SCIENTIFIC STANDARDS IN ECONOMETRIC MODELING

by

Christopher A. Sims

Discussion Paper No. 82-160, April 1982

Prepared for presentation at the celebration of the 25th anniversary of the Rotterdam Econometrics Institute of Erasmus University, January 1982. This research was supported in part by NSF Grant SES-8112023.

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

SCIENTIFIC STANDARDS IN ECONOMETRIC MODELING

By Christopher A. Sims

The increasing scale, complexity, and practical success of econometric modelling in recent years requires a rethinking of its foundations. Econometricians have made do with a formal description of the nature and objectives of their work which relies too heavily on the example of the experimental sciences, and thereby gives an incomplete and misleading picture. As a result, we have shown occasional confusion in judging or setting standards for empirical work. Perhaps worse, we have left ourselves open to apparently devastating criticisms.

The criticisms I have in mind are of two general types. One, represented by Lucas in [12], attacks the claim of standard macro-econometric models that they are "structural", and suggests attempting to replace them with models which could truly make that claim. Another, represented by Freedman [4], attacks the claim of econometric models that they are accurate probability models for the data to which they are fit. There is some justice in both of these criticisms, but they have been overdrawn. To answer them, this paper re-examines what econometricians do and ought to do.

I will argue that econometricians distinguish between subjective and objective uncertainty by casual, implicit devices. Such distinctions are necessary, but by making them badly, or failing to acknowledge that we make them, we invite misuse of our work (of the kind Lucas attacks) or dismissal of our probabilistic models as unscientific (as Freedman does). To understand why we need such distinctions and how our problems in this respect differ from those of the experimental sciences, I find it helpful to consider the parallels not only between econometrics and experimental science, but also between econometrics and accounting.

1. Personal probability and "objective" probabilistic models.

The personalist interpretation of probability, as exposited in the writings of L.J. Savage, is convincing to me. If one has to take a decision in the form of choosing a value for d given information on X, and if this will produce gains $U(d,Z)$, where Z is not known when d is chosen, I do not doubt the conclusion that a reasonable man with strong computational capacities should behave as if he had a joint probability distribution for X and Z in mind and chose d to maximize $E[U(d,Z)|X]$, using his joint distribution for X and Z to compute the expectation. For an econometrician advising a policy-maker, d might be the size of a tax cut, X an array of available historical data, U the sign-reversed inflation rate, and Z a vector of influences on the inflation rate other than the tax cut. The conclusion then is that the policy action chosen should be one which can be obtained as minimizing the expected inflation rate conditional on the data and the tax cut, using a probability model for the joint behavior of X and Z.

However, even though this conclusion is often invoked to justify a Bayesian approach to inference, Bayesian methods in practice seldom work with a joint distribution for X and Z directly. Instead the common practice is to specify the joint distribution of X and Z in a form like

$$1) \qquad p(X,Z) = g(X,Z|b)h(b)db \ ,$$

where p is the joint probability density function (p.d.f.) for X and Z, b is called the parameter vector, h is called the prior p.d.f. for b, and g is the conditional p.d.f. for X,Z given b, or the "model". Despite the insistence of personalists (and Savage in particular) that all probability is personal probability, g and h are in practice treated quite differently. Methodological discussion proceeds as if g were objective, while h is subjective.

What do I mean by "objective" and "subjective" here? Simply that over some group of individuals, variations in personal joint p.d.f.'s for X and Z are to a close approximation entirely a matter of variations in h, with a fixed g. The definition is thus constructed from a personalist point of view; but it rests on comparisons of probability judgments across individuals.

That a situation might actually occur in which individuals with different $p(X,Z)$ functions share a common $g(X,Z|b)$ is especially plausible in examples from the experimental sciences. There the nature of the experiment in which the data, X, is gathered determines g. For example, many fire departments may need to decide whether to purchase a type of rope. The value of the rope to them may depend on its breaking strength Z. The variation across fire departments in subjective distributions of Z may be adequately captured by a vector of parameters b. A testing laboratory can then construct an experiment which will be useful to all the fire departments. Assuming the fire departments know the distribution of Z to have the form $g_0(Z|b)$, the laboratory can, by making a realistic test, measure a random variable X_1 with the p.d.f. $g_0(X|b)$. By randomizing properly, they will convince all the fire departments that the X_1 they measure is independent of Z, and even that a sequence X_i, $i=1,\ldots,n$ of such random variables are mutually independent. This leads to a convenient form for $p(X,Z)$ and to the possibility of many fire departments benefiting from a single set of experiments. The form of g is not only objective in the sense of being agreed upon across fire departments, it is in some sense physical as well. Argument among fire departments over g would probably concentrate on whether physical procedures followed in the experiments met certain criteria -- whether the experimental conditions were realistic in the required sense; whether the methods for randomizing the choice of rope samples were adequate. In this respect g is different from h, which would vary across fire departments according to the experience and prejudices of their decision-making authorities. Here it is clearly useful to call g an objective p.d.f. and h a subjective one. That they can both be interpreted at a deeper level as personal is for many purposes a pedantic point.

When statistical theory takes as given an objectively true probability model, its practicioners usually have in mind, I think, something like the situation just described. The model's truth is widely acceptable because the experimental procedure has been set up in a certain way. But in economics at least, probability models seldom if ever have this kind of firm foundation. Once we recognize this, we may be tempted to take the position that distinguishing objective and subjective probability in economics is senseless. For a non-Bayesian, this may suggest that there is no legitimate role for formal, probability-based statistical analysis -- decision-making should be forthrightly based on subjective judgment. For a personalist Bayesian, the same recognition might suggest that we ought to rely directly on assessing personal $p(X,Z)$ functions without bothering with the inappropriate decomposition into g and h.

This position goes too far. While basing the distinction between objective and subjective probability on the distinction between physically true probability models and other sources of uncertainty is legitimate in the experimental sciences, it is not the only reasonable basis for such a distinction. Any situation in which a number of people have $p(X,Z)$ functions which take the form (1), with h varying across people and g constant, or even in which this condition is approximately valid, makes a distinction between objective and subjective probability useful. The likelihood principle, which states that all that we need to know about the data is the shape of the likelihood function, follows from the existence of a common g, regardless of whether the common g arises as a physically true model or out of a common psychological makeup of the people among whom the results are shared.

Probability models can, therefore, be justified in economics, even though they do not have the same claim to objective truth as formally similar models in experimental sciences. Nonetheless, the different basis needed to justify econometric probability modeling has implications for practice we will explore later on in the paper.

2. Personal probability and non-probabilistic data analysis.

Having explained how probability modeling can be justified in economics, we now ask why so much reported analysis of data for economic decision-making proceeds without any explicit reference to a probability model. A great deal of business decision-making makes heavy use of data which has been gathered, aggregated, and manipulated by the procedures we call accounting. Though probabilistic methods have been used in some aspects of accounting in recent years, most accounting is not probabilistic.

At the level of the national economy we have the national income accounts and input-output tables, among other systems of regularly collected, related, aggregated data summaries. Though we often treat these as if they were raw data produced by nature, they are in fact themselves the result of a large scale analysis of data.

When we take account of the fact that there may be many decision-makers with related but not identical decision problems, the practice of making public reports of certain statistics, without an associated probability model, finds a rationale. One can imagine that in a group of individuals indexed by $i=1,\ldots,n$ all of the joint densities for X and Z, $p_i(X,Z)$, have the form

2) $$p_i(X,Z) = g_i(f(X),Z) \,,$$

where f maps a large array of data into a short array of statistics. The vector $f(X)$ is thus a sufficient statistic for each individual, though each individual's conditional p.d.f. for Z depends on $f(X)$ in a different way. There is no point in public efforts aimed at analyzing likelihoods or posterior distributions, but public efforts at computing $f(X)$ are worthwhile.

This example can be extended to allow g_i to depend on an individual-specific vector of statistics f_i as well as on f. Then f by itself would not be a sufficient statistic for any individual, but it would still be an appropriate subject for public computation efforts. One can go still further. The occurrence of a low-dimensional $f(X)$ in the otherwise different decision problems of many individuals need not arise only out of the existence of sufficient statistics. For example, if everyone has a loss function such that the optimal choice of d is $E[Z|X]$, and if everyone has a $p(X,Z)$ such that $E[Z|X]=f(X)$, then however else $p(X,Z)$ may differ across individuals, it is worthwhile to compute $f(X)$ publicly.

When, as in these examples, individuals' decision problems have common features which cannot be expressed as a common parametric probability model, it is evident that we should not expect statistical analysis to be reported in the form of parameter estimates and descriptions of likelihood functions. That accounting and related descriptive statistical activities have a long history of apparent usefulness in economics suggests that we might find such situations in economics.

3. A general explanation of why we publish analyses of data.

The point of the first section above was that, despite the fact that probability models in economics do not have the physical character of probability models in experimental sciences, they may provide a similar basis for useful exchange among individuals of the results of empirical research. What matters about a probability model is that it captures some aspects of similarity in the way different individuals view an uncertain situation.

The second section then gave two examples to show that the case of a common probability model across individuals, with differences in loss functions and priors across individuals, is not the only way to explain why individuals may be interested in each others' empirical work or public empirical work by specialists.

When we consider a collection of individuals indexed by i, $i=1, \ldots, n$, each of whom faces a decision problem requiring him to maximize his utility $U_i(d_i, Z)$ given data on X using a personal p.d.f. $p_i(X,Z)$, any aspect of the structure of this collection of problems which allows some of the computation to be shared across individuals is a reasonable grounds for public statistical analysis. Even the standard case of a common probability model, when it is used to invoke the "likelihood principle", in fact draws not only on the existence of the model, but also on implicit restrictions on the forms of U_i and the prior h_i. The p.d.f. $p_i(X,Z)$ is assumed to have the form

$$p_i(X,Z) = \quad f_i(Z|b)g(X|b)h_i(b)\, db \tag{4}$$

The likelihood principle states that the likelihood function, $g(X|b)$ is a complete summary of the data. However, X itself will usually consist of finitely many real numbers. To report $g(X|b)$ as a function of b, even if g is continuous in b, in general requires presenting a countable infinity of real numbers. It seems strange to think of this infinite collection of numbers as a summary of a finite collection. In practice, though, we usually know that the range of relevant U_i, f_i and h_i functions is limited, and that expected loss can be computed to very high accuracy from knowledge of, say, the first few moments of the normalized likelihood function or from its "shape" as revealed in a small plot. Thus the "report of the likelihood function" is actually a report of much less than the whole function. It is implicitly assumed that we know enough about U_i, f_i, and h_i to be sure that knowledge of, say, high order derivatives, high order moments, or high order terms in the Fourier series for the likelihood function is not necessary. Particularly when the parameter vector b is of high dimension relative to the amount of data, there may be important gains from using more information about U_i, f_i, and h_i. When U_i and f_i are the same across individuals, for example, the risk function, $R(d,b)=E[U(d,Z)|X,b]$ is more useful than the likelihood function and may, depending on the structure of g, f and U, have a form that is easier to summarize accurately than that of g. Even more likely to simplify reporting is the case where h_i is common across individuals, so that the posterior p.d.f. can replace the likelihood. Thus where the form of the p.d.f. $f(Z)$ for Z is essentially unknown, so that many parameters are required to characterize it, the likelihood might have many local maxima and be, therefore, complicated to describe. A prior common across individuals which, say, made non-smooth shapes for $f(Z)$ unlikely would probably greatly downweight many of the local maxima and generate a posterior which could be well described in terms of its shape in the neighborhood of a few a priori likely local maxima.

Taking this last example further, note that constancy of h_i across individuals is more than we need to get the result of the example. We could instead have $h_i(b)=h_0(b)h_{i1}(b)$, where the factor h_0 decreases as b implies greater non-smoothness for $f(Z|b)$ and h_{i1} depends little on smoothness. The prior then has an objective component, h_0, and we can report the product of this component of the prior with the likelihood without loss of information, obtaining again the same sort of simplification of the reporting problem.

4. Bibliographical remarks.

The ideas presented to this point are new at most in marginal ways, or in the pattern of emphasis given them. They have been described tersely, in recognition that many readers will find many aspects of them familiar. But it is important to note where they can be found set forth in more detail.

Savage [14] and de Finetti [3] describe the personalist approach to probability and decision theory. This paper's emphasis on finding and using aspects of personal probability distributions which are common across a group of individuals reflects my reading of the book by Arthur W. Burks [2]. While Burks's theory is designed to apply to the natural sciences and therefore does not confront some of the central difficulties of econometric research, it does in my view successfully shift attention from individual decision making to collective advancement of knowledge, without losing or contradicting the insights of the personalist approach to decision theory.

Savage himself noted [15,p.14] that in some situations there might be approximate similarities in probability distributions across individuals which could usefully be exploited. He did not, as far as I have been able to discover, point out that on the personalist view every application of the likelihood principle must be such a situation. Geisser [5] has pointed out the questions raised by the use of parameterized probability models by a personalist. Hildreth [7] discusses the reporting of results to "the public" from a Bayesian perspective. His ideas closely parallel those presented here, except that he does not distinguish the case of an experimental science, where probability models may legitimately be labeled "objective", from cases where similarities in subjective p.d.f.'s arise in other ways.

5. Can econometrics be scientific?

The claim that econometrics, because it uses probability models without the kind of objective foundation such models can have in experimental sciences, is unscientific, is certainly incorrect in at least one sense. As we have seen, similarities in the personal probability distributions of individuals can create a basis for exchange of statistical results which is formally like reporting of results in experimental science. Whatever one calls it, it can in principle be a useful activity.

Econometrics (and possibly some other non-experimental disciplines as well) does face special problems in setting professional standards for empirical work, however. The standards for setting up an experimental probability model -- use of controls, randomization methods, etc. -- and for reporting results have developed over many years. They reflect not just common elements of personal p.d.f.'s across a few dozen or a few hundred researchers, but long experience with use of experimental results for practical purposes by non-researchers. The personalist view seems to lead to the conclusion that any probability model which any person finds a priori plausible is as legitimate as any other, which allows no explanation of the fact that in experimental science there are rigid, useful rules for determining what is a scientifically legitimate probability model. Our explication of econometrics distinguishes models about which communication is useful from purely personal models, but does it leave us with any criterion for scientific legitimacy beyond the criterion that at least two people in the world must find the model interesting?

If econometric research were carried out or directly financed by decision-makers, there would be no reason to look for a standard for econometric research broader than the criterion that those who read it or finance it are interested in it. But in fact the audience for much econometric research is other econometricians; just as in experimental sciences, the use of research results by non-researchers is often remote in time or place from the presentation of the results. Without rigid objective standards, there is the chance that models will be analyzed and discussed because of their attractiveness as puzzles, because of the prejudices of a few professional economists, or because they have become conventional -- few claiming to take them seriously themselves, but many working out their implications because of presumed interest in them by others.

These dangers are not merely hypothetical. I have argued in detail elsewhere [17] that much applied work within the standard simultaneous equations framework has amounted to working out the implications of probability models which had become conventional, justifiable only because other people had used them. This situation arose because standard simultaneous equations methods seemed to require that, if a model with numerous equations was to have a well-behaved likelihood function, it had to be more heavily restricted a priori than any reliable knowledge allowed. Hence it became common (and remains common in some quarters) to invoke, in addition to those few restrictions with some substantive justification, an array of conventional restrictions to make the model manageable.

In my earlier paper [17] I displayed an example of an approach to analyzing the same type of data ordinarily used in macroeconometric business cycle models without the usual burden of conventional restrictions. This approach is a special case of a more general strategy which can be described in the terms of this paper. While we cannot generate the simple objectively acceptable models emerging from good experimental procedure, we do have in cross section work on samples of individuals the independence and identical distribution assumption, based on confidence in the randomization procedure used in sample selection. In time series we have a fuzzier notion which is more or less objective: the notion that dependence between events at different times should weaken as their separation in time increases, and that the form of the joint distribution of events dated t and $t-s$ should change only slowly with t. When we say this is objective, we mean only that it is a common characteristic of the prior p.d.f. of nearly everyone.

The procedures in [17] showed that a model which used only the assumptions that the form of the best linear predictor for a vector of time series was stable in time and that it involved only a fixed finite number of lags (4 in that paper) could produce estimates with interesting interpretations and provide a framework within which informative tests of substantive hypotheses were possible. I have taken a similar approach in applied work with distributed lag models, showing that substantively useful conclusions can emerge from models which restrict only lag length. A model with a fixed lag length, though, is clearly only a very crude approximation to reflecting the common prior belief that dependence decays with separation in time. Litterman [11] has gone a step further, imposing a prior distribution on the parameters of a multivariate linear model which implies that it is likely that dependence decays with lag, thus allowing a much longer maximum lag length in his model. Shiller [16], much earlier, displayed a procedure appropriate for situations where a common prior would have a lag distribution smooth in shape and tending toward zero with increasing lag length.

Procedures like Shiller's and Litterman's sometimes fall between stools in the profession. Bayesians find them too mechanical: they clearly use a prior p.d.f. which does not correspond directly to the prior p.d.f. of any individual; though labeled Bayesian, they may seem to encourage analysis of data without serious assessment of the individual's full prior p.d.f. Of course, the procedures are beyond the classical framework of a nominally objective parameterized model, and therefore are often regarded as suspect by non-Bayesians as well. They are, however, a good example of non-classical objective probability modeling. Lag distributions and multivariate linear prediction models (vector autoregressions, or VAR's) are examples of models with parameter vectors for which people's prior p.d.f.'s take the form $h_i = h_0 h_{i1}$, as in the last paragraph of section 3 above, so that analysis of the likelihood multiplied by h_0 is a more useful summary of the data than analysis of the raw likelihood. This is of course equivalent to a "Bayesian" analysis with h_0 as prior, but it should be clear that a Bayesian should no more object to such a procedure than to a presentation of the likelihood function in a classical model. The fact that h_0 is not the prior of any one individual does not prevent the posterior based on that prior from being a useful way to summarize data. In particular, presentation of the peak of the product of h_0 with the likelihood, together with descriptions of the shape of this function near its peak, is likely to be much more widely useful than the corresponding display of information for the likelihood, which is what classical estimation would produce. The standard alternative to use of a common h_0 in a Bayesian analysis would be to shrink the number of parameters while preserving fit -- parsimonious parameterization. Such a procedure, paradoxically less objective than the Bayesian one, is discussed in section 6 below.

In both Shiller's and Litterman's work the standardized common prior h_0, while spread over a large finite-dimensional space, is most naturally thought of as approximating a prior spread over an infinite dimensional space. Lindley [10] expresses some puzzlement as to why there has been so little progress in developing practically useful ways to specify distributions on infinite-dimensional parameter spaces.[1] As I have noted elsewhere [18], there are deep difficulties in spreading a prior "smoothly" over an infinite-dimensional space. One way to describe the problem is to note first that in any infinite-dimensional space with a linear structure, compact sets are nowhere dense. That is, the complement of every compact set has a closure which is the whole space. Compact sets are thus in one intuitive sense small. Yet it can also be shown that any probability measure on such a space which defines the probability of every open set puts probability one on some countable union of compact sets. Thus any prior on an infinite-dimensional space puts probability zero on "most of the space" in a certain sense.

Similar formal difficulties arise with probability densities on the real line. A probability density on the real line which vanishes nowhere nonetheless gives probability one to a countable union of nowhere dense sets. But on the real line we have a natural criterion to use for what a "small" set is which can replace the notion of nowhere-denseness. Lebesgue measure, uniquely defined by the property that the measure of $S+x$ is the same as the measure of S, where S is a subset of the real line R and x is a point in R, can be used to define sets of Lebesgue measure zero as small. But in an infinite dimensional linear space, there is no measure which has Lebesgue measure's property of translation-invariance. In fact, there is not even any which has the property that if $P[S]$ is non-zero, $P[S+x]$ is also non-zero. No matter what prior we put on such a space, we rule out as a priori impossible some set which someone else, whose prior differs from ours only in its "location parameter", gives positive probability.

I believe these apparently abstract difficulties with putting priors on infinite-dimensional spaces account for people's not having been satisfied with certain simple practical solutions to the problem. For example, in a parameter space of all absolutely summable lag distributions b, where b is a real-valued function on the non-negative integers, one could propose a prior which puts probability 2^{-k} on the finite-dimensional subspace of finite-order lag distributions of length k, with some convenient density, say a Shiller prior, within each k-dimensional subspace. If we use the sum of absolute differences as our metric on this space, this prior does put positive probability on every neighborhood of every point in the space. But it has the glaring deficiency that it puts probability zero on the set of all truly infinite lag distributions, i.e. those with $b(s)$ nonzero for every s. The general results cited in the preceding paragraph tell us that this kind of thing will crop up in every probability distribution we try to specify over the b's.

That these problems are unavoidable is in a way discouraging, of course, but it is also liberating in another way. Since any approach to putting a prior on an infinite-dimensional space must put low probability on "large" sets, simple approaches to specifying such priors ought not to be discarded on the grounds that they seem to fail to spread probability smoothly over the space. Every prior will fail on this score. The general results tell us that we will always face the possibility that some people, with priors which are not unreasonable on formal grounds, will interpret evidence differently from the way we do, even when the evidence is very strong. But we may hope that in practice, the important implications of the data, which may depend on relatively few functions of the infinite-dimensional parameter, will emerge as similar even for priors which differ sharply on certain sets of parameter values. It is possible to show [18] that conclusions from two priors which both put positive probability on all open sets must differ more and more infrequently with increasing sample size, under reasonable regularity conditions.

Though it has been applied in econometrics mainly to time series, the idea of approaching data with an infinite-dimensional parameter space and a common prior h_0 can apply as well outside time series, and to aspects of models other than serial dependence within time series. General nonlinear regression models of the form $E[y|X]=f(X)$ can be set up by putting a smoothness prior over an infinite-dimensional space of candidate f's. For the case of bivariate regression such methods have attracted some attention from statisticians, and an example of a computationally tractable approach appears in Ansley and Wecker

[1]. Their prior gives f the distribution of a continuous-time random walk, or Weiner process. A possibly simpler approach, which, according to the general result on lumpiness of priors on infinite dimensional spaces is not more restrictive, is to put a distribution on the coefficients in a series expansion for f, say its Taylor expansion, using the same sort of approach which would apply to putting a prior on a lag distribution. This would lead in practice to fitting polynomial regression models of high order, with priors, much like Litterman's, specifying that it is likely that higher-order coefficients are smaller. The form of the p.d.f. of the residuals could also be given an infinite-dimensional parameterization along these lines.

In dealing with large numbers of variables, one can treat the variable list as unboundedly long in principle and use a prior which gives stronger a priori weight to simpler patterns of interaction among them. Factor analysis and its extension to time series (see Geweke [6]) can be interpreted as approximations to such an approach, but there is no work I know of implementing an explicitly Bayesian approach. The obvious approach of putting prior probabilities on k-factor models which decline with k might be feasible. However, since in economic applications k has usually been taken quite small, there may be little to be gained from an explicitly Bayesian approach.

Another approach to highly multivariate modeling is to give higher prior probability to models which show near block-recursive structure. The observed behavior of econometricians, whose large-scale models are usually built on strong strict exogeneity and predeterminedness assumptions, seems to indicate that this approach is plausible to many researchers. Recursive structures with a pre-specified partial ordering on the variables can be described with Gaussian priors (unlike prior beliefs that high-indexed factors are likely to be small), so they can be given approximate implementation with mixed estimation methods. Litterman [11] describes, in his "circle-star" prior, a way of specifying an h_0 giving higher prior probability to more block-recursive forms of a multivariate time series model.

To answer the question which forms the title to this section, econometrics can at least be much more scientific if it grounds its models more closely on the aspects of prior beliefs which economists and users of our analyses really do have in common. Models constructed on such a criterion will be "overparameterized" by conventional standards, but they can in many applications be usefully analyzed by forming and describing the posterior distribution they generate when used with a common prior. Modeling methods of this sort are not computationally difficult, they are not hard to interpret, and they yield results of substantive value. I expect they will become standard in econometric work.

6. The role of false assumptions.

Even the most complete implementations of these ideas about econometric method have used finite lists of variables, finite lag lengths, and, in generating posterior distributions, normality assumptions. They have also used stationarity assumptions. We know these assumptions are not exactly true. Why is there any difference in principle between these arbitrary assumptions and the arbitrary assumptions made in standard simultaneous equations modeling?

Though we may think of ourselves as in principle having an infinite dimensional parameter space indexing functional form, the p.d.f. of the residuals, degree of cross-variable interaction, parameter variation over time, and decay of dependence between observed data points with separation in time, the analytical complexities of dealing with all these matters at once are at least for now too much for practical data analysis. Use of simplifying assumptions amounts to examining the likelihood or posterior distribution over submanifolds of the larger parameter space. There is certainly no objection in principle to such exploration.

To be useful, however, examination of simplified models should meet two criteria: the models should fit reasonably well and the analysis should yield numerical results which are a well-defined, reproducible function of the data. The example of accounting is relevant here. Even estimates of a bad-fitting model may be of some use if treated as descriptive statistics, the way accounting data are. But a set of accounts is much less valuable if we have no objective, reliable description of the methods used in putting it together. Procedures which involve experimenting with large numbers of potential restrictions, retaining only those which give the best-looking results, and reporting only results of estimating the restricted model, fail by the second criterion. Much econometric work with large-scale models is of this type, imposing restrictions in reaction to the data by rules of thumb and intuitive judgment. The resulting reported estimates are ill-defined functions of the data, and hence of limited use to other researchers. The arbitrary simplifying restrictions imposed in use of Litterman's methods are less complicated and more explicit functions of the data than those in most standard large macroeconometric models. Thus, for example, lag length, which is always given a finite maximum in vector autoregressive models, is often set at a fixed large number at the start when a prior is used to damp distant lags. Or, if lag length is tested, it is generally tested for all equations jointly, with only a few alternative lengths examined and test statistics for all of them reported.

If a simplified model is to be taken as more than a scheme for generating descriptive statistics, but rather as a representative plausible probability model for the data, it must fit well. There can be no objective standard of good fit, but description of how high the likelihood is on this simplified model's manifold relative to other nearby parameter values and relative to the regions of highest posterior probability using a reasonable h_0 are useful. This kind of information is generated by specification error tests. Such tests are made convenient if the simplified model is constructed as a special case of a more densely parameterized model, or if the simple model can be expanded, by the addition of extra parameters, to a more plausible form. In either case the usual apparatus of likelihood ratios, Wald statistics, and Lagrange multiplier tests will provide information on fit. Even when nesting of the restricted model in a more general one is not convenient, comparison of the restricted and more general models for the data, according to some standard measures (like, say, first and second moments) can give an indication of whether the restricted model is missing something. Franz Palm [13] has recently shown that one can, starting from the standard simultaneous equations methodology as a base, go a long way toward meeting these suggested standards by careful reporting of procedures and specification testing.

7. Pitfalls of treating simple models as true.

Though for most purposes good practice in estimating and testing a simplified model will lead to useful analysis whether or not the abstract perspective of this paper is accepted, this is not true in every instance. It may require some care to avoid drawing conclusions which rest critically on arbitrary simplifying assumptions, though I think most economists are careful in this respect. Thus in an Almon polynomial lag disribution a test of an hypothesis on the sum of coefficients may be equivalent to a test on a single coefficient, for example, and most econometricians avoid the fallacy of treating such tests as they would tests of the same hypothesis in an unconstrained model. A more subtle case arises in the analysis of multivariate time series models. It is often suggested (see Palm [13] and other references he cites on this subject) that a check on the validity of a multivariate model is to compute the univariate model it implies for each variable and to check whether the implied model is compatible with a direct empirical estimate of a univariate model for the variable. Put in this general way, this is a reasonable suggestion. A good way to implement it would be to compare the fit of the implied univariate model to that of the directly estimated univariate model, using a least-squares criterion. Another good approach would be to compute and compare the moving average representations of the implied and directly estimated models. These comparisons are both ways of examining part of the likelihood, of checking fit, and neither method would be likely to suggest an incompatibility if both univariate and multivariate model fit well. However, suppose the multivariate model was an unrestricted fourth order autoregression in six variables. The order of the autoregression would have been chosen empirically, as a simplifying approximation. But if taken literally, the multivariate model would imply that each univariate model should be an ARMA(24,20) model. It is unlikely thata reasonable approach to finding a simple univariate empirical model would emerge with an ARMA(24,20) form for all or even most of the variables in the system. Finding that most of the variables could be fit very well with an ARMA(5,2), say, would not be grounds for concluding that the multivariate model was badly specified, as it is quite possible that those ARMA(5,2) models might fit about as well and imply about the same MAR as the ARMA(24,20) models derived from the multivariate model.

This last example illustrates a general point. Simplifying assumptions are most useful when they eliminate a dimension of varriation in the parameter vector about which the data tell us little in any case. That is, the simplifying assumption fixes our position along a dimension of the likelihood surface in which it is close to flat. The danger then is that we forget, or fail to notice, that this is what is going on, and therefore proceed as if the data give us sharp answers to questions which depend on this dimension of parameter variation. While astute informal reasoning will often save us from such errors, it remains useful to study infinite-dimensional estimation problems explicitly, so that a clear understanding of what functions of the parameter will be well-determined by the data can emerge.

In a univariate time series model, the likelihood depends on the sum of squared residual errors in the model's autoregressive representation. It is natural to take two models to be "close" if they give nearly the same expected squared residual forecast error in this equation. Let b represent the parameterization of the model, and define $d(b_1, b_2)$ as the expected squared forecast error when b_1 is used to generate predictions, minus that when b_2 is used, assuming b_2 is the true model. Then using arguments just parallel to those in [19], one can easily show that for b_2's implying an everywhere non-zero, finite spectral density, $d(b_1, b_2) \to 0$ if and only if $\|a(b_1)-a(b_2)\| \to 0$, where $a(b)$ is the sequence of autoregressive coefficients implied by b and $\| \ \|$ is the sum of squared deviations norm. The orders of the AR and MA components in finite-order ARMA models are functions of the parameter vector which are $\| \ \|$-discontinuous. Thus the data will never give us strong information about such functions without powerful auxiliary assumptions restricting the parameter space.

Some econometricians find the resort to infinite-dimensional parameter spaces inherently repugnant. Most points which are practically relevent which can be made elegantly in such a parameter space can be made with a little more effort by use of finite-dimensional examples. In the example at hand, the point is that an ARMA(24,20) model can give nearly the same forecasts as a model of much lower order. It is obvious that this can happen if the roots of the numerator and denominator polynomials in the lag operator in the ARMA model approximately cancel. But, because the roots themselves are discontinuous functions of the coefficients, in the squared-sum norm, the approximate cancellation is not necessary to the results. Thus the zero-order identity polynomial in the lag operator can be arbitrarily well approximated in the relevant sense by polynomials with an arbitrary root of, say, 2. For example:

$$(1-.5L)(1+.5L+.25L^2+.125L^3+.0625L^4).$$

This polynomial is very close the identity, being exactly $1-.03125L^5$, and would give close to optimal results as an autoregressive forecaster of pure white noise, though it has five roots, all of absolute value 2. A model which makes the simplifying assumption of a low-order finite parameterization may make the roots of the polynomials in an ARMA model appear to be sharply determined by the data, but they are not, unless the simplifying assumptions are more than that, having a grounding in substantive a priori knowledge.

Another important error that is made through taking simple models as literally correct is the fallacy of "evaluating" models through their forecasting, or "out of sample" performance. There is no doubt that if a model is to be used for forecasting, the best way to evaluate it and compare it to other models is to compare the models' forecasting performance. And it is true that formal and informal "overfitting" to the sample period can occur, so that out of sample tests of forecasting performance are better than in-sample tests. Indeed, Geisser [5] makes a strong argument that our inference should focus much more on such direct measures of the performance of a model in the use for which it is intended, instead of on the model's "parameters." However, comparisons of forecasting performance are no more than that; they are not measures of models' relative closeness to the truth if all the models are simplified approximations. There are applications, including all applications in which we are trying to discover things about the structure of the economy, where we really are interested in the model's parameters, not directly in its forecasting performance. It is by now well known that forecasting performance can be improved by imposition of false assumptions, if the assumptions are not "too false." A densely parameterized model will not ordinarily produce good forecasts if estimated by maximum likelihood or related methods. This does not mean that "nature has few parameters." It means only that when the data contain little information, one should forecast either with explicitly Bayesian methods or with simplified approximate models. Taking the simplified models as true can lead to serious error.

This last point may seem obvious, but, having heard good econometricians express incredulity at the idea that forecasting performance is not an appropriate criterion for the best model, I will provide a simple example. Suppose we have historical data on money supply growth m and on the deficit b. We assume these latter two variables have a structural relation to the inflation rate y which we can estimate by linear regression of y on m and b. The truth is that m and b each have a unit coefficient in the regression. Let m and b have a correlation of .99 and unit variance. Let the residual error in a regression of y on m and b have unit variance as well. Then, if m and b retain the same covariance matrix out of sample, the mean square error of out-of-sample forecasts will be $1+2/T$, where T is sample size. If we simplify the model by constraining the coefficient on b to be zero, then mean square forecast error will be $1.0199(1+1/T)$. For sample sizes below 50, the constrained model will provide better forecasts. Yet we might be most interested in a non-forecasting use of the model -- e.g. using it to decide whether we can expect to control inflation using money stock alone while letting the deficit b be very large. Clearly the unconstrained model will give us a much better answer to this question, despite its poorer out of sample forecasting performance. Furthermore, if we used a Bayesian procedure to generate forecasts from the unconstrained model and to gauge the effects of deficits on inflation, we would be likely to get good results from the unconstrained model for both purposes. Because of the strong collinearity, even a weak prior favoring a model with, say, only m entering the equation would produce point estimates similar to those of the simple model constrained to eliminate effects of b, yet the posterior p.d.f. would show clearly that large effects of b were not highly unlikely.

8. Structural models.

Econometricians use the word "structural" in several ways. Sometimes it is used as if a structural model and the "structural form" of a standard simultaneous equations model were the same thing. A better definition, I think, is that a structural model is one which remains invariant under some specified class of hypothetical interventions, and hence is useful in predicting the effects of such interventions. This definition has a long history, which is explored in more detail in [20].

Whether a model is structural depends on the use to which it is to be put -- on what class of interventions is being considered. Most econometricians would admit that loosely restricted multivariate time series models, though often labeled non-structural, are for practical purposes structural when the object is forecasting. For forecasting, the relevant "intervention" is simply advancing the date.

Loosely restricted multivariate time series models include reduced forms of simultaneous equations models as a special case, and, paradoxically, most use of simultaneous equations models for policy analysis in fact treats the reduced form as structural! That is, the usual way to use a simultaneous equation model to project the effect of a policy change is to characterize the policy change as a certain time path for a policy variable, which has in the model estimation been treated as predetermined, and then to use the reduced form to trace out the effects of this path for the policy variable on other variables. This means precisely that the reduced form is treated as invariant under the change in policy, i.e. as structural. Loosely restricted multivariate time series models can be used in exactly the same way to project the effects of policy. They cannot properly be criticized as non-structural, in the sense of not being useful for policy evaluation, by people who would use standard simultaneous equations models for policy evaluation in the usual way.

Another line of criticism of loosely restricted models is the claim that they represent mindless churning of the data. Ever since Koopmans's classic essay [8], economists have, sometimes mindlessly, insisted that "measurement without theory" is valueless. This is of course unarguably true, if broadly interpreted -- if one approaches the data with no questions in mind, one will obtain no answers. But the notion that good empirical work must involve confronting the data with a model which allows the data to answer only a narrow range of questions, i.e. with a heavily restricted model, is quite incorrect. I believe it can be argued that the most influential empirical work in economics has historically been quite "unstructured", asking the data relatively vaguely specified classes of questions and leaving it to tell the story of such regularities as were actually present. I would put Milton Friedman's statistical work on the relation of money and income in this category, as well as Simon Kuznets's work on patterns of economic growth. This issue will never be resolved, partly because it seems that people's attitude toward it is as much a function of their personality structure as of any rational argument. There is not space to pursue this question further here, though to close the discussion it might be noted that Thomas Kuhn [9] observes an important phase of "measurement without theory" in the early modern history of the physical sciences. He notes cultural and psychological patterns of division between empiricist and theoretician which may be interesting to modern economists.

Of course Lucas's persuasive paper [12] has convinced many economists that the taking of reduced forms as structural in policy evaluation is a useless procedure. He suggests that instead we should estimate the parameters determining private sector behavior as a function of the "policy rule"; he argues that when we have done so we will find that the reduced form of our model will have changed in response to a change in policy rule. The stochastic properties of the new and the old reduced forms should be compared to find the effects of policy. Just as is standard simultaneous equations methodology, the rational expectations methodology which attempts to meet Lucas's critique is not contradictory to loosely restricted empirical modeling. It does not imply that such models are false descriptions of historical data, only that the interpretation of them to yield policy prescriptions should be different from that implicit in standard econometric policy evaluation. In principle, structural parameters are functions of reduced form parameters, both in standard simultaneous equations methodology and in rational expectations methodology. Under either approach, loosely restricted time series models may provide standards of fit and a descriptive guide to formulation of good simple models whose parameters are structural.

This conciliatory note would be a comfortable one to end on. However, my view is that the rational expectations critique of econometric policy evaluation has sent the profession down a false trail. The major defects in standard econometric policy evaluation had been that it took insufficient account of policy endogeneity, and that, in exericises applying optimal control theory, it was claiming to predict the effects of policies which lay far outside the pattern of historical experience. Its practical applications had in fact largely avoided the latter of these criticisms, since it was (and still is) ordinarily used to extrapolate the effects of policy paths which are not historically unprecedented into the immediate future. The rational expectations critique provided one intellectually appealing example to illustrate how in economics nearly any claim to have found a probability model which can objectively be claimed to be structural under a drastic policy intervention is likely to prove false. But the positive program of rational expectations econometrics, to estimate identified, structural models to be used in predicting the effects of change in policy rules while taking account of induced changes in expectational mechanisms, reproduces the main faults of standard econometric policy evaluation in exaggerated form.

Standard methodology took historical time variation in policy variables as statistically exogenous and extrapolated the effects of time paths for these variables which looked somewhat like the historical data for them. The rational expectations program focuses on the parameters of the policy rule, which are taken to have changed hardly ever, or even never, in the historical data, and presumes to extrapolate the effects of once-for-all changes in them. Of course a constant is an extreme form of an exogenous variable, and a once-for-all change in a constant is always historically unprecedented. The data cannot be expected to tell us much about the effects of such interventions. One can certainly construct models consistent with the data in which the parameters are interpreted as structural relative to such interventions, but the interpretation is bound to be controversial and there will be slim objective basis for probability models of historical data which can resolve the controversy.

Shifts in expectational mechanisms are not the only, nor even the most important, source of increased uncertainty about model projections when drastic interventions are contemplated. The recent rise of monetarist views among policy makers in the U.S. and U.K., for example, has brought home the lesson that the pervasive practice of using one variable to proxy for several which move with it (e.g., one "money") in econometric modeling, a practice which is pervasive in rational expecations as well as standard models, is a source of great uncertainty about specification when the variable serving as a proxy becomes a special focus of policy concern.

But if the positive econometric program of the rational expectations school should not be the central business of econometrics, what should we be doing instead? Continue to do what, in a rough way, econometricians providing actual policy advice have been doing: evaluate the likelihood of various proposed policy scenarios for the immediate future; warn, when it is appropriate, that policy changes which really are historically unprecedented have very uncertain consequences. There is much room for improvement in these procedures -- taking account of policy endogeneity, avoiding telling misleading stories about the interpretation of model coefficients, formalizing statistical procedures so better scientific communication among researchers is possible. But, by construction, policy formulation must most of the time not involve historically unprecedented changes in rule, and econometricians should not proceed under the impression either that analysis of the data has nothing to contribute to the normal formulation of policy or that it has a great deal to contribute to speculation about what will happen if a millenial rule change does take place.[2]

What about the argument that this style of econometric policy evaluation leads to myopia, with short-run gains displayed by the econometric projections continually being pursued while long-run losses snowball? This argument applies if econometric projections are misused. It applies to control-theoretic solutions for optimal policy rules. But it is certainly possible to place more credence in a model's near-term than in its long-term projections. It is certainly possible to avoid policy choices which are projected to produce results which one recognizes as possibly producing undesired changes in model structure. To be specific, there is no reason we cannot give special weight to inflation in evaluating projected policy effects because of concern that persistent high inflation may shift the linear structure of the model so that the real costs of reducing inflation would become higher. Such a way of using econometric projections would avoid the simple fallacies of myopia, and would not make good projections of near-term consequences of policy any the less useful. In fact, it seems to me that actual historical use of econometric policy projections has for the most part avoided the myopic fallacy of which the rational expectations school accuses it in just this common-sensical way.

9. Conclusion

Economists are not physical scientists. Despite the way we sometimes talk and write, we do not estimate parameters which define the truth. If we think carefully about what we are doing, we will emerge, I think, both more confident that much of applied econometrics is useful, despite its differences from physical science, and more ready to adapt our language and methods to reflect what we are actually doing. The result will be econometrics which is more scientific, if less superficially similar to statistical methods used in experimental sciences.

FOOTNOTES

1. He also cites work on Bayesian approaches to spectral estimation by Whittle [21] which sounds similar in spirit to Shiller's work on estimating smooth lag distributions.

2. Economists of the rational expectations school sometimes argue that in fact changes in the rule will occur as a slow drift over time, e.g. with steadily increasing stability of the money supply as monetarist arguments are increasingly accepted. Then, it is argued, the proper role of economists is to determine the proper direction of drift. If drift were going to be unidirectional, and toward a fixed limit, this argument might make some sense. In fact, it appears that drift is far from unidirectional and is far from being determined mainly by the intellectual activities of economists. Drift in policy rules is endogenous and endemic. It is one source among many of drift in the probability structure of the economy. Its existence is one reason why, even when governments claim to be making once-for-all changes in rules, extrapolations for the near term (one or two years) are likely to be made more reliably by taking autoregressive structure as constant, identifying policy interventions as patterns of disturbance to variables in the system, than by attempting to use a rational expectations structure and taking the government's announcement of a permanent change seriously.

REFERENCES

[1] Ansley, Craig and William F. Wecker, "The Signal Extraction Approach to Linear and Nonlinear Regression Problems," mimeo, University of Chicago, 1980.

[2] Burks, Arthur W., Chance, Cause, Reason: An Inquiry into the Nature of Scientific Evidence, Chicago and London: University of Chicago Press, 1977.

[3] de Finetti, Bruno, Probability, Induction and Statistics, New York: Wiley, 1972.

[4] Freedman, David. (to be provided)

[5] Geisser, Seymour, "A Predictivist Primer," in Bayesian Analysis in Econometrics and Statistics, Arnold Zellner, ed., Amsterdam: North-Holland, 1980.

[6] Geweke, John, "The Dynamic Factor Analysis of Economic Time Series," in Latent Variables in Socio-Economic Models, D. Aigner and A. Goldberger, eds., Amsterdam: North-Holland, 1975.

[7] Hildreth, Clifford, "Bayesian Statisticians and Remote Clients," Econometrica, 31 (1963), p. 422-439.

[8] Koopmans, Tjalling, "Measurement Without Theory," Review of Economics and Statistics, 29(1947), 161-172.

[9] Kuhn, Thomas S., "Mathematical versus Experimental Traditions in the Development of Physical Science," Journal of Interdisciplinary History, 7(1976), 1-31. Reprinted in The Essential Tension, Chicago and London: University of Chicago Press, 1977.

[10] Lindley, D.V., Bayesian Statistics, A Review, Philadelphia: SIAM, 1972.

[11] Litterman, Robert L., "A Bayesian Procedure for Forecasting with Vector Autoregressions," forthcoming, Journal of Econometrics.

[12] Lucas, Robert E., "Macro-economic Policy Making: A Critique," Journal of Monetary Economics, 1975.

[13] Palm, Franz, "Structural Econometric Modeling and Time Series Analysis: An Integrated Approach," unpublished. Free University, Amsterdam, 1981.

[14] Savage, Leonard G., The Writings of Leonard Jimmie Savage: A Memorial Selection, Washington, D.C.: American Statistical Association and Institute of Mathematical Statistics, 1981.

[15] Savage, Leonard G., "The Shifting Foundations of Statistics," in Logic, Laws and Life, R. Colodny, ed., Pittsburgh: University of Pittsburgh Press, 1977. Reprinted in [14].

[16] Shiller, Robert, "A Distributed Lag Estimator Derived from Smoothness Priors," Econometrica, 41 (1973), 775-788.

[17] Sims, C.A., "Macroeconomics and Reality," Econometrica, 48 (1980).

[18] __________, Annals of Mathematical Statistics, 1972.

[19] __________, "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," Journal of the American Statistical Association, 1972.

[20] __________, "Exogeneity and Causal Orderings in Macroeconomic Models," in New Methods in Business Cycle Research, Federal Reserve Bank of Minneapolis, 1977.

[21] Whittle, P., "Curve and Periodgram Smoothing," Journal of the Royal Statistical Society, B, 19 (1957).