Econometrics for Policy Analysis: Progress and Regress

Christopher A. Sims (sims@princeton.edu)
Princeton University

Abstract. Progress, regress, and continuity in quantitative analysis for policy is discussed. The approach taken is to look at the present from the perspective of Tinbergen, Haavelmo, and Keynes. It is argued that probability modeling has been in retreat at central banks and elsewhere. New computational methods, though, are making Bayesian analysis of previously intractable problems possible, and at the same time appreciation of the clarity with which Bayesian data analyis integrates with decison-making is spreading.

Keywords: econometrics, methodology, policy

Keywords: econometrics, methodology, policy

A few years ago I wrote "Macroeconomics and Methodology" (1996). While I will not rehash here the views set forth in that paper, those views are the starting point for this paper. This paper discusses and draws connections among three sources:

1. Visits to central banks and interviews I conducted with people in those banks in 2002, assessing the models they use and how they use them; [1].
2. Recent technical developments that have converted theoretical advantages of Bayesian over classical approaches to inference into practical reality in some applied areas; associated applied work and methodological commentary emerging in the literature;
3. Haavelmo's 1944 paper/monograph "The Probability Approach in Econometrics", and some related previous literature.

The paper begins by discussing 3, using it as a kind of table of contents for aspects of (1) and (2).

I. The road ahead, as of 1939, and our progress on it

To undertake what Tinbergen did in the 30's, constructing a quantitative model of the economy based on tools of statistics then available, required audacity, energy, and a certain personality type. This personality type was not that of Keynes. He closes his review of the first volume of Tinbergen's League of Nations modeling effort with

> I hope I have not done injustice to a brave pioneer effort. The labour it involved must have been enormous. The book is full of intelligence, ingenuity, and candour; I leave it with sentiments of respect for the author. But it has been a nightmare to live with, and I fancy that other readers will find the same. I have a feeling that Prof. Tinbergen may agree with much of my comment, but that his reaction will be to engage another ten computors and drown his sorrows in arithmetic.

(Of course, in those days, what Keynes meant by "computors" were people hired to do computations.)

[1] These results were discussed in (Sims, 2002)

We still today have economists who love to work with computers, trace out patterns in data, confront complexity in numerical data head on, and who are impatient with theory that asks to be accepted as much on esthetic as on empirical criteria. These are "numbers men"[2]. And communication between numbers men and their counterparts, theorists who relish analytical elegance and rhetorical effectiveness and who are impatient with tedious calculation, remains imperfect. While neither Keynes nor Tinbergen is easily categorized when their work is considered in total, in this dispute Tinbergen was the numbers man.

Keynes weakened his critique by treating as equally weighty problems that we now agree were serious problems in principle with Tinbergen's approach and problems that were minor, or at least a matter of degree rather than of principle. For example, Keynes complained at length about the linearity assumption and about what he saw as the requirement that *all* true causes of the left-hand-side variable appear on the right in a regression, with no double-counting. Tinbergen answered these components of Keynes's criticisms effectively. On the latter point, an uncharitable reading of the Keynes review might suggest he did not understand how multiple regression works.

Nonetheless, a similar pattern of mixing minor issues — even these same minor issues — with more fundamental arguments about applied econometric work, are observed even today, for example in some critiques of the structural VAR literature. When a certain approach inspires revulsion in its critics, it may tend to subvert their rhetorical effectiveness by making them unresistant to the temptation to pile trivia on top of substance in their critiques.

But Keynes made important points as well. Some of these reflect problems in macroeconomic modeling that are still with us. There is the problem of "bad data" or of "castles in the air theory" (two names, from different perspectives, for the same thing). The variables in formal theories are almost never a perfect match for data series with the same name. There are always possibly important causal mechanisms on which we have no data. Keynes and Tinbergen both saw this, Haavelmo articulated the point at length, and the situation has not changed much. There is the problem of what to make of conclusions from a model whose explicit accompanying behavioral stories are known not to be literally true. Keynes knew there was no behavioral story he could believe that would precisely justify Tinbergen's specification. Tinbergen didn't think this a decisive objection. RBC modelers know that their

[2] This phrase is in quotes, because I actually heard it used, referring to a type of economist, by a distinguished colleague.

general equilibrium models with small numbers of agent types are not literally true. They don't think this a decisive objection.

I.1. Expectations

On the other hand, in some areas where Keynes saw problems, we have made progress. After listing variables that Tinbergen takes to explain investment, (Keynes, 1939, f.n.1,p.563) Keynes observes, "I should like to have said 'the *expected* rate of profit'. But there is no room for expectations, so far as I can discover, in the theory of investment with which the economists have supplied Professor Tinbergen."

Haavelmo in his 1958 presidential address to the Econometric Society, argued that econometricians had a responsibility to take a stronger role in the direction of economic theory. One example he gave was expectations; he appeared to believe that econometricians, working with data, were strongly aware of the importance of expectations, and that (in 1958) they should have been pushing theorists to consider theories in which expectations appeared explicitly, as well as pushing statistical agencies to collect better data on expectations. (His other example might be read as a suggestion in 1958 to better develop the field of behavioral economics.)

Whatever our current problems in macroeconomic theory, the lack of theories that give explicit treatment to expectations is not one of them. And the development of these theories, and the beginnings of matching them to data, are a major achievement, clearly only dimly foreseen even as late as the 50's.

I.2. Simultaneity

Both Keynes and Tinbergen recognized the problem of simultaneity. Keynes included it under the heading "spurious correlation" in his critique, and thereby mixed it up with other, less fundamental, problems. Of course Haavelmo in his 1943 *Econometrica* paper showed the path toward resolving this problem, and he and the Cowles Foundation worked it out in subsequent years. This was again a major advance over what either Tinbergen or Keynes understood in 1939.

I.3. The probability approach

Haavelmo wrote his 118 page 1944 article on this topic in an atmosphere where the application of probability-based statistical methods to economic models was a controversial idea. His argument, briefly was:

- We use models in economics that do not make precise predictions that allow rejection of the theory based on precise measurements.

- If our models are to represent science, they must be confronted with data.

- We need therefore to provide commonly accepted principles by which we weight errors, treating some as unimportant or expected, while others are treated as casting strong doubt on the model. He argues that this is always implicitly a matter of putting probability distributions on the errors, and ought to be explicitly so.

Keynes concluded that Tinbergen's modeling effort could *at best* merely give "quantitative precision to what, in qualitative terms, we know already as the result of a complete theoretical analysis." He concluded therefore that "the method is one neither of discovery nor criticism." Keynes could not see how fitting a model could ever result in producing evidence about the validity of a theory! Tinbergen knew this was not correct, but did not have a formal, articulated, counterargument.

Haavelmo embraced Neyman-Pearson testing theory in part because it so clearly answered this objection of Keynes (and similar objections by others). Econometric analysis of time series data *could* result in formal tests of theories. What was required was that the theories be formulated so as to yield probability models for the observed data.

II. Regress

Though we have made progress, our movement has not been monotonic. Progress on expectations has not been completely reversed. Progress on simultaneity has been almost completely reversed. And the probability approach, though not disappeared, is under siege.

II.1. Simultaneity

None of the four central banks I visited (US FRB, ECB, Bank of England, Swedish Riksbank) use as the primary model one in which estimation takes account of simultaneity. The same is true of the major commercial macro models. The topic occupies a much smaller part of the econometrics curriculum in most graduate programs than it did 30 years ago, and many economics graduate students have a weak understanding of the subject: of general principles; not just technical details. That is, they avoid research topics or theories that seem to lead to the need to analyze and fit a model that has multiple equations and shows simultaneity.

II.2. Expectations

Graduate students in macro these days certainly learn how to handle several classes of models (though often only small ones) that include expectations. Only two of the four banks I studied have models that claim to allow for assuming "model consistent expectations". But these models include expectations in ways that few academic macro theorists would defend. Because they are based almost entirely on single-equation inference, they ignore from the start one of the first-noted implications of the rational expectations approach: RE implies cross-equation restrictions and precludes equation-by-equation approaches to estimation and equation specification.

II.3. The probability approach

Much of what Haavelmo wrote on this could be lifted from his 1944 paper and applied directly to the dispute today between econometricians and "extreme calibrationists". By the latter I mean economists who would claim that calibration, i.e. inference without formal appeal to probability-based statistical methods, is not just an occasionally, arguably, necessary expedient when probability-based inference is too complicated, but instead an improved replacement for probability-based inference. In other words, there is serious dispute today about Haavelmo's argument. The dispute is no doubt greater today than it was in, say 1965.

The fact that the models in use at central banks use single-equation estimation methods reflects a flaw at a deeper level: They are not formulated as, and never tested as, models for the joint time series behavior of the data used to estimate them. In other words, we have gone from a day when policy-oriented macroeconometric modelers tried to apply the probability approach, as in the 60's, to a day when the most important policy macro models in actual use have completely forsworn it. They have gone back in many respects to Tinbergen's single-equation modeling approach.

III. How did we reach this position?

Some of the blame has to go to the Luddites of the rational expectations revolution. Use of quantitative models as a guide to real-time policy advice was cast into such deep disrepute that academic research on the topic nearly completely ceased. Real-time policy advice was still needed, and policy makers still wanted staff to help them understand what recent data implied about their policy options, so quantitative

macro modeling continued, always in the time-pressured atmosphere of actual rounds of policy advice and with little cross-fertilization from academia.

Some of the blame must also be placed at the doorstep of Neyman-Pearson theory and Haavelmo's wholesale swallowing of it. It envisions science as a process of sifting through theories, rejecting one false one after another until the truth crops up (after which only one out of 20 published papers rejects it). It therefore is silent when many theories are "unrejectable", as well as when all the interesting theories (say DSGE models) are rejected in favor of an uninteresting (say reduced form VAR) model. In pursuit of an "objective" theory, it insists on never putting probabilities on what is labeled as the "parameter space", or on competing models. It has no mechanism for introducing prior probability beliefs, even when these are of a standardized form, "objective" in the sense that they downweight portions of the parameter space that nearly everyone agrees are uninteresting.

Neyman-Pearson based econometrics then had no answer when RBC modelers saw, as they started their enterprise, that the notion that theorists hand econometricians candidates for "the true theory", and that the econometricians then have the job of returning the message "accept" or "reject" at the 5% level, made no sense for what they were doing.

Decision makers naturally use the language of probability to characterize and discuss major sources of uncertainty. They want to discuss the weight of evidence across models or across values of key parameters, not which of the models or which sets of parameter values are "accepted at the 5% level". For example in the last decade monetary authorities in the US have had to assess whether the rate of productivity growth has shown a sustainable rise. If it has, growth rates of output that might otherwise imply inflationary pressure may be benign. But evidence on this issue of necessity accumulates slowly, and policy decisions cannot be delayed until uncertainty has dissipated. Policy makers need an assessment of the weight of evidence in the data now. A statement that the null hypothesis of no change in the growth rate is "accepted at the 95% confidence level", is of no use to them. They need an assessment of the weight of the currently available evidence. They understand well the language of risk, odds ratios, and loss functions that goes with decision theory. And they therefore are skeptical of "confidence intervals" and "significance levels" that, if explained carefully, are clearly not assessments of odds or probabilities conditional on current data.

More than one central bank staffer, at more than one central bank, gave me almost exactly this argument for why "econometrics" could not be brought to bear in actual discussions with policy makers. But

of course it is only the Neyman-Pearson framework that is irrelevant to policy discussion; not econometric data analysis in general.

Simultaneous equation theory, when applied to large macro models, seemed to say "apply 2SLS or 3SLS", or "apply FIML or LIML". If all instruments suggested by the model structure were used, 2SLS and 3SLS tended to be identical to, or nearly identical to, OLS. In models this big, FIML and LIML tended to be "ill-behaved". Though there were ad hoc resolutions of these difficulties, academic econometric research paid them little attention, and results of careful attention to simultaneity with existing tools tended to be close to OLS, but at much greater computational and staff time expense.

IV. How Bayesian thought patterns can help

- In the many-instruments, few degrees of freedom environment, use of weak priors that rule out bizarre corners of the parameter space can regularize the problem.
- Integrating the posterior, as suggested by a Bayesian approach, is likely to avoid some of the difficulties created for maximum likelihood by sharp, isolated peaks in the likelihood.
- Of course Bayesian approaches lead directly to analysis of how data and prior affect posterior odds across models and posterior distributions of parameters. This fits with the natural language and thought patterns of both decision makers and econometricians. Indeed here is Tinbergen, trying to explain to Keynes what can be learned from Tinbergen's model fitting:

  > certain details on the probability distribution of the [explanatory variables'] 'influences' can be given. These details are the central (most probable) values and the standard deviations of the regression coefficients, measuring the 'influences'.

  Note that Tinbergen is making probability statements about parameter values, not estimators. He was taking a Bayesian perspective.

Bayesian thinking does not help very much if all the interesting (DSGE) models available have very low posterior odds, while reduced form VAR's have high posterior odds. Contrary to assertions one sometimes hears, a Bayesian viewpoint does not comfortably handle choice among models *known* to be false. This misconception may arise because a Bayesian viewpoint does help guide analysis and decision making

when there is uncertainty, given the data, about *which one* of several models is correct. My own view is that when it is clear that all the interesting models are false, it should nearly always be possible to diagnose their faults and get them to fit better. But if there really is no possibility of doing this, it may still be possible to do something useful in analyzing fit. (Schorfheide, 2000) has interesting analysis of these issues.

There are recent signs of progress. Work by (Kloek and Van Dijk, 1978), (Chib, 1995; Chib and Greenberg, 1995), (Geweke 1989, 1999) and others on importing into econometrics Monte Carlo integration methods like importance sampling and Markov Chain Monte Carlo methods, which were applied successfully in other fields, see (Casella and George, 1992), (Hastings, 1970), (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953), (Tanner and Wong, 1987) and (Tierney, 1994), has made possible complete Bayesian analyses of models for which that would have been impossible a few years ago. This approach relaxes the constraint on model complexity somewhat, so that DSGE models that tell appealing stories about behavior can at the same time be complex enough to fit the data about as well as the best Bayesian reduced form VAR's. (Smets and Wouters, 2003) have given us an example of exactly this situation. While I disagree with some of its specific arguments, recent work on the relation of econometric modeling and model choice to policy analysis, for example that by (Brock, Durlauf and West, 2003) and (Leeper and Zha, 2001), as well as by Schorfheide, suggests that the Neyman-Pearson framework may be loosening its grip.

V. What distinguishes "Bayesian thinking" about inference?

Up to this point I have been writing as if the reader understands the fundamental distinctions between Neyman-Pearson theory and Bayesian approaches to inference. It is perhaps worthwhile to close by making these distinctions explicit, since they are not explained well in most econometrics textbooks.

Bayesian inference is distinguished by its recognition that the aim of analysis of data is to guide decision making, and that this will implicitly or explicitly involve putting probability distributions across every source of uncertainty, which includes competing models and parameter values as well as the usual model "disturbance terms". This leads to an emphasis on characterizing the shape of the likelihood function in scientific reporting of data analysis, see e.g. (Kleibergen and Van

Dijk, 1994) and (Kleibergen and Van Dijk, 1998). Bayesian analysis for decision-making weights the likelihood function with a prior probability density function, normalizes the result to integrate to one, and treats the result as an updated ("posterior") probability density function for the parameters. For scientific reporting, where readers may have varying prior distributions, Bayesian analysis either characterizes the likelihood shape directly, or uses standardized prior weighting functions (reference priors) where they are helpful, see e.g. (Strachan and Van Dijk, 2003). A reference prior can be helpful, for example, when nearly all readers agree that certain regions or points in the parameter space are extremely unlikely.

Neyman-Pearson inference insists on never putting a probability distribution on what are labeled as "parameters", even though these may in fact be uncertain and crucial to decision-making. This approach uses probability distributions on *estimators* of parameters to make probability statements that sound as if they are characterizing uncertainty about parameters, and are probably nearly universally interpreted by users as if they are doing so. But in fact, rigorously interpreted, the significance levels and confidence levels of this style of analysis are probabilities only *before* the data are observed. Often in practice they turn out to be approximately the same as Bayesian probability statements about parameters, so it is not common to insist on the distinction. But recently we have become aware that in many cases of importance, including the analysis of possibly non-stationary time series and the analysis of large simultaneous equations models, Bayesian and Neyman-Pearson approaches give systematically different results.

The characteristics listed above — that Bayesian inference leads to probability statements about parameters and models conditional on the data, while non-Bayesian approaches only make probability statements about data before it is observed, never about parameters or models — are really the *only* distinctions between Bayesian and non-Bayesian inference. This point requires emphasis because textbooks often characterize the difference as being that Bayesian inference uses a "subjective" notion of probability, in contrast to the "objective" notion underlying non-Bayesian approaches. Bayesian approaches do insist that decision-making must ultimately involve the decision-maker's assessing his or her prior beliefs. But few doubt that this is in fact true in practice, whatever the methodology of the data-analyis the decision-maker takes account of. In scientific reporting, the day-to-day paper-writing and memo-writing of economic researchers, the Bayesian approach is no more subjective than the Neyman-Pearson approach. Indeed, if anything it is less so, as it exposes formally the mechanism by which

subjective information is or is not invoked, while such information is in fact invoked, but implicitly, in use of non-Bayesian data-analysis for decision-making.

References

Brock, W.A., Durlauf, S.N. and K.D. West, 2003, Policy evaluation in uncertain economic environments, *Brookings Papers on Economic Activity* (1), 1-67.

Casella, G. and E. George, 1992, Explaining the Gibbs sampler, *The American Statistician*, 46(3), 167-174.

Chib, S., 1995, Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, 90(432),1313-1321.

Chib, S. and E. Greenberg, 1995, Understanding the Metropolis-Hastings algorithm, *The American Statistician*, 49(4), 327-335.

Geweke, J., 1989, Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, 57, 1317-1339.

Geweke, J., 1999, Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication, *Econometric Reviews*, 18(1), 1-73.

Haavelmo, T., 1943, The statistical implications of a system of simultaneous equations, *Econometrica* 11(1), 1-12.

Haavelmo, T., 1944, The probability approach in Econometrics, *Econometrica* 12 (supplement), iii-vi+1-115.

Haavelmo, T., 1958, The role of the econometrician in the advancement of economic theory, *Econometrica* 26(3), 351-357.

Hastings, W.K., 1970, Monte Carlo Sampling Methods using Markov Chains and their Applications, *Biometrika*, 57, 97-109.

Keynes, J.M., 1939, Professor Tinbergen's method, *Economic Journal* 49(195), 558-577, book review.

Kleibergen, F.R. and H.K. van Dijk, 1994, On the Shape of the Likelihood/Posterior in Cointegration Models, *Econometric Theory*, 10(3-4), 514-551.

Kleibergen, F.R. and H.K. van Dijk, 1998, Bayesian Simultaneous Equations Analysis using Reduced Rank Structures, *Econometric Theory*, 14(6), 701-743.

Kloek, T. and H.K. van Dijk, 1978, Bayesian estimates of equation system parameters: an application of integration by Monte Carlo, *Econometrica*, 46, 1-19.

Leeper, E. and T. Zha, 2001, Models policy interventions. Technical report, Indiana University and Federal Reserve Bank of Atlanta. http://php.indiana.edu/ eleeper/Papers/lz0101Rev.pdf.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and E. Teller, 1953, Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, 21, 1087-1091.

Schorfheide, F, 2000, Loss function-based evaluation of DSGE models, *Journal of Applied Econometrics* 15(6), 645-670.

Sims, C.A., 1996, Macroeconomics and methodology, *Journal of Economic Perspectives* 10, 105-120.

Sims, C.A., 2002, The role of models and probabilities in the monetary policy process, *Brookings Papers on Economic Activity* 2002(2), 1-62.

Smets, F. and R. Wouters, 2003, An estimated stochastic dynamic general equilibrium model of the euro area, *Journal of the European Economic Association* 1(5), 1123-1175.

Strachan, R.W. and H.K. van Dijk, 2003, The value of structural information in the VAR, Econometric Institute Report EI 2003-17, Erasmus University Rotterdam.

Tanner, M.A. and W.H. Wong, 1987, The Calculation of Posterior Distributions by Data Augmentation (with discussion), *Journal of the American Statistical Association*, 82, 528-550.

Tierney, L., 1994, Markov Chains for Exploring Posterior Distributions, *Annals of Statistics*, 22, 1701-1762.