Critical Review
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rcri20

MACROECONOMETRICS: THE SCIENCE OF HUBRIS

Arnold Kling [a]

[a] Cato Institute and a member of the Financial Markets Working Group, Mercatus Center, George Mason University, 3351 North Fairfax Drive, Arlington, VA, 22201

Available online: 10 Jun 2011

Arnold Kling

MACROECONOMETRICS: THE SCIENCE OF HUBRIS

ABSTRACT: *Macroeconometric models are built on astonishingly precarious grounds and yet are used by policy makers to project precision and certainty. Econometricians use lagged dependent variables, "add factors," and other techniques to make their models more accurate—at the expense of the integrity of the models. The reason for the unscientific nature of macroeconometric models is that, unlike the objects of controlled experimentation, real-world events are often unique and non-repeatable. Models that use repeatable events are poorly suited to accurate prediction or historical explanation.*

Ten days prior to President Obama's inauguration in 2009, two of his economists, Christina Romer and Jared Bernstein (2009), published a memorandum analyzing the effects of fiscal stimulus proposals. The appendix to the memo described a central part of their computations:

> For the output effects of the recovery package, we started by averaging the multipliers for increases in government spending and tax cuts from a leading private forecasting firm and the Federal Reserve's FRB/US model. The two sets of multipliers are similar and are broadly in line with other estimates.

Arnold Kling, arnold@arnoldkling.com, adjunct scholar with the Cato Institute and a member of the Financial Markets Working Group, Mercatus Center, George Mason University, 3351 North Fairfax Drive, Arlington, VA 22201, is the author, *inter alia, of Unchecked and Unbalanced: How the Discrepancy between Knowledge and Power Caused the Financial Crisis and Threatens Democracy* (Rowman and Littlefield, 2010).

Critical Review 23(1–2): 123–133 ISSN 0891-3811 print, 1933-8007 online

DOI: 10.1080/08913811.2011.574475

The appendix went on to list the values of these multipliers. For example, an increase in government purchases of 1 percent of GDP would, if sustained for eight quarters, raise real GDP by 1.57 percent, according to Romer and Bernstein.

In the introduction to the memo, the authors caution,

> Our estimates of economic relationships and rules of thumb are derived from historical experience and so will not apply exactly in any given episode. Furthermore, the uncertainty is surely higher than normal now because the current recession is unusual both in its fundamental causes and its severity.

However, the rest of the memo conveys certainty and precision. The multipliers are reported with two decimal places. The estimated employment effects are reported with four significant figures—table 1 states that the stimulus will increase employment by 3,675,000 jobs. All figures are reported as exact point estimates, rather than ranges plus or minus.

Such figures give the impression of exquisite control. The memo conveys that with the proper manipulation of the dials and levers of fiscal policy, any level of employment can be achieved.

The reality is somewhat different. How unreliable are the multiplier estimates in the Romer-Bernstein memo? There would be no reason to quibble if an estimate of 1.57 were indicative of a range of, say, 1.27 to 1.77 for the multiplier. In fact, it would be satisfactory if the range of possibilities for the multiplier were between 0.57 and 2.57. However, the scientific basis for multiplier analysis is so poor that hardly any value for the multiplier can be ruled out. We cannot say for certain that a 1 percent increase in government purchases will not raise GDP by more than 5 percent or actually *decrease* it by more than 2 percent.

Neither the economists who report multipliers nor the policy makers who study them have an interest in questioning their scientific basis. Economists who doubt the validity of macroeconometrics cannot compete in the market for policy advice. Those of us who have no faith in such models come across as nihilistic and unhelpful. When challenged to come up with better estimates, we can only murmur that this is not possible.

People are more confident if they believe that they are guided by precise information. Nobel Laureate Daniel Kahneman tells a story[1] of a group of Swiss soldiers who got lost while hiking in the Alps. They managed to find their way down the mountain, and they attributed their good fortune to having a map. However, on examination, it turned out that their map was of the Pyrenees. Not surprisingly, policy decisions are made on the basis of macroeconometric models that may be as far removed from reality as the Pyrenees are from the Alps.

Within the economics community, it is not some tiny group of skeptics that doubts the usefulness of macroeconometric models. On the contrary, the model-builders themselves have been relegated to the outer fringe of the profession.

From the late 1950s through the mid-1970s, leading academic economists, including Nobel Laureates Lawrence Klein and Franco Modigliani, were active contributors to the design and evaluation of macroeconometric models. Since then, the connection between this type of modeling and academia has been severed. A few models continue to be maintained, notably at a couple of consulting firms and at the Federal Reserve Board. But analysis based on traditional macroeconometric models has disappeared from the peer-reviewed academic literature of the past thirty years. This is part of a larger gulf between macroeconomic research and macroeconomic policy making, lamented by N. Gregory Mankiw in his paper, "The Macroeconomist as Scientist and Engineer":

> The fact that modern macroeconomic research is not widely used in practical policymaking is *prima facie* evidence that it is of little use for this purpose. The research may have been successful as a matter of science, but it has not contributed significantly to macroeconomic engineering.

In the academy, the highly mathematical macroeconomic theorists (Mankiw's scientists) are respected. The macroeconometric modelers (his engineers) are not. I believe that neither approach yields useful research, but this paper is confined to explaining why the engineers' attempt to acquire knowledge through statistical computations using macroeconomic data is futile.

In the first section, I will recall a few of my experiences with the macroeconometric models of the 1970s in order to highlight some of the alarming characteristics of the model-building process. Next, I will discuss how macroeconometric models fell into disrepute with the mainstream profession. Finally, I will look at the fundamental reason that macroeconometric modeling has no scientific foundation: it employs methods designed for quasi-experimental data to macroeconomic quantities that do not remotely approximate a quasi-experimental data set.

Inside the Sausage Factory

Macroeconometric models consist of equations that relate various macroeconomic quantities—such as GDP, the unemployment rate, and inflation—to one another. The model-builder specifies the variables that enter each equation, and the causal structure for each equation, a priori. This structure is then fit to historical data, using statistical computation techniques, primarily linear and nonlinear regression. Typically, the model's confrontation with the data produces awkward or surprising results, in which case the modeler adjusts the a priori structure and re-estimates the model. This process goes through many iterations.

The estimated equations are subject to a number of ad hoc adjustments. When I was working with models as a researcher at the Congressional Budget Office and the Federal Reserve Board in 1975-76, and as an economist with the Federal Reserve Board in the early 1980s, we were aware that various time periods were affected by unusual factors such as steel strikes in the 1950s or the imposition of wage and price controls in the early 1970s. Special adjustments were made to account for such specific distortions.

In addition, in any period some of the model's equations would be off track, in which case the model proprietor would make a judgment as to how much of an "add factor" or "constant adjustment" to insert into the equation. For example, suppose that the model would predict a value for consumer spending in the latest quarter that is 2 percent below the actual value. If the modeler's instinct is that the error will persist, the modeler will insert an "add factor" that raises the model's forecast for consumer spending by 2 percent each subsequent quarter.

In the 1970s, three consulting firms dominated the private market in macroeconometric models: Chase Econometrics, Wharton Economic Forecasting Associates, and Data Resources. Each of the consulting firms made extensive use of "add factors," and subscribers to the forecasts were in effect paying for the hunches of the model jockeys rather than for the unadjusted models per se. Studies of forecasting performance that were published periodically by researchers at the Federal Reserve

Bank of Boston showed that the models were not very accurate with the "add factors" and hopelessly off base without them.

More recently, the model-based forecasts of early 2008 called for the unemployment rate to peak at about 8 percent, when in fact it reached 10 percent. This discrepancy was too public for the model to be tweaked after the fact, but the Obama administration nonetheless claimed that the stimulus spending that had been premised on the accuracy of the model had "saved or created" 2.5 to 3.6 million jobs (Council of Economic Advisers 2010). More important than the questionable nature of this estimate is the fact that the stimulus had been based on the model in the first place.

In effect, the use of "add factors" amounts to a selective, judgmental inclusion of lagged dependent variables. Lagged dependent variables are the previous quarter's value of that variable. By using an "add factor" for the unemployment rate, a model proprietor is saying, in effect, "Last period, the value of unemployment was higher than the equation would have predicted. I believe that there is information in this error, and I am going to assume that the equation would continue to predict unemployment if I were not to include an add factor."

The most accurate forecasting models use lagged dependent variables. However, the resulting equations often have structural properties that are not desired by the model builders. Including the lagged dependent variable serves to vitiate the model's structure. On the other hand, failure to include the lagged dependent variable tends to cause large prediction errors. The use of "add factors" can be viewed as an attempt to keep the modelers' preferred equations while judgmentally including the information in the lagged dependent variable so as to make a somewhat accurate prediction.

The dilemma of what to do about the lagged dependent variable has no satisfactory resolution. Statistical analysis of the properties of equations estimated on nonstationary data (data where the lagged dependent variable is a very important predictor) shows that scientific rigor requires incorporating lagged dependent variables.[2] However, the rigorous approach (known as vector autoregression) that has enjoyed some popularity among academics tends to yield equations that are defective for the purpose of making policy, in part because they lack the clear structure of the traditional models. The macroeconometric modeling fraternity instead continues to impose a priori views of economic relationships while attempting to keep lagged dependent variables suppressed.

Why Academics Split from Modelers

In 1976, Robert Lucas published a paper that eventually caused most leading economists to sour on macroeconometric models. The "Lucas critique," as it became known, was a particular challenge to the way that macroeconometric modelers dealt with causal structure.

Lucas pointed out that if individuals were economically rational in their approach to forming expectations about economic variables, then macroeconometric models that assumed otherwise would break down.

As it happened, macroeconometric models did perform poorly in the 1970s. The reasons for this breakdown had little or nothing to do with the Lucas critique. The problem, in part, was the weakness described in the previous section, and was in part the more fundamental problem discussed in the next section.

However, the economics profession behaved as if the Lucas critique was the overwhelming problem with macroeconometrics and banned it from their midst. The past thirty years of macroeconomic theory can fairly be described as one large gloss on the Lucas critique. Today's academic theorists (Mankiw's "scientists") pride themselves on building models that are robust with respect to the Lucas critique. By the same token, models that fail to address the critique (those maintained by Mankiw's "engineers") are absent from the professional journals, and in consequence, so is scholarly research based on traditional macroeconometric models. But as Mankiw points out, the macroeconometric fraternity and the policy advising community have brushed off the Lucas critique without refuting it.[3] After all, if the critique is valid, then the models are not.

A further reason for the break between macroeconometricians and academic economists is that some of the latter disagree with the intellectual basis of the Lucas critique—namely the assumption of rational behavior (cf. Krugman 2009; Akerlof and Shiller 2009). Although this position has merit, one can reject the Lucas critique and still view macroeconometric models as empty, as I will explain in the next section.

Non-Experimental Science

Economists aspire to scientific rigor. The gold standard for scientific rigor is the controlled experiment. Traditionally, economists have not been able to conduct controlled experiments. Instead, we use statistical techniques and primarily regression analysis.

Suppose, for example, that you want to estimate the effect of the minimum wage on teenage unemployment rates. One way to do that is to compare teenage unemployment rates in different communities with different minimum-wage levels. This would be an observational study rather than a controlled experiment because the minimum-wage levels would not be assigned randomly to cities by the economist. The teenage unemployment rate in any one city could be affected by many factors, including the size of the city, the types of industry located in that city, the relative size of the teenage labor force, and so on. These other factors could be correlated with the minimum wage rate in some way that distorts the outcome of the study. Hence, it is the job of the economist to identify and control for such factors.

In theory, once the other factors have been controlled for, the only variable that can account for differences in teenage unemployment rates across cities is the minimum wage. Thus, introducing the proper controls results in a quasi-experiment.

If there were only a handful of cities but a dozen variables that must be controlled for, the study cannot be conducted. Under such circumstances, no quasi-experiment is possible.

One solution is a "natural experiment," such as the division of Germany after the Second World War into a Communist East and a non-Communist West. Natural experiments and related techniques have become the preferred empirical methods in economics. Joshua D. Angrist and Jörn-Steffen Pischke (2010) call this the "credibility revolution" in econometrics, although they limit the scope of this revolution to empirical microeconomics. Macroeconomic research tends not to lend itself to the newer design-oriented approaches to empirical work.

Macroeconometrics takes as its unit of observation not a city but a unit of time, typically a calendar quarter. When a macroeconometrician uses regression, he or she is implicitly saying, in effect, that the third quarter of 2007 is the same as the first quarter of 1988, once all factors that might be different between those two quarters are controlled for. The idea is that the economist is conducting an intertemporal quasi-experiment. But because there is only one economic history with which to work, there is a lack of experimental control. The "natural experiment" and related techniques that have enhanced the credibility of econometric studies of microeconomic issues are not available to the macroeconomist.

An almost limitless number of factors could affect key macroeconomic variables from quarter to quarter. For each factor, moreover, there are several potential specifications for the variable representing that factor. The variable might be entered into the equations as linear or nonlinear, de-trended or not, current or lagged, and so forth. Perhaps the effect of the level of average house prices is different from the effect of the *rate of change* of house prices. *Ceteris paribus*, a higher level might reduce demand, but a higher rate of increase might increase demand. The number of factors to be controlled for is further enlarged by "special factors," such as the steel strikes or wage/price controls alluded to earlier. All things considered, there are thousands of plausible specifications of equations.

On the other hand, the number of data points is limited. Time periods more than twenty years before the present are typically dropped because the meaning of the data has been altered, either by new collection techniques or by technological innovation. For example, how does the propensity to spend on durable goods in 1975 relate to this propensity in 2009, given that in 1975 personal computers and microwave ovens were much less prevalent? How does labor demand in 1970 relate to labor demand today, given that manufacturing production workers now account for less than half of the proportion of the labor force that they did then, and given that the proportion of the labor force with at least some college education has more than doubled?

If we have twenty years of quarterly data, then how many observations do we have? If you answered "eighty," then you know how to multiply, but you did not give the correct answer. In fact, there are fewer observations, because of the problem of time aggregation. Simply from the standpoint of measurement, the data in any one quarter are influenced by the data in the preceding quarter. That is one reason that the lagged dependent variable is such a powerful predictor. To filter out the noise in adjacent quarters, it might be best to limit our sample to the first quarter of every year, which would leave only twenty observations in twenty years. (This is a crude way of looking at time aggregation. The literature has much more rigorous, albeit mathematically dense, treatments.)

To summarize, in macroeconometrics, we have, in effect, no more than a few dozen data points, with thousands of plausible control variables (taking into account alternative specifications). There is not even a remote resemblance to a quasi-experiment.

To put this another way, suppose that an alien with a strong belief in scientific empiricism were to visit our planet. Observing macroeconomists arguing over the structure of macroeconomic equations, the alien might propose the Ultimate Modeling Contest. That is, the alien would say, "Look, why don't you just let the data decide? Build a model that includes every possible specification, and then see which model fits the best."

The problem with the Ultimate Modeling Contest is that there are not enough data to let them decide. With thousands of model specifications entering the contest, and less than a few dozen points of meaningful data, the computer is asked to solve a system of many equations with only a few unknowns. The printout should consist of an error message. In statistical terminology, there are no degrees of freedom.

By analogy, consider the problem of finding the genetic pattern that predicts a person's propensity to get a particular disease or to have a particular trait. Because there are many more possible genetic patterns than there are people, this problem cannot be solved simply by loading data and having the computer search without any a priori theory. Instead, the investigator's know-how must be used to tell the computer which sorts of patterns are most likely to be predictive.

Now, imagine how much more difficult this problem would be if the relationships between genes and the diseases or traits that they influence were constantly in flux. That would mean that the number of patterns that the computer must search through is much greater. The task would become far more difficult, indeed, almost certainly impossible. That is what happens in macroeconomics.

Explaining the past is almost as difficult as predicting the future. For instance, economists continue to write papers about what caused the severity and duration of the Great Depression, and on what caused the eventual recovery. For the most part, we are not discovering new data. Rather, we are coming up with new explanations of old data. We are no closer to having a definitive explanation than historians are to having a definitive explanation for the causes of the First World War. This is because, like historians, economists are dealing with individual episodes rather than repeatable experiments.

With macroeconomic data, the degrees of freedom belong to the modeler, not to the data. The data do not dictate the properties of the model. Instead, the modeler can choose the properties of the model and then fit the data to those properties. In fact, models have been constructed that have only classical supply-side impacts for policies, and no Keynesian demand-side impacts whatsoever. Such models are scorned by policy makers and by Keynesian academics, but they actually have a greater presence in the recent peer-reviewed literature than do the traditional macroeconometric models. (Which is not to say that the supply-side models are any better than Keynesian models.)

The academic literature of recent decades has not abandoned the use of data. Economists with a particular model of the economy will "calibrate" that model using observed values of variables in order to be able to illustrate numerically how the model works. However, the project of trying to produce a single statistical model that definitively represents the behavior of the economy is no longer considered worth pursuing.

* * *

Macroeconometric models of the sort used by Christina Romer and Jared Bernstein to project the impact of a fiscal stimulus are pure fabrications. To get the multiplier of 1.57, a computer cranks through the equations first under a baseline scenario, and then under an alternative scenario with an assumed increase of government purchases amounting to 1.00 percent of GDP. This process contains no more information than if Romer or Bernstein were to type the number 1.57 into a computer and then print it out. Either way, what we know is that someone's a priori theory, or rather conclusion, was entered into a computer, and the number 1.57 appeared on the printer.

The next step in this process comes one year later, when credulous legislators and journalists ask the modelers how many jobs were created by the stimulus. The modelers simply repeat the exercise of simulating their models with and without the stimulus. Lo and behold, the computer gives the same answer that it did before. The press reports that the stimulus worked exactly as planned, notwithstanding that the unemployment rate turned out to be 2-1/2 percentage points higher than was predicted.

If macroeconometric models are fabrications, then where does that leave us? Imagine that two hundred years ago, you questioned the use of bleeding by physicians. Many doctors might share your concerns, particularly in private. But then the doctors would say, "Well, what can we do? We can't just stand by and appear helpless when the patient is so ill.

Unless you can show us something better, we will just keep doing what we are doing." And so it is with macroeconometric models.

Like the Swiss hikers lost in the Alps, policy makers are in need of a map. The macroeconometric models purport to represent a map, but it is unlikely that their map refers to the Alps. If the policy makers knew that it might be the wrong map, they might not use it with such confidence.

NOTES

1. The story is related in a video at http://www.edge.org/3rd_culture/kahneman_taleb_DLD09/ kahneman_taleb_DLD09_index.html
2. One of the first articles to highlight this issue was Nelson and Plosser 1982. There is an extensive follow-up literature.
3. Mankiw (2006, 15) writes: "Recent developments in business cycle theory, promulgated by both new classicals and new Keynesians, have had close to zero impact on practical policymaking."

REFERENCES

Akerlof, George A., and Robert J. Shiller. 2009. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton: Princeton University Press.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24(2) (Spring): 3–30.

Council of Economic Advisers. 2010. "The Economic Impact of the American Recovery and Reinvestment Act of 2009." 14 July. http://www.whitehouse.gov/files/documents/cea_4th_arra_report.pdf

Krugman, Paul. 2009. "A Dark Age of Macroeconomics." January 27. http://krugman.blogs.nytimes.com/2009/01/27/a-dark-age-of-macroeconomics-wonkish/

Lucas, Robert E., Jr. 1976. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy* 1(1): 19–46.

Mankiw, N. Gregory. 2006. "The Macroeconomist as Scientist and Engineer." Harvard Institute of Economic Research Discussion Paper No. 2121. http://papers.ssrn.com/sol3/papers.cfm?abstract_id = 916527

Nelson, Charles R., and Charles R. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10(2): 139–62.

Romer, Christina, and Jared Bernstein. 2009. "The Job Impact of the American Recovery and Reinvestment Plan." January 9. http://otrans.3cdn.net/45593e8ecbd339d074_l3m6bt1te.pdf