself.
5. **Statistical Decision Theory**

The previous discussion in effect presupposes that the object of analysis is the interpretation of data, usually attaching some indication of precision to estimates and predictions. There is another strand to statistical thinking that emphasizes decision making. The essence was set out in a remarkable early paper by Neyman and Pearson (1933), motivated probably in part by industrial inspection problems. The book by Wald (1950) was highly influential for a period leading to claims in theoretical circles that all formal statistical issues should be viewed in decision-theoretic terms. While vestiges of this view remain, and the notion of clarifying the purposes of analysis is clearly important, many statisticians probably see statistical decision theory as a relatively small part of their subject.

Wald, who had a strong interest in economics, supposed that utilities are clearly defined but prior probability enters his analysis only as a technical device for producing complete classes of admissible decision functions. There is nowadays a fairly broad agreement that the separate problems in determining prior probabilities and utilities are related and similar and that a satisfactory quantitative treatment requires a fully Bayesian approach.

Perhaps the main difference between the treatments of decision problems in the economic and in the statistical literatures is that mathematical economists, although perhaps not econometricians, seem to relate utility strongly, if not exclusively, to money, often in effect if not in principle linearly, whereas statistical discussions do so less, thereby to some extent reducing the concept of utility to circularity.

6. **Interpretation of Observational Studies**

In a sense a number of the desiderata raised above aim to mitigate the limitations on the interpretation of observational as opposed to experimental studies. Often this centres on the question: what is the substantive meaning of a regression coefficient (using regression in a very broad sense as concerned with the dependence of one or more response variables on one or more explanatory variables)?

Two perhaps rather extreme examples of overinterpretation of regression coefficients are the conclusion that each implementation of the death penalty in the US saved about seven lives and, secondly, inferences from a time series study of tobacco consumption, prices and advertising expenditure concerning the likely consequences of making tobacco advertising illegal.

While the mathematical interpretation of a regression coefficient is clear as the statistical analogue of a partial derivative, the difficulties in interpreting regression coefficients in terms of the effect of interventions in the system stem from a number of sources:

1. the "laws" governing the system may change, especially under large interventions, cf. the Lucas critique;
2. explanatory variables omitted from the regression system, and perhaps not measured, or even whose existence is not appreciated, may not change under intervention in the way that they have changed in the data under analysis;

3. other explanatory variables included in the regression equation, and therefore implicitly held fixed when the intervention variable changes, should themselves have imposed changes whose form may not be easily appreciated.

The three points are, of course, interrelated. The last is in a sense a technical error and can be largely corrected but the first two points are more worrying.

One approach to the clarification of complex relations is the use of graphical representations, stemming from the geneticist Sewell Wright's notion of path analysis, introduced into econometrics by H.O.A. Wold. Recent work on this has moved away from the notion of decomposing dependencies into components, via generalizations of the formula relating total to partial regression coefficients,

\[ \beta_{Y1} = \beta_{Y1,2} + \beta_{12,1} \beta_{21} \]

where \( Y \) is regressed on two explanatory variables \( X_1, X_2 \) and, for example, in Yule's notation, \( \beta_{Y1} \) is the total regression coefficient of \( Y \) on \( X_1 \) and \( \beta_{Y1,2} \) is the corresponding partial regression coefficient given \( X_2 \). This is a statistical generalization of the elementary calculus formula relating ordinary and partial derivatives.

Rather the emphasis is now on representing relatively complex systems of conditional independencies either in the context of probabilistic expert systems (Spiegelhalter et al, 1994) or for analysis of empirical data (Cox and Wermuth, 1996).

The interpretation of regression coefficients is often connected with issues of causal interpretation. Statisticians have tended to take a very cautious position over this, especially as regards observational studies, some taking the line that causal inference is possible only from randomized experiments. (Randomized experiments have, in my judgment, made a massive contribution to human welfare in connection with clinical trials but they too can raise major problems of interpretation). To a limited extent differences of opinion over causality are issues of definition. Experience does, however, suggest that the traditional view in epidemiology that risk factors and causal agents should be firmly distinguished is wise.

Bradford Hill (1965) gave eight conditions under which a causal interpretation of observational studies becomes more plausible. King, Verba and Keohane (1984) have discussed these from a political science viewpoint; see also Cox and Wermuth (1996, pp 219-226). It would be interesting to have a parallel econometric analysis. By contrast there is coming from the artificial intelligence literature a much more positive attitude (Pearl, 1995; Spirtes, Glymour and Scheines, 1993). These include the development of computer
programs that will determine causality even from single cross-sectional studies: Despite the considerable interest of these developments, the use of the word causality with its normal scientific connotation of some understanding of underlying process seems unwise. Similar remarks apply to the time series notion of Granger-causality; the special structures of conditional independence encapsulated in that notion can be of much interest but ambiguities of interpretation preclude the immediate inference of causality in any deep sense.

One of Bradford Hill's conditions concerned the possibility of drawing stronger inferences from so-called natural experiments. In his context these were typically massive interventions, such as natural or man-produced disasters, with such strong effects that anomalous behaviour following the intervention can safely be ascribed to that intervention and not to unobserved confounders. In econometrics the term is used more widely to cover situations in which, even though the effects involved may be small, there are reasonable grounds for assuming that something approaching randomization has taken place. This is typically expressed via plausible assumptions about instrumental variables. An interesting example is the careful analysis by Angrist and Krueger (1991) via highly aggregated data of the effect of duration of schooling on earnings in which the instrumental variable month of birth has an association with both with duration of schooling and earnings. It is argued that the association is evidence that duration of schooling.

Rosenbaum (1995) has given a careful account of the general problems of design, analysis and interpretation arising with observational studies.

7. Testing Goodness of Fit

Models are by their nature idealized and it is unreasonable to expect them to describe all aspects of the data. In very empirical approaches to statistical analysis consistency with the data is often regarded as of overwhelming importance but it seems doubtful if this is wise. Tests can be regarded as ones of

1. the adequacy of the underlying theory or formulation, failure to fit the data indicating a need to rethink the whole basis of analysis;
2. details of formulation, such as number of lags needed, failure indicating a need for relatively minor modification;
3. important aspects of error structure, affecting in a critical way the assessment of precision and the efficiency of estimation;
4. relatively minor aspects of error structure.

A difficulty common in some fields of application is the clash between statistical significance and substantive significance and in particular the possibility that an effect may be highly significant statistically but yet too small to be of substantive importance. With the possible exception of the analysis of some kinds of financial data this difficulty seems likely to be uncommon in econometrics, where the models fitted are often somewhat elaborate relative to
the amount of information in the data, so that any effect found statistically significant is likely to be of subject matter concern.

A general strategy is that all significant lack of fit is to be reported and any argument that the discrepancy is substantively unimportant explicitly stated and as far as feasible justified. The obligation to report any lack of fit found is protection against extreme arbitrariness in model choice.

Issues of goodness of fit are connected with what in the econometric literature, but not elsewhere, is called calibration. With complex systems it may be helpful to divide the response features, i.e. those which are not to be held fixed in the model, into those of direct interest, those which are of indirect interest and those which are unimportant or which are accepted as poorly represented in the proposed model. The first set are to be used in fitting, for example by some form of the generalized method of moments, and in testing goodness of fit. Appreciable lack of fit will not be acceptable. If the second set of features, not used in fitting, are well represented the model gains in general credibility. The fit of the third set of features is not expected to be good and this is not regarded as important. Of course the division depends totally on context.

These remarks are of most relevance for fairly complicated problems. A relatively simple example would be the fitting of a continuous time diffusion-like model to share prices. The first set of features would be the mean and covariance structure of returns over time spans of say one hour and longer. The second set of features might concern the extremes of the process. The third might be very short-term properties of the process on a time scale where individual transactions are relevant, calling for a different type of continuous time model in which the point process structure is recognized.

8. Sensitivity Analysis

Large computer models are now quite widely used in many fields certainly including economics and a number are mentioned in companion papers. Their use may be either to gain insight by exploring the behaviour of complex systems under idealized conditions or to develop predictions for policy decisions, typically by estimating the consequences of intervention. It is easy to feel considerably more comfortable with the former than with the latter, certainly if the latter is not preceded by the former!

By contrast with the analytical solution of simple models, whose value for developing qualitative understanding can be great, the computer models require the assignment of numerical values to a large number of unknown parameters (coefficients). Some numerical values may be derived from explicit statistical analysis of relevant data, others may be chosen based on relatively informal "historical" analysis while yet others may be assigned via informal expert judgment, which, if not derived from well-formulated evidence, can be particularly suspect.
In all cases, whether the model is deterministic or stochastic, study of the reliability of the conclusions to the numerical assumptions made, is highly desirable, indeed should be mandatory.

This raises some technical statistical issues in particular regarding the model as a multidimensional system to be explored by highly fractionated design (Sachs et al, 1989; Welch et al, 1992). It is an open issue as to how far dependence on assumptions is best explored in this way by designed sensitivity analysis and how far by so-called elicitation of priors for the unknowns, leading to a posterior distribution for the quantities of interest giving a composite measure of uncertainty. It seems to me that some element at least of the former is desirable.

9. Panel Data

A very interesting field in which there are parallel and largely complementary literatures) emanating from the econometric, the sociological, the biostatistical and the industrial reliability fields, is the analysis of panel data) taken broadly to mean the analysis of many short time series. Technically the statistical properties come from consideration of a large number of supposedly independent individuals rather than from the reliance on the ergodic properties of a single long series.

A special kind of panel data arises with data defined by point events such as the beginning and ending of periods of employment. There is a vast predominantly biostatistical literature on survival data, the situation where there is essentially a single interval under analysis, the focus being its dependence on explanatory variables; see, for example, Cox and Oakes (1984) and from a viewpoint emphasizing the connection with martingales (Andersen et al, 1993). The sociological literature, under the name event-history analysis (Blossfeld, Hamerle and Mayer, 1992; Blossfeld and Rohwer, 1995), emphasizes the analysis of short series of events possibly of different types. There are strong connections also with the study of Markov and semi-Markov processes with discrete states; see, for example, Janssen (1986). Lancaster (1990) has given a very interesting account from an econometric perspective.

10. Conclusion

The literatures on econometric and statistical techniques have a good deal of naturally arising overlap, with inevitably some substantial duplication and some differences of emphasis. Of more interest in many ways is the relation between the approaches to specific problems. The statistical aethos, superficially at least, puts more emphasis on problems of study design, on issues of measurement and data quality, on a desire for simple models and on a purely empirical approach, and moreover is strongly influenced, indirectly at least, by an interest in the experimental as well as the observational sciences. The emphasis on assessment of precision tends to lead to a cautious, sometimes
overcautious, attitude to the interpretation of data. The econometric tradition appears to involve more willingness to fit rather complex models to limited data. The present paper touches on just a few of the issues involved.

If there is a unifying theme to late 20th century applied mathematics it is nonlinearity. With the exception of the special models developed to study volatility most of the models discussed in the very applied econometric literature appear to be linear and it is perhaps in moving away from linear models that the most scope for methodological development lies.

I am very grateful to Dr Neil Shephard for helpful suggestions.

11. References


Received: July 1999