

STATISTICAL MODELING OF MONETARY POLICY AND ITS EFFECTS

Prize Lecture, December 8, 2011

by

CHRISTOPHER A. SIMS

Princeton University, Princeton, NJ, USA.

The science of economics has some constraints and tensions that set it apart from other sciences. One reflection of these constraints and tensions is that, more than in most other scientific disciplines, it is easy to find economists of high reputation who disagree strongly with one another on issues of wide public interest. This may suggest that economics, unlike most other scientific disciplines, does not really make progress. Its theories and results seem to come and go, always in hot dispute, rather than improving over time so as to build an increasing body of knowledge. There is some truth to this view; there are examples where disputes of earlier decades have been not so much resolved as replaced by new disputes. But though economics progresses unevenly, and not even monotonically, there are some examples of real scientific progress in economics. This essay describes one – the evolution since around 1950 of our understanding of how monetary policy is determined and what its effects are. The story described here is not a simple success story. It describes an ascent to higher ground, but the ground is still shaky. Part of the purpose of the essay is to remind readers of how views strongly held in earlier decades have since been shown to be mistaken. This should encourage continuing skepticism of consensus views and motivate critics to sharpen their efforts at looking at new data, or at old data in new ways, and generating improved theories in the light of what they see.

We will be tracking two interrelated strands of intellectual effort: the methodology of modeling and inference for economic time series, and the theory of policy influences on business cycle fluctuations. The starting point in the 1950s of the theory of macroeconomic policy was Keynes's analysis of the Great Depression of the 1930s, which included an attack on the Quantity Theory of money. In the 30s, interest rates on safe assets had been at approximately zero over long spans of time, and Keynes explained why, under these circumstances, expansion of the money supply was likely to have little effect. The leading American Keynesian, Alvin Hansen, included in his (1952) book *A Guide to Keynes* a chapter on money, in which he explained Keynes's argument for the likely ineffectiveness of monetary expansion in a period of depressed output. Hansen concluded the chapter with, "Thus it is that modern countries place primary emphasis on fiscal policy, in whose service

monetary policy is relegated to the subsidiary role of a useful but necessary handmaiden.”

The methodology of modeling in the 1950s built on Jan Tinbergen’s (1939) seminal book, which presented probably the first multiple-equation, statistically estimated economic time series model. His efforts drew heavy criticism. Keynes (1939), in a famous review of Tinbergen’s book, dismissed it. Keynes had many reservations about the model and the methods, but most centrally he questioned whether a statistical model like this could ever be a framework for testing a theory. Haavelmo (1943b), though he had important reservations about Tinbergen’s methods, recognized that Keynes’s position – doubting the possibility of any confrontation of theory with data via statistical models – was unsustainable. At about the same time, Haavelmo published his seminal papers explaining the necessity of a probability approach to specifying and estimating empirical economic models (1944) and laying out an internally consistent approach to specifying and estimating macroeconomic time series models (1943a).

Keynes’s irritated reaction to the tedium of grappling with the many numbers and equations in Tinbergen’s book finds counterparts to this day in the reaction of some economic theorists to careful, large-scale probability modeling of data. Haavelmo’s ideas constituted a research agenda that to this day attracts many of the best economists to work on improved successors to Tinbergen’s initiative.

Haavelmo’s main point was this: Economic models do not make precise numerical predictions. Even if they are used to make a forecast that is a single number, we understand that the forecast will not be exactly correct. Keynes seemed to be saying that once we accept that models’ predictions will be incorrect, and thus have “error terms”, we must give up hope of testing them. Haavelmo argued that we can test and compare models, but that to do so we must insist that they include a characterization of the nature of their errors. That is, they must be in the form of probability distributions for the observed data. Once they are given this form, he pointed out, the machinery of statistical hypothesis testing can be applied to them.

In the paper where he initiated simultaneous equations modeling (1943a), he showed how an hypothesized joint distribution for disturbance terms is transformed by the model into a distribution for the observed data, and went on to show how this allowed likelihood-based methods for estimating parameters. After discussing inference for his model, Haavelmo explained why the parameters of his equation system were useful: One could contemplate intervening in the system by replacing one of the equations with something else, claiming that the remaining equations would continue to hold. This justification of – indeed definition of – structural modeling was made more general and explicit later by Hurwicz (1962).

Haavelmo’s ideas and research program contained two weaknesses that persisted for decades thereafter and at least for a while partially discredited the simultaneous equations research program. One was that he adopted the frequentist hypothesis-testing framework of Neyman and Pearson. This

framework, if interpreted rigorously, requires the analyst not to give probability distributions to parameters. This limits its usefulness in contributing to analysis of real-time decision-making under uncertainty, where assessing the likelihood of various parameter values is essential. It also inhibits combination of information from model likelihood functions with information in the beliefs of experts and policy-makers themselves. Both these limitations would have been overcome had the literature recognized the value of a Bayesian perspective on inference. When Haavelmo's ideas were scaled up to apply to models of the size needed for serious macroeconomic policy analysis, the attempt to scale up the hypothesis-testing theory of inference simply did not work in practice.

The other major weakness was the failure to confront the conceptual difficulties in modeling policy decisions as themselves part of the economic model, and therefore having a probability distribution, yet at the same time as something we wish to consider altering, to make projections conditional on changed policy. In hindsight, we can say this should have been obvious. Policy behavior equations should be part of the system, and, as Haavelmo suggested, analysis of the effects of policy should proceed by considering alterations of the parts of the estimated system corresponding to policy behavior.

Haavelmo's paper showed how to analyze a policy intervention, and did so by dropping one of his three equations from the system while maintaining the other two. But his model contained no policy behavior equation. It was a simple Keynesian model, consisting of a consumption behavior equation, an investment behavior equation, and an accounting identity that defined output as the sum of consumption and investment. It is unclear how policy changes could be considered in this framework. There was no policy behavior equation to be dropped. What Haavelmo did was to drop the national income accounting identity! He postulated that the government, by manipulating "G", or government expenditure (a variable not present in the original probability model), could set national income to any level it liked, and that consumption and investment would then behave according to the two behavioral equations of the system. From the perspective of 1943 a scenario in which government expenditure had historically been essentially zero, then became large and positive, may have looked interesting, but by presenting a policy intervention while evading the need to present a policy behavior equation, Haavelmo set a bad example with persistent effects.

The two weak spots in Haavelmo's program – frequentist inference and unclear treatment of policy interventions – are related. The frequentist framework in principle (though not always in practice) makes a sharp distinction between "random" and "non-random" objects, with the former thought of as repeatedly varying, with physically verifiable probability distributions. From the perspective of a policy maker, her own choices are not "random", and confronting her with a model in which her past choices are treated as "random" and her available current choices are treated as draws from a probability distribution may confuse or annoy her. Indeed economists

who provide policy advice and view probability from a frequentist perspective may themselves find this framework puzzling.¹ A Bayesian perspective on inference makes no distinction between random and non-random objects. It distinguishes known or already observed objects from unknown objects. The latter have probability distributions, characterizing our uncertainty about them. There is therefore no paradox in supposing that econometricians and the public may have probability distributions over policy maker behavior, while policy makers themselves do not see their choices as random. The problem of econometric modeling for policy advice is to use the historically estimated joint distribution of policy behavior and economic outcomes to construct accurate probability distributions for outcomes conditional on contemplated policy actions not yet taken. This problem is not easy to solve, but it has to be properly posed before a solution effort can begin.

I. KEYNESIAN ECONOMETRICS VS. MONETARISM

In the 1950s and 60s economists worked to extend the statistical foundations of Haavelmo's approach and to actually estimate Keynesian models. By the mid-1960s the models were reaching a much bigger scale than Haavelmo's two-equation example model. The first stage of this large scale modeling was reported in a volume with 25 contributors (Duesenberry, Fromm, Klein, and Kuh, 1965), 776 pages, approximately 150 estimated equations, and a 50 × 75cm foldout flow chart showing how sectors were linked. The introduction discusses the need to include a "parameter" for every possible type of policy intervention. That is, there was no notion that policy itself was part of the stochastic structure to be estimated. There were about 44 quarters of data available, so without restrictions on the covariance matrix of residuals, the likelihood function would have been unbounded. Also, in order to obtain even well-defined single-equation estimates by standard frequentist methods, in each equation a large fraction of the variables in the model had to be assumed not to enter. There was no analysis of the shape of the likelihood function or of the model's implications when treated as a joint distribution for all the observed time series.

The 1965 volume was just the start of a sustained effort that produced another volume in 1969, and then evolved into the MIT-Penn-SSRC (or MPS) model that became the main working model used in the US Federal Reserve's policy process. Important other work using similar modeling approaches and methods has been pursued in continuing research by Ray Fair described e.g. in his 1984 book, as well as in several central banks.

While this research on large Keynesian models was proceeding, Milton Friedman and Anna Schwartz (1963b, 1963a) were launching an alternative view of the data. They focused on a shorter list of variables, mainly measures of money stock, high-powered money, broad price indexes, and measures of real activity like industrial production or GDP, and they examined the behavior of these variables in detail. They pointed out the high correlation between money growth and both prices and real activity, evident in the data

over long spans of time. They pointed out in the 1963b paper that money growth tended to lead changes in nominal income. Their book (1963a) argued that from the detailed historical record one could see that in many instances money stock had moved first, and income had followed. Friedman and Meiselman (1963) used single-equation regressions to argue that the relation between money and income was more stable than that between what they called “autonomous expenditure” and income. They argued that these observations supported a simpler view of the economy than that put forward by the Keynesians: monetary policy had powerful effects on the economic system, and indeed that it was the main driving force behind business cycles. If it could be made less erratic, in particular if money supply growth could be kept stable, cyclical fluctuations would be greatly reduced.

The confrontation between the monetarists and the Keynesian large scale modelers made clear that econometric modeling of macroeconomic data had not delivered on Haavelmo’s research program. He had proposed that economic theories should be formulated as probability distributions for the observable data, and that they should be tested against each other on the basis of formal assessments of their statistical fit. This was not happening. The Keynesians argued that the economy was complex, requiring hundreds of equations, large teams of researchers, and years of effort to model it. The monetarists argued that only a few variables were important and that a single regression, plus some charts and historical story-telling, made their point. The Keynesians, pushed by the monetarists to look at how important monetary policy was in their models, found (Duesenberry, Fromm, Klein, and Kuh, 1969, Chapter 7, by Fromm, e.g.) that monetary policy did indeed have strong effects. They argued, though, that it was one among many policy instruments and sources of fluctuations, and therefore that stabilizing money growth was not likely to be a uniquely optimal policy.

Furthermore, neither side in this debate recognized the centrality of incorporating policy behavior itself into the model of the economy. In the exchanges between Albert Ando and Franco Modigliani (1965) on the one hand, and Milton Friedman and David Meiselman on the other, much of the disagreement was over what should be taken as “autonomous” or “exogenous”. Ando and Modigliani did argue that what was “autonomous” ought to be a question of what was uncorrelated with model error terms, but both they and their adversaries wrote as if what was controlled by the government was exogenous.

Tobin (1970) explained that not only the high correlations, but also the timing patterns observed by the monetarists could arise in a model where erratic monetary policy was not a source of fluctuations, but he did so in a deterministic model, not in a probability model that could be confronted with data. Part of his story was that what the monetarists took as a policy instrument, the money stock, could be moved passively by other variables to create the observed statistical patterns. I contributed to this debate (1972) by pointing out that the assumption that money stock was exogenous, in the sense of being uncorrelated with disturbance terms in the monetarist regres-

sions, was testable. The monetarists regressed income on current and past money stock, reflecting their belief that the regression described a causal influence of current and past money stock on current income. If the high correlations reflected feedback from income to money, future money stock would help explain income as well. It turned out it did not, confirming the monetarists' statistical specification.

The monetarists' views, that erratic monetary policy was a major source of fluctuations and that stabilizing money growth would stabilize the economy, were nonetheless essentially incorrect. With the right statistical tools, the Keynesians might have been able to display a model in which not only timing patterns (as in Tobin's model), but also the statistical exogeneity of the money stock in a regression, would emerge as predictions despite money stock not being the main source of fluctuations. But they could not do so. Their models were full of unbelievable assumptions² of convenience, making them weak tools in the debate. And because they did not contain models of policy behavior, they could not even be used to frame the question of whether erratic monetary policy behavior accounted for much of observed business cycle variation.

II. WHAT WAS MISSING

Haavelmo's idea, that probability models characterize likely and less likely data outcomes, and that this can be used to distinguish better from worse models, fits neatly with a Bayesian view of inference, and less comfortably with the Neyman-Pearson approach that he adopted. Since standard statistics courses do not usually give a clear explanation of the difference between Bayesian and frequentist inference, it is worth pausing our story briefly to explain the difference. Bayesian inference aims at producing a probability distribution over unknown quantities, like "parameters" or future values of variables. It does not provide any objective method of doing so. It provides objective rules for updating probability distributions on the basis of new information. When the data provide strong information about the unknown quantities, it may be that the updating leads to nearly the same result over a wide range of possible initial probability distributions, in which case the results are in a sense "objective". But the updating can be done whether or not the results are sensitive to the initial probability distribution.

Frequentist inference estimates unknown parameters, but does not provide probability distributions for them. It provides probability distributions for the behavior of the estimators. These are "pre-sample" probabilities, applying to functions of the data before we observe the data.

We can illustrate the difference by considering the multiplier-accelerator model that Haavelmo³ used to show that probability-based inference on these models should be possible. Though it is much smaller than the Keynesian econometric models that came later, at the time many fewer data were available, so that even this simple model could not have been sharply estimated from the short annual time series that were available.

The model as Haavelmo laid it out was

$$C_t = \beta + \alpha Y_t + \varepsilon_t \quad (1)$$

$$I_t = \theta(C_t - C_{t-1}) + \eta_t \quad (2)$$

$$Y_t = C_t + I_t \quad (3)$$

He assumed $\varepsilon_t \sim N(0, \sigma_c^2)$ and $\eta_t \sim N(0, \sigma_i^2)$ and that they were independent of each other and across time. He suggested estimating the system by maximum likelihood.

He intended the model to be useful for predicting the effect of a change in government spending G_t , though G_t does not appear in the model. This was confusing, even contradictory. We will expand the model to use data on G_t in estimating it. He also had no constant term in the investment equation. We will be using data on gross investment, which must be nonzero even when there is no growth, so we will add a constant term. Our modified version of the model, then, is

$$C_t = \beta + \alpha Y_t + \varepsilon_t \quad (1')$$

$$I_t = \theta_0 + \theta_1 (C_t - C_{t-1}) + \eta_t \quad (2)$$

$$Y_t = C_t + I_t + G_t \quad (3')$$

$$G_t = \gamma_0 + \gamma_1 G_{t-1} + v_t \quad (4)$$

We will confront it with data on annual real consumption, gross private investment, and government purchases from 1929 to 1940.⁴

The model does not make sense if it implies a negative multiplier – that is if it implies that increasing G within the same year decreases Y . It also does not make sense if γ_1 , the “accelerator” coefficient, is negative. Finally, it is hard to interpret if γ_1 is much above 1, because that implies explosive growth. We therefore restrict the parameter space to $\theta_1 > 0$, $\gamma_1 < 1.03$, $1 - \alpha(1 + \theta_1) > 0$. The last of these restrictions requires a positive multiplier. The likelihood maximum over this parameter space is then at

α	β	θ_0	θ_1	γ_0	γ_1
0.566	166	63.0	0.000	10.7	0.991

Note that the maximum likelihood estimator (MLE) for θ_1 is at the boundary of the parameter space. At this value, the investment equation of the model makes little sense. Furthermore, the statistical theory that is used in a frequentist approach to measure reliability of estimators assumes that the true parameter value is not on the boundary of the parameter space and that the sample is large enough so that a random sample of the data would make finding the MLE on the boundary extremely unlikely.

θ_1 probability density

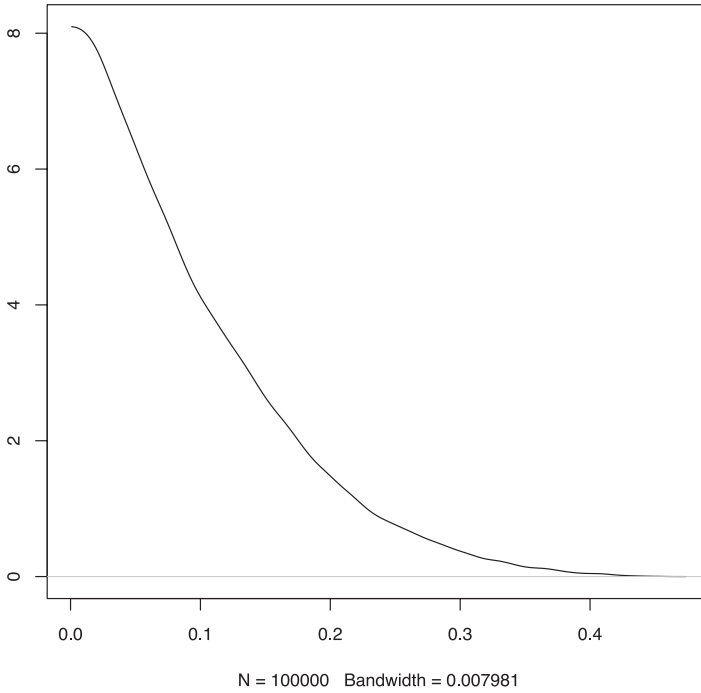


Figure 1. θ_1 probability density.

A Bayesian approach to inference provides a natural and reasonable result, though. The probability density over the parameter space after seeing the data is proportional to the product of the likelihood function with a prior density function. If the prior density function is much flatter than the likelihood, as is likely if we began by being very uncertain about the parameter values, the likelihood function itself, normalized to integrate to one, characterizes our uncertainty about the parameter values. With modern Markov Chain Monte Carlo methods, it is a straightforward matter to trace out the likelihood and plot density functions for parameters, functions of parameters, or pairs of parameters. Under a flat prior, the density function for γ_1 has the shape shown in Figure 1. While the peak is at zero, any value between 0 and 0.25 is quite possible, and the expected value is 0.091. The system's dynamics with $\theta_1 = 0.2$ would be very different from dynamics with θ_1 close to zero. So the data leave substantively important uncertainty about the value of θ_1 and do not at all rule out economically significant accelerator effects. The within-year multiplier in this model, that is the effect of a unit change in G_t on Y_t , is $1/(1 - \alpha(1 + \theta_1))$. Its flat-prior posterior density is shown in Figure 2. Note that the maximum likelihood estimate of the multiplier, shown as a vertical line in the figure, is 2.30, well to the left of the main mass of the posterior distribution. This occurs because the multiplier increases with θ_1 , and the MLE at zero is unrepresentative of the likely values of θ_1 .

Probability density of multiplier

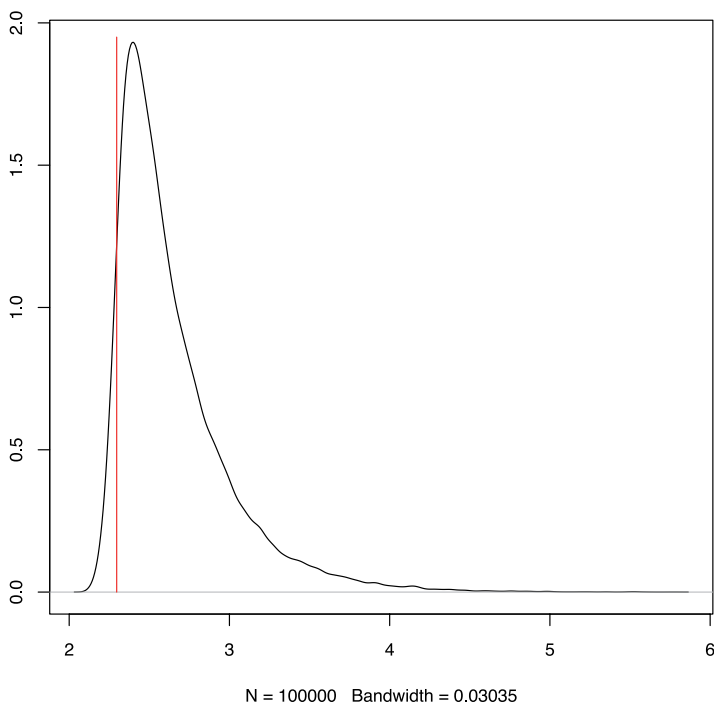


Figure 2. Probability density of multiplier.

In calculating the “multiplier” here, I am looking at the impact of a change in G_p in the context of a model in which G_t is part of the data vector for which the model proposes a probability distribution. There are several ways of thinking about what is being done in this calculation. One is to say that we are replacing the “policy behavior equation” (4) by the trivial equation $G_t = G^*$, holding the other equations fixed, and considering variations in G^* . Another, equivalent, way to think of it is that we are considering choosing values of v_p , the disturbance to the policy equation. The latter approach has the advantage that, since we have an estimated distribution for v_p , we will notice when we are asking about the effects of changes in v_t that the model considers extremely unlikely.⁵ While there is nothing logically wrong with asking the model to predict the effects of unlikely changes, simplifying assumptions we have made in setting up the model to match data become more and more questionable as we consider more extreme scenarios.

Neither of these ways of looking at a multiplier on G is what Haavelmo did in his hypothetical policy experiment with the model. In fact he did not calculate a multiplier at all. He instead suggested that a policy-maker could, by setting G (which, recall, was not in his probability model), achieve any desired level Y^* of total output. He recognized that this implied the policy-maker could see ε_t and η_t and choose G_t so as to offset their effects. He noted that under these assumptions, the effects of changes in Y^* on C_t and

I_t could be easily calculated from equations (1) and (2). He said that what he was doing was dropping the accounting identity (3) and replacing it with $Y_t = Y^*$, but one cannot “drop” an accounting identity. What he was actually doing was replacing an implicit policy equation, $G_t \equiv 0$, with another, $G_t = Y^* - C_t - I_t$, while preserving the identity (3'). Since policy-makers probably cannot in fact perfectly offset shocks like η_t and ε_p and since they are more likely to have seen themselves as controlling G_t than as directly controlling Y_t , this policy experiment is rather artificial.

If Haavelmo had tried to fit his model to data, he would have had to confront the need to model the determination of his policy variable, G_t . My extension of Haavelmo's model in (1')-(4) specifies that lagged values of C_t and I_t do not enter the G_t equation (4) and that the disturbance of that equation is independent of the other two disturbances. This implies, if this equation is taken as describing policy behavior, that G_t was determined entirely by shifts in policy, with no account being taken of other variables in the economy. This would justify estimating the first two equations in isolation, as Haavelmo suggested. But in fact the data contain strong evidence that lagged C_t and I_t do help predict G_t .⁶ If the model was otherwise correct, this would have implied (quite plausibly) that G_t was responding to private sector developments. Even to estimate the model properly would then have required a more complicated approach.

This discussion is meant only as an example to illustrate the difference between frequentist and Bayesian inference and to show the importance of explicitly modeling policy. It is not meant to suggest that Haavelmo's model and analysis could have been much better had he taken a Bayesian approach to inference. The calculations involved in Bayesian analysis of this simple model (and described more fully in the appendix) take seconds on a modern desktop computer, but at the time Haavelmo wrote were completely infeasible. And the model is not a good model. The estimated residuals from the MLE estimates show easily visible, strong serial correlation, implying that the data have richer dynamics than is allowed for in the model.

In large macroeconomic models it is inevitable that some parameters – some aspects of our uncertainty about how the economy works – are not well-determined by the data alone. We may nonetheless have ideas about reasonable ranges of values for these parameters, even though we are uncertain about them. Bayesian inference deals naturally with this situation, as it did with the prior knowledge that γ_1 should be positive in the example version of Haavelmo's model. We can allow the data, via the likelihood function, to shape the distribution where the data are informative, and use pre-data beliefs where the data are weak.

When we are considering several possible models for the same data, Bayesian inference can treat “model number” as an unknown parameter and produce post-sample probabilities over the models. When a large model, with many unknown parameters, competes with a smaller model, these posterior probabilities automatically favor simpler models if they fit as well as more complicated ones.

The models the Keynesians of the 1960s were fitting were orders of magnitude larger than Haavelmo's, with many hundreds, or even thousands, of free "parameters" to be estimated. Asking the data to give firm answers to the values of these parameters was demanding a lot, too much, in fact. The statistical theory that grew out of Haavelmo's ideas, known as the Cowles Foundation methodology, provided approximate characterizations of the randomness in estimators of these parameters, on the assumption that the number of time series data points was large relative to the number of parameters being estimated, an assumption that was clearly untrue for these models. It was frequentist theory, and therefore insisted on treating unknown quantities as non-random "parameters". To provide reasonable-looking estimates, modelers made many conventional, undiscussed assumptions that simplified the models and made ad hoc adjustments to estimation methods that had no foundation in the statistical theory that justified the methods.⁷

The result was that these models, because they included so many ad hoc assumptions that were treated as if they were certain a priori knowledge and because they were estimated by methods that were clearly dubious in this sort of application, were not taken seriously as probability models of the data, even by those who built and estimated them. Measures of uncertainty of forecasts and policy projections made with the models were not reliable.

III. NEW EVIDENCE AND NEW MODELING IDEAS

While the Keynesian modelers were working on their unwieldy large models, and while Friedman and Meiselman and Anderson and Jordan (1968) were estimating single-equation models explaining income with money, other economists were estimating single equations explaining money with income and interest rates. These latter were labeled "money demand" equations. If, as my 1972 paper implied, M did in fact behave in the regressions of income on money like a legitimate explanatory variable, it seemed likely that the empirical money demand equations were mis-specified. So Mehra (1978) set out to check whether that was so. He found, surprisingly, that that equation also passed tests of necessary conditions that income and interest rates were explaining money causally. The only way to reconcile these results was to put the variables together in a multiple-equation model to study their dynamics.

In my 1980a paper I estimated such a model, as a vector autoregression, or VAR, the type of model I was suggesting in "Macroeconomics and Reality" (1980b). VAR's are models of the joint behavior of a set of time series with restrictions or prior distributions that, at least initially, are symmetric in the variables. They include, for example, restrictions on lag length – the same restrictions on all variables in all equations – or prior distributions favoring smooth evolution of the time series in the model. Of course to make the results interesting, they require some interpretation, which brings in economic theory at least informally. In the 1980a paper I included data on interest rates, production, prices, and money. The results showed clearly that much of the variation in money stock is predictable from past values of the

interest rate, particularly so after World War II. Furthermore, a considerable part of the postwar variation in the interest rate was in turn predictable based on past values of industrial production. Since interest rates were in fact a concern of monetary policy makers, this made it hard to argue that money stock itself was an adequate representation of policy or to argue that money supply variation consisted largely of erratic mistakes. If monetary policy was in part systematic and predictable historically, it was no longer clear that shutting down its variation would reduce business cycle fluctuations.

Of course in the 1970s there was increased attention to modeling policy behavior from another side. The rational expectations hypothesis was applied to macroeconomic models. It required that models that included explicit expectations of the future in behavioral equations should base those expectations on the full model's own dynamic structure. A simple version of a rational expectations model was already common in studies of financial markets: the "efficient markets" hypothesis stated that excess returns on assets should be approximately unpredictable, as otherwise there would be profit opportunities from trading on the predictable returns. This meant that the kind of test I had applied to the money-income regressions would tend to imply that any asset price from a smoothly functioning market "causes" movements in any other publicly observable variable. Some economists (Fischer Black, in particular, in conversation with me) thought from the start that money appeared to have a unidirectional causal impact on income in the monetarist regressions for the same reason that stock prices would, and that the result was therefore not strong support for the monetarist causal interpretations of those regressions. But neither the monetarist nor the Keynesian models at the time implied that quantity of money had the same kind of properties as an asset price. While my 1980a paper and subsequent work with VAR models made clear that monetary policy responded to the state of the economy and that the money stock was predictable in a model including interest rates, an explicit theoretical model that validated Fischer Black's intuition arrived later in my 1989 paper, which showed that a monetary policy of making interest rates respond positively to the growth rate of the money stock would lead to an apparent causal ordering from money to income, even in a model with negligible effects of monetary policy on real variables.

Another main conclusion from applying the rational expectations hypothesis was that a model of policy behavior was required to accurately model expectations in the private sector. Since the large Keynesian models had devoted no attention to careful modeling of policy behavior, they could not easily take this criticism into account. While the emphasis from this viewpoint on modeling policy behavior was valuable, the effect on policy modeling of the rational expectations "critique" of the large Keynesian models was for a few decades more destructive than constructive.

Some of the early proponents of rational expectations modeling, such as Sargent (1973), presented it as implying cross-equation restrictions on models that were otherwise similar in structure to the then-standard ISLM

models. But even in the small model Sargent used in that article, estimating the complete system as a model of the joint time series behavior was not feasible at the time – he used the model to derive single equations that he estimated and used to test implications of rational expectations. Maximum likelihood estimation of complete systems embodying rational expectations at the scale needed for policy modeling was not possible.

Probably even more important to the inhibiting effect on policy-oriented econometric inference was the emphasis in the rational expectations literature on evaluating non-stochastic alterations of a policy rule within the context of a rational expectations model. The intuition of this point could easily be made clear to graduate students in the context of the Phillips curve. If one accepted that only surprise changes in the price level had real effects, then it was easy to show that a negative correlation between unemployment and inflation might emerge when no attempt was made to control unemployment by manipulating inflation (or vice versa), but that there might nonetheless be no possibility of actually affecting unemployment by deliberately changing the inflation rate, assuming those deliberate changes were anticipated. If one thought of the Phillips curve in this story as standing in for a large Keynesian model, the implication was that what policy makers actually did with econometric models – use them to trace out possible future paths of the economy conditional on time paths for policy variables – was at best useless and at worst might have been the source of the US inflation of the 1970s. This story was widely believed, and it supported a nearly complete shutdown in academic economists' interest in econometric policy modeling.

There was in fact no empirical support for this story. The Phillips curve negative correlation of unemployment with inflation did indeed disappear in the 1970s, but this quickly was reflected in Keynesian empirical policy models, so those models implied little or no ability of policy to reduce unemployment by raising inflation. Rising inflation was not a deliberate attempt to climb up a stable Phillips curve. I would support this argument, which is no doubt still somewhat controversial, in more detail if this essay were about the evolution of macroeconomics generally, but for current purposes, we need only note the effect of the story on research agendas: few economists paid attention to the modeling tasks faced by the staffs of monetary policy institutions.

The emphasis on changes in rules as policy experiments was unfortunate in another respect as well. As we have noted, it was a major defect in Haavelmo's framework and in the simultaneous equation modeling that followed his example that policy changes were always modeled as deterministic, coming from outside the stochastic structure of the model. Recognition of the importance of modeling policy behavior ought to have led to recognition that policy changes should be thought of as realizations of random variables, with those random variables modeled as part of the model's structure. Instead, the mainstream of rational expectations modeling expanded on Haavelmo's mistake: treating policy changes as realizations of random variables was regarded as inherently mistaken or contradictory; attention was focused

entirely on non-stochastic, permanent changes in policy behavior equations, under the assumption that these equations had not changed before and were not expected to change again after the intervention.⁸

Large econometric models were still in use in central banks. New data continued to emerge, policy decisions had to be made, and policy makers wanted to understand the implications of incoming data for the future path of the economy, conditional on various possible policy choices. Modelers in central banks realized that the frequentist inference methods provided by the Cowles Foundation econometric theorists were not adapted to their problems, and, in the absence of any further input from academic econometricians, reverted to single-equation estimation. No longer was any attempt made to construct a joint likelihood for all the variables in the model. There were attempts to introduce rational expectations into the models, but this was done in ways that would not have been likely to stand up to academic criticism – if there had been any.⁹

Vector autoregressive models were not by themselves competitors with the large policy models. They are statistical descriptions of time series, with no accompanying story about how they could be used to trace out conditional distributions of the economy's future for given policy choices. In my earliest work with VAR's (1980a; 1980b) I interpreted them with informal theory, not the explicit, quantitative theoretical restrictions that would be needed for policy analysis. It was possible, however, to introduce theory explicitly, but with restraint, so that VAR's became usable for policy analysis. Blanchard and Watson (1986) and my own paper (1986) showed two different approaches to doing this. Models that introduced theoretical restrictions into VAR's sparingly in order to allow them to predict the effects of policy interventions came to be known as structural vector autoregressions, or SVAR's.

IV. CONSENSUS ON THE EFFECTS OF MONETARY POLICY

By 1980 it was clear that money stock itself was not even approximately a complete one-dimensional measure of the stance of monetary policy. Interest rates were also part of the picture. Policy makers in the US and most other countries thought of their decisions as setting interest rates, though perhaps with a target for a path of money growth in mind. They were also concerned about the level of output and inflation, trying to dampen recessions by lowering rates and restrain inflation by raising rates. But there are many reasons why interest rates, money, output and prices are related to output and inflation other than the behavior of monetary policy makers. Interest rates tend to be higher when inflation is high, because lenders require to be compensated for the loss in value of the loan principal through inflation. Interest rates will change when the real rate of return to investment changes, which can happen for various reasons not related to monetary policy. Private sector demand for money balances can shift, because of financial innovation or fluctuating levels of concern about liquidity. Untangling these patterns of

mutual influence to find the effects of monetary policy is inherently difficult and can at best produce results that leave some uncertainty.

In my 1986 paper, I attempted this untangling by a combination of strategies. I postulated that interest rate changes could affect private sector investment decisions only with a delay, and also that the Federal Reserve could not, because of delays in data availability, respond within the quarter to changes in output or the general level of prices. The model also included an attempt to identify a money demand equation, using two different sets of additional restrictions. The effects of monetary policy identified this way were quite plausible: a monetary contraction raised interest rates, reduced output and investment, reduced the money stock, and slowly decreased prices. The effects on output of unpredictable disturbances to monetary policy were non-trivial, but accounted for only a modest fraction of overall variability in output. That the responses emerged as “reasonable” was in fact part of the identification strategy. The precise zero-restrictions on some coefficients were interacting with qualitative views as to what a response to a monetary policy contraction should look like.

This pattern of results turned out to be robust in a great deal of subsequent research by others that considered data from other countries and time periods and used a variety of other approaches to SVAR-style minimalist identification. A summary of some of this research appeared in Leeper, Sims, and Zha (1996). It was widely accepted as a reasonable quantitative assessment of how monetary policy changes affect the economy.

SVAR's that isolated an equation for monetary policy behavior could be used for making conditional policy projections, but they did not become widely used as the main model in central bank policy discussions. Future policy is not the only future event that policy makers like to condition on in making projections. Scenarios involving high or low commodity prices due to supply disruptions, high or low productivity growth, decline in the value of the dollar, fiscal policy changes, etc. are often important to policy discussion. Since SVAR's were limiting themselves to isolating monetary policy, treating the rest of the economy as a single “black box” system, they could not easily provide these types of conditional forecasts. Whether accurate or not, the existing large scale models could at least provide answers.

Christiano, Eichenbaum, and Evans (2005) developed a complete dynamic, stochastic general equilibrium model (DSGE), in which all disturbances and equations had economic interpretations and reflected assumed optimizing behavior, yet which also could reproduce the pattern of responses to a monetary policy shock that had emerged from SVAR models. Extending their work,¹⁰ Frank Smets and Raf Wouters (2007; 2003) showed that the model could be used to form a likelihood function and that formal Bayesian inference to estimate the parameters and characterize uncertainty about them was possible. They could compare their models to VAR's and SVAR's, and they showed that their models were competitive with VAR's in terms of fit to the data. They thus finally delivered on Haavelmo's project: a macroeconomic model usable for policy analysis that was in the form of an

asserted probability distribution for the data. It could be compared in fit to other models, and proposed improvements to it could be evaluated on the basis of their effects on the full system, not just one equation. Central bank research staffs around the world recognized the potential value of this type of model, and many have since developed models of this type. In some cases this type of model has become the main model used in policy discussions.

V. ARE WE THERE YET?

Haavelmo advocated formulating economic models as probability models not as an end in itself, but because he saw this as the only way it would be possible to test models against one another and thereby make progress. We now have, in Bayesian DSGE models, the first step in Haavelmo's program. But these models in their current form are ripe for improvement, in several directions.

The recent financial crash and recession was not predicted by the DSGE models. Predictions from probability models by construction will be subject to error. It is therefore not the existence of the errors, or even their size, that is a problem; it is that the errors in the forecasts from the DSGE's (and standard SVAR's, for that matter) were of a size that the models' probability structures implied should almost never occur. That too many large errors occur, compared to the predictions of the models, was verifiable before the recent crisis, even though the scale of errors in the recent crisis was greater. Recognizing the true likelihood of large errors in economic models is a technical challenge, which is one reason they received little attention from modelers. Perhaps more important was that unusually large errors are by definition rare events. It is therefore not possible for the historical data by some mechanical statistical procedure to generate a probability model for the large errors. That is, the role of prior distributions will be important in modeling these errors and uncertainty about them will be great. An approach based on the usual tools of frequentist, large-sample-approximate econometrics will not suffice.

A related gap in the current crop of DSGE's is that they are nearly all based on linearization around a steady-state growth path. Users of the models are well aware that the models therefore are likely to become unreliable in the face of large deviations of the economy from its trend line. Large disturbances push us well off the trend line, however, and we need to identify and model the important nonlinearities that then arise.

Already many economists are working at integrating more financial data and more complete models of the financial sector into DSGE models, in hopes that these improvements will help avoid another crash like that of 2008, or failing that make policy choices in the wake of such a crash more accurately apparent. Equally important, in fact probably more important looking ahead, is extending DSGE's to more accurately model fiscal dynamics. The Bayesian DSGE's mostly do not model, or do not model accurately, national debt, the connection of deficits to debt, and the wealth effects on

private sector behavior of debt and expected future fiscal policy. These aspects of the economy are central to current policy discussions.

The existing DSGE's are mainly variants on the Christiano, Eichenbaum and Evans model. This is a "micro-founded" model, in which inertia and stickiness, instead of being injected into the model in a purely ad hoc way, are pushed one level down, into the constraints of optimizing representative agents. But these micro foundations are in many instances clearly at odds with empirical micro evidence or common sense. Price stickiness, for example, is modeled by assuming that there is a constraint or a cost associated with changing a nominal price, and that firms setting prices are all individually negligible in size. Everyone understands that actual industries are nearly always characterized by wide divergences in firm size, with a few large firms being strategically important. Everyone understands that the "cost" of changing a nominal price or the constraint that nominal prices can only be changed at times whose occurrence the firm cannot control are at best a strained metaphor. If one thinks of DSGE's as a set of stories that make economists and policy-makers more comfortable with policy projections that basically reproduce what would be implied by SVAR's, the implausibility of the micro-foundations of the DSGE's are of secondary importance. But when the models are used to evaluate welfare effects of alternative policies, these issues become more important. It is unlikely that macroeconomic models that are explicitly built up from micro data will be feasible in the near future. But it is feasible to experiment with variations on the inertial mechanisms in the Christiano, Eichenbaum and Evans framework, investigating whether other specifications can fit as well and might have different policy implications.

Existing DSGE's are too tightly parameterized. The Smets and Wouters US model (2007) for example uses data on seven variables and has 37 free parameters, about five per "equation", which is a parameter density that is probably lower than characterized the large-scale Keynesian models of the 60s and 70s. The result is that Bayesian VAR's fit better than the DSGE's, by a substantial margin.¹¹ The DSGE's could be made to fit better by adding parameters allowing more dynamics in the disturbances or more flexible specifications of various sources of inertia. Since we think of the theory in these models as at best approximate, though, a more promising approach may be that of DelNegro and Schorfheide (2004), extended by DelNegro, Schorfheide, Smets, and Wouters (2007), who use a DSGE as the source for a prior distribution on the parameters of a SVAR. In their procedure the result can in principle be nearly identical to the DSGE, if the DSGE fits the data well. Park (2010) has extended the Del Negro/Schorfheide framework to make it more realistically reflect uncertainties about identification. This approach to modeling, since it does not treat the DSGE as directly explaining the data, makes using the model's microtheory for evaluating welfare effects of policy impossible. But as I've noted above, this may be all to the good. And use of the model to trace out the distribution of the future of the economy conditional on various sources of disturbance remains possible, and likely more accurate than using the DSGE model directly.

VI. CONCLUSION

Despite some early confusion about how to bring macroeconomic time series data to bear on the issue, the controversy between Keynesian and Quantity Theory views of the effects of standard monetary policy is at least for the time being largely resolved. Interest rate changes engineered by open market operations do have substantial effects on the economy, both on real output and inflation. Erratic shifts in monetary policy are not the main source of cyclical variation in the economy. The quantity of money is not a good one-dimensional index of monetary policy. Effects of monetary policy on output are fairly quick; effects on inflation take longer to play out. The methods of inference that have been developed in resolving these issues have brought us close to realizing Haavelmo's goals for a scientific approach to macroeconomics.

Nonetheless, there remain uncertainties even about the new consensus view, and the models now at the frontier still contain major gaps. Much remains to be done.

APPENDIX: INFERENCE FOR THE HAAVELMO MODEL

The model defined by (1')–(4) is in the form of a slightly non-standard simultaneous equations model:

$$\Gamma_0 z_t = \phi + \Gamma_1 z_{t-1} + \zeta_t, \quad (5)$$

with

$$\Gamma_0 = \begin{bmatrix} 1 & 0 & -\alpha & 0 \\ -\theta_1 & 1 & 0 & 0 \\ -1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ -\theta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_1 \end{bmatrix},$$

$$\phi = \begin{bmatrix} \beta \\ \theta_0 \\ 0 \\ \gamma_0 \end{bmatrix}, z_t = \begin{bmatrix} C_t \\ I_t \\ Y_t \\ G_t \end{bmatrix}, \zeta_t = \begin{bmatrix} \varepsilon_t \\ \eta_t \\ 0 \\ v_t \end{bmatrix}.$$

The non-standard aspect of the model is that the covariance matrix of ζ_t is constrained to be diagonal. The simultaneous equations literature that emerged from Haavelmo's insights treated as the standard case a system

in which the joint distribution of the disturbances was unrestricted, except for having finite covariance matrix and zero mean. It is interesting that Haavelmo's seminal example instead treated structural disturbances as independent, as has been the standard case in the later structural VAR literature.

Likelihood-based inference is straightforward, if we condition on the initial (1929) observation. The distribution of ζ_t is Normal, with a diagonal covariance matrix, i.i.d. across t . So the log pdf of $\zeta_{1930}, \dots, \zeta_{1940}$ is

$$-10 \log \sigma_\varepsilon - 10 \log \sigma_\eta - 10 \log \sigma_v - \frac{1}{2} \sum_t \frac{\varepsilon_t^2}{\sigma_\varepsilon^2} - \frac{1}{2} \sum_t \frac{\eta_t^2}{\sigma_\eta^2} - \frac{1}{2} \sum_t \frac{v_t^2}{\sigma_v^2}. \quad (6)$$

To arrive at the log likelihood, we first substitute $C_t + I_t + G_t$ for Y_t to reduce the system to three variables and three equations. Then we write the shocks as functions of the parameters and the data, using (5) and multiply by the Jacobian of the transformation from the data vector z_t to the shock vector ζ_t , which is

$$|\Gamma_0^*|^T,$$

where Γ_0^* is the contemporaneous coefficient matrix from the 3×3 version of the system solved to eliminate Y_t .

The likelihood can be integrated analytically with respect to the three η parameters. The integration treated the prior on these parameter as flat in $1/\sigma_\varepsilon^2, 1/\sigma_\eta^2$ and $1/\sigma_v^2$. With those parameters integrated out, the resulting function of the six parameters $\alpha, \beta, \theta_o, \theta_I, \gamma_o$ and γ_I can be treated as a marginal posterior pdf for those parameters.

The maximum likelihood estimates displayed in the text are arrived at by numerical maximization of this marginal posterior. (Thus they are probably somewhat different from the values of those parameters at the joint likelihood peak, were maximization over the σ parameters and the other six jointly.) To generate the full posterior distribution, a random walk Metropolis algorithm was used, with jump distribution $N(0, \Omega)$, where Ω was 0.3 times a crude estimate of the second derivative of the log posterior density near its peak. The crude estimate was just that produced as the approximate inverse-Hessian during the likelihood maximization computation. Despite the maximum of the posterior pdf being on the boundary, both the maximization and the MCMC iterations converged fairly easily. 100,000 draws were used for the MCMC computations. The coefficients α and β in the consumption function showed effective sample sizes (using the **coda** package for R) of 297 and 284, while for the other parameters the effective size was over one thousand.

REFERENCES

- Anderson, L. C., and J. L. Jordan (1968), "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization," *Federal Reserve Bank of Saint Louis Review*, pp. 11–24.
- Ando, A., and F. Modigliani (1965), "The Relative Stability of Monetary Velocity and the Investment Multiplier," *American Economic Review*, 55(4), 693–728.
- Blanchard, O.J., and M.W. Watson (1986), "Are Business Cycles All Alike?," in *The American Business Cycle: Continuity and Change*, ed. by R. Gordon, pp. 123–156, University of Chicago Press, Chicago, Illinois.
- Christiano, L., M. Eichenbaum, and C. Evans (2005), "Nominal Rigidities and the Dynamics Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.
- Delnegro, M., and F. Schorfheide (2004), "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.
- Delnegro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit and Forecasting Performance of New Keynesian Models," *Journal of Business and Economic Statistics*, 25(2), 123–162.
- Duesenberry, J. S., G. Fromm, L. R. Klein, and E. Kuh, eds. (1965), *The Brookings Quarterly Econometric Model of the United States*. Rand Mc-Nally.
- eds. (1969), *The Brookings Model: Some Further Results*. Rand McNally.
- Fair, R. C. (1984), *Specification, Estimation, and Analysis of Macroeconometric Models*, Harvard University Press.
- Friedman, M., and D. Meiselman (1963), "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958," in *Stabilization Policies*, Commission on Money and Credit.
- Friedman, M., and A. J. Schwartz (1963a), *A Monetary History of the United States, 1867–1960*, Princeton University Press.
- Friedman, M., and A. J. Schwartz (1963b), "Money and Business Cycles," *The Review of Economics and Statistics*, 45(1), 32–64.
- Haavelmo, T. (1943a), "The statistical implications of a system of simultaneous equations," *Econometrica*, 11(1), 1–12.
- (1943b), "Statistical Testing of Business-Cycle Theories," *The Review of Economics and Statistics*, Vol. 25, No.1 (Feb., 1943), pp. 13–18, published by: The MIT Press Stable. URL: <http://www.jstor.org/stable/1924542>.
- (1944), "The Probability Approach in Econometrics," *Econometrica*, 12 (supplement), iii–vi + 1–115. Hansen, A. H. (1952), *A Guide to Keynes*, McGraw-Hill, New York; Toronto; London.
- Hurwicz, L. (1962), "On the Structural Form of Interdependent Systems," in *Logic, Methodology and Philosophy of Science*, pp. 232–239, Stanford University Press, Stanford, CA.
- Keynes, J. M. (1939), "Professor Tinbergen's method," *Economic Journal*, 49(195), 558–577, book review.
- Leeper, E. M., C. A. Sims, and T. Zha (1996), "What Does Monetary Policy Do?," *Brookings Papers on Economic Activity*, (2), 1–78.
- Leeper, E. M., and T. Zha (2003), "Modest Policy Interventions," *Journal of Monetary Economics*, 50(8), 1673–1700.
- Liu, T.-C. (1960), "Underidentification, Structural Estimation, and Forecasting," *Econometrica*, 28(4), pp. 855–865.
- Mehra, Y. P. (1978), "Is Money Exogenous in Money-Demand Equations," *The Journal of Political Economy*, 86(2), 211–228.
- Park, W. Y. (2010), "Evaluation of DSGE Models: With an Application to a Two-Country DSGE Model," Discussion paper, Princeton University, <http://www.sef.hku.hk/~wypark/>.

- Samuelson, P. A. (1939), "Interactions Between the Multiplier Analysis and the Principle of Acceleration," *Review of Economics and Statistics*.
- Sargent, T. J. (1973), "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment," *Brookings Papers on Economic Activity*, 1973(2), 429–480.
- (1984), "Autoregressions, Expectations, and Advice," *American Economic Review*, 74, 408–15, Papers and Proceedings.
- Sims, C. A. (1972), "Money, Income, and Causality," *The American Economic Review*, 62(4), 540–552. — (1980a), "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered," *American Economic Review*, 70, 250–57.
- (1980b), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- (1986), "Are Forecasting Models Usable for Policy Analysis?," *Quarterly Review of the Minneapolis Federal Reserve Bank*, 10, 2–16.
- (1987), "A rational expectations framework for short-run policy analysis," in *New approaches to monetary economics*, ed. by W. A. Barnett, and K. J. Singleton, pp. 293–308, Cambridge University Press, Cambridge, England.
- (1989), "Models and Their Uses," *American Journal of Agricultural Economics*, 71, 489–494.
- (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 Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.
- (2007), "Shocks and frictions in US business cycles: a Bayesian DSGE approach," *American Economic Review*, 97(3), 586–606.
- Tinbergen, J. (1939), *Statistical testing of business-cycle theories*, no. v. 1–2 in Publications: Economic and financial, League of Nations, Economic Intelligence Service, <http://books.google.com/books?id=E8MSAQAAMAAJ>.
- Tobin, J. (1970), "Money and Income: Post Hoc Ergo Propter Hoc?," *The Quarterly Journal of Economics*, 84(2), 301–317.

(ENDNOTES)

- 1 An example of a sophisticated economist struggling with this issue is Sargent (1984). That paper purports to characterize both Sargent's views and my own. I think it does characterize Sargent's views at the time, but it does not correctly characterize my own.
- 2 This fact, which everyone in some sense knew, was announced forcefully by Liu (1960), and much later re-emphasized in my 1980b paper.
- 3 Haavelmo's model differs from the classic Samuelson (1939) model only in using current rather than lagged income in the consumption function.
- 4 We use the chain indexed data, which did not exist when Haavelmo wrote. We construct Y as $C + I + G$, since the chain indexed data do not satisfy the accounting identity and we are not using data on other GDP components.
- 5 This is the point made, with more realistic examples, by Leeper and Zha (2003).
- 6 I checked this by fitting both first and second order VAR's.
- 7 For example, two-stage least squares was widely used to estimate equations in these models. According to the theory, the more exogenous and predetermined variables available, the better the (large-sample) properties of the estimator. But in these models there was an embarrassment of riches, so many instruments that two-stage least squares using all of them reduced to ordinary least squares – the same method Tinbergen used pre-Haavelmo. So in fact, the modelers used fewer instruments, without any formal justification.
- 8 I argued against this way of formulating policy analysis at more length in 1987.
- 9 I surveyed the state of central bank policy modeling and the use of models in the policy process in a 2002 paper.

- 10 The work by Christiano, Eichenbaum and Evans circulated in discussion paper form for years before its final publication.
- 11 Smets and Wouters report in their tables only marginal likelihoods for Bayesian VAR's that are at least slightly worse than the marginal likelihood of the DSGE. In footnote 13, however, they report that a BVAR with a different prior produces a log marginal likelihood better than that of the DSGE by 9.2 in log units, a substantial advantage in fit for the BVAR.