SOCIAL EPISTEMOLOGY, 1993, VOL. 7, NO. 3, 245–248

Open peer commentary on the scientific status of econometrics

Comment

ARTHUR M. DIAMOND, Jr.

On the side of the Social Science Research Building at the University of Chicago is a shortened version of this quote from Lord Kelvin: 'When you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre [*sic*] and unsatisfactory kind'. Knight used to say that the way the dictum was implemented by Chicago economists was: 'if we cannot measure a thing, go ahead and measure it anyway' (Merton *et al.* 1984). In spite of Knight's sarcasm, his student George Stigler took Kelvin's message to heart. Stigler believed in theoretical clarity, but above all else he believed in systematic empirical evidence. This view, shared by most of the luminaries at Chicago during the Friedman-Stigler era, was once sung to a startled audience by Stigler's friend George Schultz (then Director of Management and Budget, later Secretary of the Treasury and Secretary of State):

> A fact without a theory
> Is like a ship without a sail,
> Is like a boat without a rudder,
> Is like a kite without a tail.
> A fact without a figure
> Is a tragic final act.
> But one thing worse
> In this universe
> Is a theory without a fact.
> George Schultz (1973).

Feigenbaum and Levy are solidly in the old Chicago tradition: boldly quantifying where no academic has quantified before. Economists frequently seek to make institutions more efficient. However, too seldom do they focus on the institutions that produce the knowledge that is the main input to human progress and economic growth. Before institutions can be creatively reformed, it may be useful to know more about how they currently function. The authors think hard about what the incentives are in the current system of journal refereeing and academic career advancement. After thinking hard, they roll up their sleeves and get their hands dirty in the attempt to dig up relevant data to test their conjectures. In doing this they are creative and diligent and courageous (but not rich and famous).

One of the authors, David Levy, once wrote a paper entitled 'The market for fame

Author: Arthur Diamond, Frederick Kayser Professor of Economics, University of Nebraska, Omaha, NE 68182, USA.

and fortune' in which he argued that academics, like everyone else, are seeking fame and fortune. Since this is a journal well-known to be open to reflexive exercises, it might be interesting to inquire whether Feigenbaum and Levy are efficiently pursuing what Levy claims everyone pursues.

Although the authors discuss with sophistication and *élan* the faults of the reward structure of the economics profession, I suspect that the paper will not be well-received by the profession for several reasons.

The authors are examining the profession itself. This is viewed as non-serious by much of the economics profession. The editor of one of the leading journals of economics routinely rejects such articles by smugly saying that his journal only accepts 'economics' and not the 'sociology of economics'.

The authors emphasize the flaws in the current reward structure of the economics profession. There is a receptive audience to such a message, mainly on the fringes of the profession, but those who are receptive to this message tend largely to be very unsympathetic to mathematical modelling and sophisticated econometrics. The audience that appreciates the subject and message of Levy and Feigenbaum's paper will be antagonistic to its method.

The authors do not achieve many 'positive' results, in the sense that in the paper's most important regression (Table 4) only a single variable is statistically significant (even though only three independent variables are included in the regression). Ignoring McCloskey's complaints (1985), the economics profession judges that if a variable is shown to be statistically significant at a traditional level (usually 0.05), then the variable has explanatory significance. As Leamer (1983) and others have emphasized, this leads to the re-running of regressions with many specifications, and then reporting the one or two that look 'best' (in the sense of having the highest number of significant *t*-statistics). Studies where this is not done are frequently unpublishable (or published in less prestigious journals). The profession seems to assume that evidence that something matters is more interesting or convincing than evidence that something does not matter.

I conclude that Feigenbaum and Levy's paper will not be well-received by the economics profession and, hence, will not put them on the sure path to fame and fortune. Although I believe that the profession grossly undervalues what Feigenbaum and Levy are doing, even I, if I look hard, can still find a few points to take issue with. Unlike with profession, I like the picture they are painting and, in broad strokes, the way in which they are painting it.

I have a few suggestions, however, on how to improve the picture's shading and tone. For instance, I would have preferred to have seen more space given to empirical detail, and perhaps a little less to theoretical elaboration. On the theory, although my papers are graciously (and correctly) cited as evidence for a small reputation effect, I am not sure that this result is a general one. I argue in those papers (Diamond 1988, 1992a, 1992b) that the polywater case is one where the mistake was a reasonable one, fully justifiable even for normally careful researchers. The finding that there is no large reputation effect in this case is not sufficient evidence to conclude that 'a radical devaluation would occur only in cases where it is believed that replication discrepancies did not result from carelessness, but from fraud'. I believe that major carelessness may indeed have a significant reputation effect. This is an issue that should be settled by further impirical investigation and, to their credit, Feigenbaum and Levy are currently conducting just such an investigation (Feigenbaum and Levy 1991). On the other side of the issue, Aloysius Siow in 'Are first impressions important in academia?', has presented some arguments, and empirical evidence, that reputation effects can be very strong, even when the only misdeed is to have written an uninteresting paper (Siow 1991).

I also do not see how the authors can claim that 'the benefits of data publication are an "all-or-nothing" proposition'. Does not the publication of some data and summary statistics have some value to the would-be replicator? Surely if the hardest-to-find data are published this makes it easier to replicate. In addition, publishing summary statistics increases the likelihood that a would-be replicator can judge which replications will be fruitful (based on reported summary statistics that 'don't seem right'). One of the most important lessons that I learned at the Applications Workshops at the University of Chicago, was from Nobel Prize winner Theodore Schultz. In workshop after workshop he would gently direct our attention to the tables of descriptive statistics to point out some mean that was too high or some variance that was too low, indicating a likely problem in the presenter's data or statistical method.

In addition, I would have liked to see some more descriptive statistics in the Feigenbaum and Levy paper itself. I would like to know something about the average length of the papers, and how many of them appeared in top-ranked journals. In the regression reported in Table 4, the authors find, counter to their expectations, that RANK of journal does not matter in explaining the MAD measure of failure to replicate. The authors note earlier that the articles that make up the data set appeared in 'first-rate journals from the 1960s and early 1970s'. Perhaps the RANK variable is insignificant because it exhibits too little variation in the selected sample, rather than for any more substantive reason. If we had been provided with a table of descriptive statistics we could judge for ourselves whether the 'perhaps' should be dismissed or accepted (and whether a replication with a broader data set would be likely to yield different results).

Near the beginning of my career, I told George Stigler (Levy's mentor) that I was going to be involved with the Center for Human Resource Research at Ohio State. This was a Center that was doing standard, mainstream research on standard, mainstream topics within labor economics. I remember him advising me that I should not spend too much time there because 'it won't be any fun'.[1]

Fun? But what of fame and fortune? Aristotle (I. vii.) claimed that the key to happiness was not to seek happiness directly, but to do other things well. Similarly, it may be (reflecting a higher order of fun) that fame and fortune go to those who do not seek them directly, but who pursue the fun topic and the neat idea.[2] If they break through, those who teach us something we did not know, and at first did not even want to hear, may be the ones most highly esteemed. We may not know whether Feigenbaum and Levy will ever achieve fame and fortune, but it is certain that they are having good fun. In the process they are making fundamental contributions to our understanding of the reward structure of science.

Notes

1. I am not sure that I correctly remember Stigler's exact words, but I am sure of the gist of his comment and that the word 'fun' was part of it.
2. Munevar (1981, p. 69) and others have argued that scientists frequently have a sense of playfulness. Adam Smith wrote of scientists having a sense of curiosity (1976, p. 124). Richard Feynman (1985, 1988) seems to have been a contemporary exemplar of both. Feigenbaum and Levy are too.

Bibliography

ARISTOTLE. *Nicomachean Ethics*, Harvard University Press, Cambridge, Mass. (1934).

DIAMOND, A. M., Jr. 'The polywater episode and the appraisal of theories', in *Scrutinizing Science: Empirical Studies of Scientific Change*, A. DONOVAN, L. LAUDAN, and R. LAUDAN (eds), Kluwer Academic Publishers, Dordrecht, Holland (1988), pp. 181–198.

DIAMOND A. M., Jr. 'The career consequences for a scientist of a mistaken research project' (working draft, 1992a).

DIAMOND, A. M., Jr. 'The determinants of a scientist's choice of research projects', in *Scientific Failure*, T. HOROWITZ and A. I. JANIS (eds), Rowman & Littlefield Publishers Inc., Savage, MD, (forthcoming 1992b).

FEIGENBAUM, S. and LEVY, D. M. 'Fraud, error & reputational capital'. Paper presented at the Southern Economic Association meeting, 22 November, 1991.

FEYNMAN, R. P. *Surely You're Joking, Mr. Feynman!*, W. W. Norton, New York (1985).

FEYNMAN, R. P. *What Do You Care What Other People Think?*, G. K. Hall & Co., Boston (1988).

LEAMER, E. 'Let's take the con out of econometrics', *American Economic Review* 73 (1983), pp. 31–43.

LEVY, D. M. 'The market for fame and fortune', *History of Political Economy*, 20, No. 4 (1988), pp. 615–625.

MCCLOSKEY, D. N. 'The loss function has been mislaid: The rhetoric of significance tests', *The American Economic Review*, 75 (May 1985), pp. 201–205.

MERTON, R. K., SILLS, D. L., and STIGLER, S. M. 'The Kelvin dictum and social science: An excursion into the history of an idea', *Journal of the History of the Behavioral Sciences*, 20 (October 1984); pp. 319–331.

MUNEVAR, G. *Radical Knowledge*, Hackett Publishing Co., Indianapolis (1981).

SCHULTZ, G. As quoted in 'Another professor with power', *Time* (26 February 1973), p. 80.

SIOW, A. 'Are first impressions important in academia?', *Journal of Human Resources*, 26, No. 2 (Spring 1991), pp. 236–255.

SMITH, A. *The Theory of Moral Sentiments*, Clarendon Press, Oxford (1976) (first published in 1759).