Report on David Colander's "Economists, Incentives, Judgment, and Empirical Work" by Kevin D. Hoover

General Comments

In the interest of full disclosure, I should say that I am a British-trained macroeconomist whose professional career has largely been in the United States. I am an advocate of what Colander refers to as the "European" approach – both in its cointegrated vector autoregression (CVAR) and London-School of Economics (LSE) forms. In fact, I have contributed substantively to the LSE form.[1] And I have helped in small way to proselytize for the CVAR approach.[2]

Thus, Colander and I advocate for the same side as far as macroeconometric methodology goes. Yet, I am largely unconvinced by this paper.

The argument of the paper seems to be this. There are two established approaches to macroeconometrics – European and American. The European approach represents craftsmanship and requires the careful application of laborious and hard-to-master techniques and great sensitivity to the data. The American approach can be implemented using by poorly trained technicians on readily available data by pushing the buttons on econometric software. Americans deploy their techniques in the service of highly simplified *a priori* theory, while Europeans – at least the ones that have not become Americanized – learn about the complexity of the economy from the data. The dominance of the American approach is explained by its ease of implementation and by the fact that American economists are under much greater pressure to publish than was – at least until recently true in Europe.

I just don't buy it.

I grant that there are real differences between Europe and North America. Within Europe there are a number of indigenous approaches to macroeconomics and macroeconometrics that are distinct from the dominant approaches in the U.S. What Colander frequently calls the "European general-to-specific approach" is not *the* European approach. There is variety in Europe. In fact, Colander conflates two related, but nonetheless not identical, methodologies: the CVAR methodology of Johansen and Juselius, among others, and the LSE approach of David Hendry and many colleagues, which includes, but is not fully defined by, Hendry's general-to-specific methodology. Indeed, Juselius has been a pointed advocate of specific-to-general approaches in the CVAR context. Properly construed, this need not conflict with Hendry's approach, but it shows how tricky the labels are. Equally, there is no single American approach:

---

[1] See, particularly, Hoover and Perez, "Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search," *Econometrics Journal* 2(2), 1999, pp. 1-25, which is the ancestor of Hendry and Krolzig's *PcGets* and Doornik and Hendry's *Autometrics* automatic specification search programs.

[2] See Hoover, Johansen, and Juselius, "Allowing the Data to Speak Freely: The Macroeconometrics of the Cointegrated Vector Autoregression," *American Economic Review,* May 2008.

structural estimation, structural vector autoregressions (SVAR), and calibration approaches represent quite different views about how to do econometrics. There is communication not only among the advocates of these approaches, but also to a lesser extent between the American approaches and the European approaches that Colander advocates. Nonetheless, some blurring about the edges eliminates neither the differences nor the variety.

Colander argues that Americans are badly trained – they use techniques that they don't understand mechanically. The evidence for this is circular, in that he implicitly defines understanding the techniques as deploying them in a "European" manner. If they fail to do so, they clearly don't understand. But he offers no independent evidence on the quality or nature of the instruction in American graduate schools or on the mastery of any techniques by American practitioners. Colander is well known for his investigations into graduate education in economics. But that evidence is at most lightly referred to here and probably needs to be seriously amplified to make it specifically relevant to the issue at hand.

Colander also argues that Americans start with theory, because theory is easier, and the application to data is merely window dressing. It is not clear that relevant theory is easier – different problems arise; different skills are needed. I see, for instance, dynamic stochastic general equilibrium (DSCG) modelers employing vastly difficult computational methods to increasingly complex theoretical models. Personally, I think that their efforts are misdirected, but it is absurd to think that they are easier than doing econometrics in a "European" mode. And while we all know work in which quantitative applications are window dressing, I have seen a large amount of work in which the quantitative applications are painstaking, time-consuming, and taken very seriously. To make this point stick, Colander has to do either (preferably both) detailed case studies and a statistical investigation.

We must also admit that the pressure to publish is high in the United States and that, most probably, it is higher than in Europe – at least until recently. That is not sufficient, however, to explain the differential success of different approaches. The incentive to publish does not operate in a vacuum. Colander's explanation requires that American approaches be easier to implement than European approaches – a point claimed but not demonstrated. Yet, even if it were demonstrated it would be insufficient. For publication is not one-sided. I must not only produce a paper, but must also get it accepted. So Colander must also explain why top journals accept what he regards as inferior products. The journals are under a quite different pressures than are the authors – namely, to publish papers that are good for their reputation. He needs to demonstrate why inferior work is held in esteem by the editors and referees of the journal, and the mere pressure to publish cannot in itself explain that.

Colander's explanation here has the same form as one that says that the profit motive places firms under pressure to cheapen production – producing inferior products being one way to cheapen – so that we should expect inferior products. (Marx goes on in *Capital* about the adulteration of bread to cheapen production in exactly this way.) But a firm has to sell its products, and consumers don't want to buy inferior products. So surprisingly to some – but it should not surprise an economist – products frequently get better in quality, not worse, precisely because of the pressure to cheapen under the constraint of salability (certainly true for bread in the U.S., which is vastly better today on average than in 1960). Why does not the similar process work here?

Colander writes as if publication pressure alone implies bad work and as if publication pressure is necessarily a bad incentive. And even sometimes, as if one must have good motives to do good work – thus, the romantic distinction between the leisurely European econometric craftsman and the frenetic American technician. But the motives of scientists are famously mixed. Watson in his well known *Double Helix* comes close to characterizing his own motives as just wanting to a win a Nobel prize. But the effective way to do so was to make a genuine contribution to science. And an economist who just wants to get published may well make a genuine contribution, especially if the way to get published requires good work. Incentives can certainly misdirect effort. But they do not do so necessarily. It all depends on whether they are properly aligned and within what other context or constraints they operate. Publication pressure is constrained by the editorial process. Just talk to colleagues about how hard it is to get a paper published in a "good" journal. Editors reject papers and they call for papers to be substantially revised. What determines the standards that they apply? Publication pressure cannot answer that question. And it cannot, therefore, answer what it is that makes adopting an "American" methodology more likely to succeed than adopting a "European" methodology.

Let me offer an alternative conjecture. The real difference between Europe and the United States is not a direct product of publication pressure but of genuine intellectual differences. Colander cites the "European" approach as best suited to dealing with the genuine complexity of the economy. I happen to agree, but not everyone does. Why not?

What Colander refers to as the theory-first approach goes back, at least, to John Stuart Mill. Mill argued that the economy was too complex for controlled experiment – the only sort of inductive approach that he recognized. He argued that instead economics must start with first principles known by direct acquaintance with our own minds and proceed deductively. Comparison to empirical data is at best a cross-check on the practical relevance of these theoretical deductions. A not dissimilar view developed in the 1940s with the Cowles Commission. Business cycle research in the 1930s had thrown up a number of problems in trying to disentangle the underlying autonomous processes that drive the macroeconomy. Frisch came to the view that the problem was insoluble on the basis of data alone and advocated that *apriori* economic theory had to play an important role. This was developed by the Cowles Commission into the canonical methodology of theoretical identification and statistical estimation that forms the basis for the development of American econometrics. This is where "theory first" comes from – not from its ease or cheapness, but from a reaction to the overwhelming complexity of the economy. So, while I agree with Colander that the CVAR and LSE approaches are reactions to complexity, so is the "theory first" approach – it is just a different reaction. (Interestingly, the dominant voices in the Cowles Commission were

European émigrés; so “European” and “American” are geographical not personal designators.)

The Cowles Commission methodology has been challenged and transformed in various ways. There are independent debates about the nature of the appropriate theory (e.g., highly simplified (for example, Friedman, who rejected the Cowles Commission view) or highly complex (for example, Klein, who embraced it).[3] But in the end, it is the after-effect of the Cowles Commission that explains the development of American econometrics, including calibration. (I have myself argued that “theory first” versus “data first” explains some of the characteristic disagreements in econometrics.[4]

The CVAR and LSE approaches are much less affected by the Cowles Commission than by Fisherian statistics. These different roots, rather than publication pressure, explains why econometrics is taught the way it is in different countries, why different aspects of research seem to be more critical in the different traditions (getting the theory right or getting the statistics right), what sort of work the best practitioners aspire to, and what sort of standards the best journals apply when they are working at their best.

While the essential differences are intellectual, there still is – and I think that this is true in Europe as well as America – a lot of bad work published, because of dysfunctional editorial processes, mendacity, and ignorance. What I am arguing is that, even if we set these to one side, there is still a debate. Publication pressure does not explain the debate; removing publication pressure will not resolve the debate; and indeed there is no evidence that removing publication pressure would favor the European side of the debate.

Detailed Comments

p. 2, para. 1: “theory first” not uniquely American nor modern. Goes back, at least, to Mill. (See General Comments.)

p. 3, para. 1: a) is it not the case that Italy and France still have national competitions for university posts, which, while they might rely on “vague and idiosyncratic” judgments, are nonetheless not “informal”?

b) the Ph.D was originally, I believe, a German degree – certainly not an English one – and it is only in the postwar period that it became fully established in England. The claim that “until recently” no Ph.D was needed depends a great deal on what is meant by “recently.” Certainly a Ph.D has been the normal path to an academic career in economics in the U.K. for several decades. The suggestion that only the second rate took them earlier needs further evidence.

[3] Consult Marcel Bouman’s *How Economists Model the World in Numbers*, Routledge, 2005 on this history.
[4] See Hoover, “Lost Causes,” *Journal of the History of Economic Thought*, 26(2), 2004, pp. 149-164.

p. 3, para. 3: I am probably in the same boat as many readers in not knowing what the nature and substance of "the European common educational policy" is.

p. 3, para. 4: a) while it is perhaps creeping Americanization in spirit, European countries appear to be more apt than Americans to establish canonical ranked lists of journals to be used in formal assessment exercises. My own experience in American promotion and hiring processes is that, while journal quantity and perceived quality matter, rankings are in fact contestable, not only department by department, but case by case. Appeals are made to various published rankings (which frequently conflict), but, unlike in many European countries, these seldom have an official status. Everybody knows which journals are in the top 10; it is just that everybody's top ten list is somewhat different.

b) the notion that economists in the U.S. don't address the big questions anymore or that, before the incentive system got established, economists paper per paper addressed more important issues, calls out for some evidence – for example, comparison of journal articles across countries and time. In any case, is the assertion here about macroeconomics or about economics in general? The paper appears to focus on macro, but the incentives claimed to be at work are general. As long as we are trading in anecdotes, I have seen many papers by U.S. authors that are deep in data, serious about the relationship of theory to data, and painstaking in their work. This work is not mechanical, but thoughtful. Barring evidence to the contrary, and Colander presents no evidence, I maintain my belief that the real differences have to do with conceptions about the relationship of theory to evidence and not about incentives to publish. As for work not addressing big questions. In most sciences at most times most work does not address big questions. About that, Kuhn was surely right. The progress of science is mainly through the pedestrian plodding of fairly ordinary folk within a conceptual and methodological framework. He called it *normal science*. Most economics is normal science. And as with other normal science, substantial progress is made by focusing on the little questions and not being worried by the big ones.

p. 4, para. 2: a) My own experience in marketing my students and hiring the products of other programs makes me question the generality of the anecdote about blowing off two of three thesis papers.

b) The dichotomy drawn here between training "general researchers" and "highly efficient journal article writers" is suspect. At the least, Colander needs to show how the absence of general research skills makes a student into a more efficient writer of journal articles, recalling that it is not enough to write them, but that one must also get them through the editorial process. The incentive of a student may be just to get published, but the incentive of the journal is not just to publish anything.

p. 4, para. 3: again, I doubt the generality of the books-are-bad-for-your-career anecdote:

i) counter-anecdote: my first book, *The New Classical Macroeconomics*, made my career;
ii) many economics books are written, and many are written by top guys, not only

after they are firmly established;
iii) as the former chair of the UC Davis Economics Department I did a study of the factors governing merit advancement in the Economics and the Agricultural and Resource Economics Departments (the University of California merit system requires regular, frequent review of even tenured, senior faculty). Journals were divided into three tiers based on a recently published ranking and articles in edited volumes or other non-journal outlets were counted as third tier. Taking the second-tier journal as the standard unit, the influence of various publications on merit was well approximated by an index with the weights: 5 for books, 3 for top tier (top 10) journals; 1 for second tier journals (ranked 11-30); 1/3 for all other articles in journals or other outlets. Books matter. This is just one department, but I don't think that it was all that idiosyncratic.
iv) to say to a younger economist that efforts are better spent on articles than books is probably good advice – more immediate visibility, lower risk because of greater diversification, lower risk because of time delays, etc. That does not mean that the same advice applies to older economists.

p. 4, para.4: a) both the U.S. and the "European" systems are said to have advantages and disadvantages. Colander does not make clear the advantages of the U.S. nor the disadvantages of Europe. If this is not just a throw-away remark, a more searching discussion is in order.

b) surely it is not a general theorem that any time a measure is used for assessment "it distorts choices and does not measure what it is supposed to measure." That really depends on the alignment of the measure with the goals. If they are properly aligned, no problem. The real problem is when incentives lead to the maximization of a measure that does not (or not quite or not completely) measure the true target.

p. 5, para. 2: as a student in the U.K. in the late 1970s and early 1980s, I can attest to the deep engagement of British economics with rational expectations and the new classical economics. If "primarily U.S. phenomena" means that Americans more than Europeans adopted the Chicago/Minnesota paradigms, that is surely true. But the British, at least, certainly did react and examine the implications of these developments.

p. 5, para. 3: "researchers create an analytically solvable model that captures some of the key features of the key elements of the economy, and then use that model for thinking about macro policy." That could easily describe John Hicks's IS-LM model. And I could think of many other examples. So, it is hard to see this as differentiating, in itself, Americans from Europeans. What's more, the claim that Americans adorn their models with quantitative clothing after the matter is settled theoretically in advance is not really closely related to that characterization of models.

p. 6, top: Solow quotation needs more context. What is the evidence that this "swindle" is related to the pressures to publish.

p. 6, para. 2: "*the* European approach" is really *a* European approach:

i) Americanization has already gone a long way in European economics;

ii) there are a variety of European approaches. Colander is singling out the cointegrated vector autoregression (CVAR) approach of Johanssen and Juselius, which is surely more highly regarded in Europe than in the U.S., but which is not the uniform view of all Europeans – even of the not Americanized;
iii) indeed, there are differences even between the CVAR approach and its closer allies, such as the LSE methodology of David Hendry and others.

p. 6, para. 4: conflation of the European (CVAR) approach with the general-to-specific approach of Hendry is not helpful. The CVAR approach is, in fact, often used in a specific-to-general way, building up from a few cointegrating relations to a larger set (see Juselius's 2006 book and Hoover, Johanssen, and Juselius's 2008 *American Economic Review* paper.) General-to-specific is an important element of Hendry's methodology, but it does not by itself adequately define that methodology. Aspects general-to-specific can be used even in the CVAR context, but they are not the same.

p. 7, para. 4: "There is nothing technically wrong . . ." Of course there is something technically wrong as the situation is described here. Colander seems to have in mind that a set of assumptions $\{A_1, A_2, \ldots A_n\} \rightarrow C$, a conclusion; that C is compared to the data; and that, on some metric, it fits. From this, he concludes that there is nothing technically wrong. But as a point of logic, $\{A_1, A_2, \ldots A_n\} \rightarrow A_1$ and, indeed, each A_i. So, if A_1 is shown to be false and if A_1 is essential to the derivation of C, then finding that C fits provides no support for the theory (here construed as the set of assumptions). It is the fallacy of affirming the consequent. That is just a mistake of logic – and quite technically wrong. Notice that this is not the same as offering a counterfactual. It may well be correct that, if $\{A_1, A_2, \ldots A_n\}$, then C. That is just a logical deduction that may be counterfactual and true. But as soon as we try to use the relationship of C to the data, we have made a technical mistake if a *relevant* assumption is *relevantly* false.[5] Real science has to put a lot of weight on the "relevant" and the "relevantly" here to account for approximation, idealization, simplification, which are necessary to deal with complexity. Friedman is frequently tarred with having supported the view that Colander seems to accept as "technically" correct, even as he argues against its desirability. That is summed up in the frequently stated view that Friedman said that the truth of the assumptions does not matter. This is, I have argued elsewhere a misreading of Friedman, who is actually concerned with understanding the "relevant" and the "relevantly."[6] Colander advocates Marshall as a touchstone for his complexity approach (p. 6). Nobody is so Marshallian as Friedman, and Friedman correctly sees Marshall as an advocate of simplification and limiting the scope of inquiry as a way of dealing with complexity. Friedman takes the empirical data seriously, so that despite many profound methodological differences, there is more kinship between Friedman and the CVAR approach and Hendry's LSE approach, than Hendry or Friedman (were he still with us) would be comfortable in acknowledging.

[5] See Hoover, "Methodology: A Comment on Frazer and Boland, II," *American Economic Review, September* 1984.
[6] Hoover, "Milton Friedman's Stance: The Methodology of Causal Realism" http://www.econ.duke.edu/~kdh9/research.html

p. 7, para. 5: Ireland's motives are, I think, not known to us. Colander overreaches in attributing the motive of just wanting to get a published paper without more substantial evidence. Individual case studies are, in fact, unlikely to ever provide satisfactory evidence for Colander's thesis. A more statistical approach might give some insight. After all, motives tend to be mixed and mixed motives may be no barrier to good science (see General Comments above).

p. 8, para. 1: a) it is pretty easy to tell a $10 from a $100 bottle of wine – look at the price tag. What a wine critic brings is the ability to tell a good wine from a poor one The price is a marker – but an imperfect marker – of the quality.

b) The problem is that the reviewing system does not insist on the standards that Colander would like to see. But it could so insist. There is nothing intrinsic in the reviewing process that would prevent it. It is rather that the profession has not yet been convinced that these are the standards that need to be enforced. God knows it enforces all sorts of other standards.

p. 8, para. 2: the cited critiques are a mixed bag and don't really support Colander's push-for-publication diagnosis, even if he finds them congenial on other points. Personally, I find these critiques to be seriously deficient and have argued against them directly or indirectly in various articles.[7]

pp. 8-11, new section: the section really addresses Johansen not Hendry. There situations are not closely similar. Johansen is a primarily a statistician who comes to economics through collaboration with economists. Most of his citations are, in fact, to his theoretical work on tests for cointegration. While it is true that he advocates a methodology closely related to Hendry's, that does not explain the bulk of the citations. Hendry, on the other hand, while he does contribute to econometric theory, is *principally* an applied econometrician and an advocate of an econometric methodology. The bulk of his citations derive from applications and methodological contributions rather from theory. There are two quite different success stories to explain here. They no doubt require different explanations.

p. 11, para. 2, last line: "the experts have better things to do with their time." Such as? Since using time this way would improve science, the real questions are 1) why the incentives militate against this use of their time? 2) is there another activity that actually has higher marginal payoff?

[7] Hoover and Perez "Three Attitudes towards Data-mining," *Journal of Economic Methodology*, 7(2), June 2000, pp. 195-210; Hoover and Perez, "Truth and Robustness in Cross-country Growth Regressions," *Oxford Bulletin of Economics and Statistics* 66(5), 2004, pp. 765-798; Hoover and Siegler, "Sound and Fury: McCloskey and Significance Testing in Economics," *Journal of Economic Methodology*, March 2008 and Hoover and Siegler, "The Rhetoric of 'Signifying Nothing' : A Rejoinder to Ziliak and McCloskey," *Journal of Economic Methodology*, March 2008.