Asking Impossible Questions About the Economy and Getting Impossible Answers

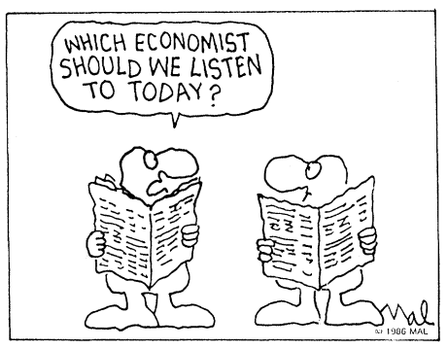
Some economists say that large-scale computer models of the economy are no better at forecasting than economists who simply use their best judgment

"THERE are two things you are better off not seeing in the making—sausages and econometric estimates," says Edward Leamer, an economist at the University of California at Los Angeles. These estimates are used by policymakers to decide, for example, how the new tax law will affect the economy or what would happen if a new oil import tax were imposed. They also are used by businesses to decide whether there is a demand for a new product. Yet the computer models that generate these estimates, say knowledgeable critics, have so many flaws that, in Leamer's words, it is time to take "the con out of econometrics."

Not everyone agrees. But even the defenders of econometric models concede that the critics have a point. Economists Kenneth Arrow of Stanford and Stephen McNees of the Federal Reserve Board in Boston say they believe the models can be useful but also say that one reason the models are made and their predictions so avidly purchased is that people want answers to impossible questions and are overly impressed by answers that come out of a computer. Arrow makes an analogy with the theory of evolution. Asking an economist to accurately forecast next year's energy demand is like asking an evolutionary biologist what species will evolve next. It is not a slur on the theory to say that it can't be done.

Yet whether or not it can be done, it is done. And sometimes everyone pays for it. The Department of Energy and the electric utilities relied on models to predict future demand for electricity a few years ago, says economist Douglas Hale who is director of quality assurance at the Energy Information Administration. The models forecasted that there would be far more energy demand than actually materialized. Power companies, believing the models, over-built and now consumers are paying for this excess capacity.

Econometric models have been around for decades and are a mainstay of economic forecasting. Essentially, researchers fit a set of equations to economic data and then use these equations to make forecasts and analyses. The problem, says statistician David Freedman of the University of California at Berkeley, is that "there is no economic theory that tells you exactly what the equations should look like."



Reprinted by special permission of Malcolm Hancock © 1986 Mal

Some model-builders do not even try to use economic theory. Others try to use as much theory as they can to at least estimate what the equations should look like. But most end up by curve-fitting—a risky business since there are an infinite number of equations that will fit any particular data set. Eventually, they get a set of several dozen to several hundred equations that model data that they have at hand from past years. The model's equations may fit, for example, the curves for gross national product from 1960 until 1980. Then the investigators test it to see if it accurately predicts the gross national product for 1981. If it does not, says Freedman, the researchers "fiddle some more." They continue to "fiddle and fit" until the model seems to be working well on data from the past.

Now it is time to ask the model to make a prediction. The economists may ask it to predict next year's gross national product, for example. More often than not they will get an answer that they know must be wrong. It will predict that the gross national product will go up by 7%, for example, when they know it changes only glacially from year to year. So they go back to their model and do a "subjective adjustment" by readjusting the parameters so that the predictions come out as expected.

This does not mean that the model is finished. The fiddling and fitting go on each year and the models are continually altered to fit the modelers' views of economic reality. "What you really have," says William Ascher of Duke University, "is a man-model system." And this system, say the critics, is hardly scientific. Wassily Leontief of New York University remarks, "I'm very much in favor of mathematics, but you can do silly things with mathematics as well as with anything else."

Defenders of the models point out that economists are just making the best of an impossible situation. Their theory is inadequate and it is impossible to write down a set of equations to describe the economy in any event. "No one really expects to model the economy," says Ascher. "There are many things that are not amenable to equations, such as investment climate."

"Why aren't the models better?" asks McNees. "Because forecasting is a pretty formidable task. Modelers are trying to capture the structure of the economy with a few hundred equations. A lot of things must be left out. There also is a data problem. The models are based on, at the most, the postwar experience. There are not enough data to get the coefficients right." Moreover, McNees says, the economy keeps changing so that the models are confounded. " The economy now is very different than it was in the 1950's. We have variable rate loans, for example, and a financial infrastructure that no one ever heard of in the 50's."

In addition, says Arrow, even if the data were better, "the statistics couldn't handle them. The statistical methods were built up because they were easy to build up." For example, he notes, "nonlinear relationships should be in the models but the statistics are not equipped for them. There is a conflict between theory and statistics and you make of it the best you can."

But the critics of the models say that none of these defenses makes up for the fact that the models are, as Leontief says, "hot air." Very few of the models can predict accurately, the economic theory behind the models is extremely weak if it exists at all, in many cases the data used to build the models are of such poor quality as to be essentially useless,

and the model-builders, with their subjective adjustments, produce what is, according to Leamer, "an uncertain mixture of data and judgment."

Hale, whose agency is one of the few that regularly assess models to see how they are doing, reports that, "in many cases, the models are oversold. The scholarship is very poor, the degree of testing and peer review is far from adequate by any scientific measure, and there is very little that you can point to where one piece of work is a building block for the next."

For example, the Energy Information Administration looked at the accuracy of short-term forecasts for the cost of crude oil. These were forecasts done 1½ years in advance, predicting, in May of 1984, the cost of crude oil in the second quarter of 1985. The models were from the Energy Information Association and also from three hugely popular commercial modeling firms—Data Resources, Incorporated, Chase Econometrics, and Wharton Economic Forecasting Associates. At first glance, it looks as if they did not do too badly. They all had similar estimates of about $34 a barrel, an increase of only $7 from the actual figure of $27.04. But, says Hale, "what we really are interested in is how much does the price change over time. The error in predicting change is 91%."

Some forecasts, of course, are much better than others. Oil prices are particularly hard to predict, and, because energy prices affect the whole economy, their effects are also hard to guess.

Others who have looked at how well the models do are Victor Zarnowitz of the University of Chicago and Ascher. Neither reports that the models are particularly good. Zarnowitz finds that "when you combine the forecasts from the large models and take an average, they are no better than the average of forecasts from people who just use their best judgment and do not use a model." But, he cautions, "this does not necessarily mean there are no advantages to models. Models can produce an enormous number of predictions quickly whereas it takes longer with just judgment. But, on the other hand, revising a model is costly and slow."

Ascher, in contrast, finds that "econometric models do a little bit worse than judgment. And for all the elaboration over the years, they haven't gotten any better. Refining the models hasn't helped." Ascher says he finds it "somewhat surprising" that the models perform worse than judgment since judgment is actually part of the models; it is incorporated in when modelers readjust their data to conform to their judgment.

Another way of assessing models is to ask whether you would be better off using them or just predicting that next year will be like this year. This is the approach taken by McNees, who first looked at the sizes of the errors that a group of major models made in predicting such things as gross national product and inflation rates. The errors were quite large for the periods 1974 to 1975 and 1981 to 1982 when there were recessions but, in other periods, such as the late 1970's the errors were smaller. "I would argue that if you average over all the periods, you would make smaller errors with the models than you would by simply assuming that next year will be just like this year," he says. "But the errors would not be tremendously smaller. We're talking about relatively small orders of improvement."

"People are overly impressed by answers that come out of a computer."

Other investigators are asking whether the models' results are reproducible. When the modelers publish their equations and their data, can someone else then plug in the data and be sure to get the same forecast? Surprisingly, the answer seems to be no. "There is a real problem with scholarship in the profession," says Hale of the Energy Information Administration. "Models are rarely documented well enough so that someone else can get the same result."

For example, William Dewald, an economist at the State Department's Bureau of Economic and Business Affairs, with Jerry Thursby and Richard Anderson of Ohio State University, attempted to replicate results published in the *Journal of Money, Credit, and Banking*. They asked 62 authors whose papers were published in the journal within the past few years for data. About two-thirds of the authors were unwilling to supply their data in enough detail for replication. In those cases where the data and equations were available, Dewald and his colleagues succeeded in replicating the original results only about half the time.

Anderson points out that their failure to replicate does not mean all the studies had errors. Instead, he says, "the authors could not furnish us enough information from their own records to tell us what they did. We could not tell whether their results are right or wrong." It is impossible, he notes, to learn from or build on models that are so poorly documented.

Walter Kramer, Harold Sonnberger, Johann Maurere, and Peter Havlik of the Institute for Advanced Studies in Vienna, Austria, looked at the simplest of econometric models—the single equation models—and asked whether they could replicate published results and whether the models passed simple diagnostic tests. These were models forecasting such things as long-term money demand or unemployment rate or the value of stocks. They found that the models failed the tests and that it was difficult or impossible to replicate many of the predictions.

Hale and Freedman are trying to assess a number of large forecasting models by developing methods for determining confidence intervals. Then they want to see whether the actual events that are predicted fall into these intervals. In addition, they are doing sensitivity analyses—determining how much the results of the models are affected by small changes in the inputs. "So there is work going on. This agency hasn't been blind to the problems with models," says Hale.

Hale also notes that the Department of Energy regularly publishes a directory of all its models and describes each of them. "Other agencies don't do that. The numbers just appear," he says. By describing the models, the energy department also avoids buying the same models over and over again—something that happens in the rest of the government, according to Hale. But Hale does not even hint that the energy department is getting more reliable forecasts for its efforts. "It's not that the results are any better. It's just that it's a lot easier to audit and see where the results came from," he says.

Since no one, not even the supporters of the models, seems happy with them, what should be done? Some say that there is no real problem in continuing to go along the way we are. "I don't think any policy-maker takes predictions from models as God's truth," says Raymond Fair of Yale University. "People in government understand the limitations of models."

Others disagree. But no one expects the modelers to go out of business. "It is very difficult to make any changes," says Leontief. "With the gigantic investment in econometrics, what are you going to do?" McNees also expects the modeling business to continue. "Even if you think the models are complete garbage, until there is an obviously superior alternative, people will continue to use them," he says. ■ GINA KOLATA

ADDITIONAL READING

W. G. Dewald, J. G. Thursby, R. G. Anderson, "Replication and scientific standards in empirical economics," *Am. Econ. Rev.*, Sept. 1986, pp. 587–603.

S. K. McNees and J. Ries, "The track record of macroeconomic forecasts," *New Engl. Econ. Rev.*, Nov./Dec. 1983, pp. 5–18.

D. A. Freedman, "As others see us: A case study in path analysis," *J. Educ. Stat.*, in press.