



Illustration by BINAY SINHA

# Statistical snakeoil

*There is a deep divide among econometricians and economists concerning causality, says DEEPAK LAL*

In the mid 80s, when I was the Research Administrator at the World Bank, a talented econometrician on the research staff came to my office saying that he had produced a comprehensive econometric model for all aspects of development policy, and that after looking at the paper he was leaving behind, I should get all the research staff to work on producing the data for the various modules that fed into his model, as no further research on development policy was needed. I politely said I would look at the paper and get back to him. After he left, my secretary came in to say that his wife had rung to say that she was very worried about his meeting with me, as he had failed to take his daily dose of lithium, which kept his manic-depression in check, and he was in a manic mood. Of course I took no further action on his manic scheme.

It, however, seems that the mania for finding quantitative certainty from various statistical models has not ceased and is now infecting many aspects of macro-economic, development and health policy. The recent credit crunch has exposed the failings of the quantitative risk management models, which had previously led to the LTMC crisis. Curbs on carbon emissions are being promoted on the basis of highly insecure climate change

models (see Chapter 11, in my *Against Dirigisme*). The mindless correlations of medical epidemiology are leading to numerous ridiculous health scares (as discussed in my earlier columns). The US push for democracy in the world is based on political scientists' purported statistical correlations showing that democracies do not fight each other (a claim shown to be spurious in my *In Praise of Empires*). Closer home, the macro-economic debate is being hijacked by self-proclaimed empiricists whose simple-minded atheoretical statistical correlations are largely meaningless.

For, when I was learning my econometrics at the feet of Terence Gorman and Alan Walters at Oxford in the early 1960s, it was drummed into us that one had to begin with a theoretical model whose predictions depended on statistical estimation of its parameters, taking account of standard estimation problems like identification, serial correlation and multi-collinearity. As most of the regressions in those days had to be done on handheld calculators, or by punching hundreds of cards to be used on gigantic computers whose power was less than any standard PC today, one had to take immense care to ensure the robustness of the data one was using and the functional form one was estimating, in order not to waste one's

time. The PC revolution has meant that any researcher can now run millions of regressions on large, readily available data sets of highly variable quality, covering countries of different sizes, institutions and histories and obtain some statistically significant result or the other, which one of the myriad of social science journals will be happy to publish. Common sense, the qualitative judgments which allow us to recognise meaningless or unpersuasive results, and specific knowledge (including historical) of the numerous countries thrown into the statistical regression sausage machine, seem to have gone out of the window.

The problem of course is the difficulty of assigning causality in the social sciences where J S Mill's "method of difference", which underlies the controlled experiments of the physical sciences to determine cause and effect, cannot be applied. It would take me too far afield to outline the different stances on causality in economics and econometrics (but the article by Kevin Hoover in the *New Palgrave Dictionary of Economics*, 2nd edn, provides a succinct discussion). The two major approaches to causality in economics go back to Hume and J S Mill. Hume believed that all ideas are based either on logic or sense experience, and that

our inductive inferences based on constant conjunction of particular temporal sequences do not give us secure grounds from observing instances to inferring a general rule. The a priorism implicit in this Humean scepticism underlies the a priori approach in econometrics best exhibited by the structural model estimation framework pioneered by the Cowles Commission. The alternative atheoretical inferential approach to causality is the Granger causality pioneered by Clive Granger, which is purely data-based and does not rely on any background economic theory. This is part of the modern probabilistic approach to causality, "which is a natural successor to Hume. Where Hume required constant conjunction of cause and effect, probabilistic approaches are content to identify cause with a factor that raises the probability of the effect" (Hoover, *op. cit*).

Thus there continues to be a deep divide among econometricians and economists concerning causality. "There are those who believe that economic logic itself provides privileged insight into economic behavior (a priori approaches) and those who believe that we must learn about economic behavior principally through observation and induction (the inferential approaches)" (Hoover, *op. cit*). In our *The Political Economy of Poverty and Growth*, Hal Myint and I argued for a third approach which was supported by Maynard's less well known father J N Keynes, which combines these two approaches in the comparative historical approach which we label "analytical economic history". This approach allows the pattern predictions which Hayek argued are alone possible in the complex orders of the social sciences and biology, as contrasted with the specific predictions possible in simple orders — such as physics. It combines the deductive analysis of economic theory with both the quantitative inductive analysis of statistics (to the extent that is possible) with the qualitative inductive analysis of comparative history.

Thus whilst the inferences drawn about causality from either of the two econometric and analytical economic history approaches can be persuasive, the simple-minded and atheoretical correlations being peddled in the current economic policy and development debates should be looked upon as statistical snake oil to be taken at one's peril.