

Replication in Labor Economics: Evidence from Data, and What It Suggests

By DANIEL S. HAMERMESH*

There is a paucity of pure “hard-science” style replication in applied economics (Hamermesh, 2007), although a few role models have used data that others had made available to re-examine conclusions that had appeared solid enough to pass muster with editor(s) and referees. My questions here are whether the profession can or should produce more pure replications, whether the market for economic research has created good substitutes for formal replication, and what the incentives are for this kind of work. To provide an empirical basis for the discussion I first examine the citation histories of ten leading articles in empirical labor economics published between 1990 and 1996, thus with at least twenty-year citation histories. This sample allows examination of ways in which earlier scholarship affects or stimulates subsequent research and consideration of how these impacts change as time passes since the original article appeared.

I. The Pathways of Scholarly Influence

I selected ten of the most heavily-cited publications in empirical labor economics from this period, with the selection criteria requiring each to have been published in a so-called “Top 5” journal and to have accumulated at least 1000 Google Scholar (GS) citations as of Summer 2016.¹ The studies, along with the number of GS and Web of Science (WoS) citations of each and its rank among all articles published in that volume of the journal, are listed in Table 1. As the Table shows,

*Royal Holloway University of London, Egham, UK TW20 0EX, and IZA

(Daniel.Hamermesh@rhul.ac.uk). I thank Frances Hamermesh, Andrew Leigh and David Sibley for helpful comments.

¹Several other articles met the criteria but were excluded as being as much methodological as making a purely empirical point.

each was sufficiently well-cited to be in the top 10 percent of all articles published in the particular volumes of the journals, themselves the most heavily cited in the economics profession (Fourcade, 2015; Hamermesh, 2017). For each of these articles I examined every publication that the WoS had recorded in September 2016 as having cited the work, in each case reading first the abstract and then, if necessary, skimming through the citing paper itself. I classified each citing article by year post-publication of the principal article and by whether it was: 1) Related to; 2) Inspired by; 3) Very similar to but using different data; or 4) A direct replication at least partly using the same data.² These classifications are obviously arbitrary, but since one person (this author) made all the classifications, at least they should be consistent across the ten articles.

The final column of Table 1 shows the percentages of citations to each paper that were merely related to the new article (Category 1 above). The overwhelming majority of citations to these highly-cited papers were based on their important roles in the relevant literatures. Few of the citing papers were inspired by the original paper, and fewer still involved a replication. The distribution of the over 3000 citing papers in the four categories was: Related, 92.9 percent; inspired, 5.0 percent; similar, 1.5 percent; replicated, 0.6 percent. (The ranges in the last three categories are 2.3 to 7.0 percent; 0 to 5.2 percent; and 0 to 2.3 percent.) Replication, even defined somewhat loosely, is extremely rare even of these most highly visible studies.

The life cycle of replication is also interesting. Figure 1 shows a scatter and linear fit of the annual fractions of citing articles that are similar to the original article or are direct replications (Categories 3 and 4) by post-publication year of the original article, leaving out years 21+ so that

²To be classified as “inspired” the citing paper had to refer repeatedly to the original paper and/or had to make clear that it was inspired by the methodology of the original work. To be noted as “similar” the citing paper had to use the exact same methodology but on a different data set, while a study classified as a “replication” went further to include at least some of the data in the original study. Thus even a “replication” in many cases involved more than simply re-estimating models in the original article using the same data.

citations to all ten original articles appear in the annual averages. The probability that a citing article is either similar to or explicitly replicates the original paper diminishes over the years after the original publication. Moreover, adding a quadratic term to the fit adds nothing—the relationship is linearly decreasing. Estimating the same equations (using either least squares or probit analysis) on the underlying micro data yields almost identical coefficients to the linear model estimated over averages across post-publication years.

Figure 2 presents a similar scatter for the fraction of citing articles that were classified as inspired by the original piece. A quadratic model in post-publication year fits this scatter better than a linear model: This relationship is significantly convex. As with the citing articles that were similar or replications, the results are almost identical if we estimate these models over the underlying micro data.³

One might be concerned that the relative paucity of replications of these important papers results from selectivity—less important papers were replicated, found wanting and henceforth essentially ignored. To examine this possibility I collected citations to the five least-cited empirical articles in labor economics published in these four journals between 1990 and 1996, classifying each citation to them in one of the same four categories used for major articles. One of these five had 155 GS citations, but references to the other four ranged from 37 to 89 GS citations. 92.8 percent of the 111 WoS citing articles were merely related to the original article (Category 1), almost identical to the percentage of articles in this Category that cited the ten major papers. Not one of the WoS citations to these five articles could be classified either as similar to or a direct replication of the original study. Sparsely-cited articles in major journals are not killed by replications that cast doubt on their results; rather, they “die” from neglect.

³The results using the micro data underlying the scatters in Figures 1 and 2 are only barely altered if we include fixed effects for each original article.

II. Implications for Replication

Does this evidence show that the replication glass is 2 percent full, or 98 percent empty? Replications are not published for most studies, even those published in Top 5 journals, nor should they be: The majority of articles in those journals are, as I showed (Hamermesh, 2017), essentially ignored, so that the failure to replicate them is unimportant. Even as a fraction of citations to major papers, replication is quite rare; but 8 of the 10 articles examined above were replicated at least 5 times, so published replications of these most heavily-cited papers are performed; and an unknown number of additional replications may have been made but never published, perhaps because they corroborated the original results.

Replication in the case of labor economics (and presumably of other areas of applied microeconomics) does not take the form of repetition of a particular protocol on laboratory rats (or their equivalent, undergraduate students). Instead, in most cases it proceeds by taking the economic idea that motivated the original empirical study in a literature and examining its predictions and implications using a set of data describing a different time and/or economy. Applied microeconomics is not a laboratory science—at its best it consists of the generation of new ideas describing economic behavior, independent of time or space. The empirical validity of these ideas, after their relevance is first demonstrated for a particular time and place, can only be usefully replicated at other times and places, since if they are general descriptions of behavior they should hold up beyond their original testing ground. Simple laboratory-style replication is important in catching errors in influential work, as in the recent case of Reinhart and Rogoff (2010); but the more important replication goes beyond this and is, as I have shown, usually undertaken for the most important work in labor economics.

I believe, and the evidence suggests, that the system is not broken and does not need fixing. Despite this evidence and these arguments, what if one firmly believes that more replication, using mostly the same data as in the original study, is necessary? First, a bit of history: During the 1960s the *American Economic Review* was replete with replication-like papers, in the form of sequences of Comments (often in the form of replications on the same or other data), Replies and even Rejoinders.⁴ For example, in the four regular issues of the 1966 volume nearly 40 percent of the articles were of this form. In the first four regular issues of the 2013 volume only 10 percent of articles were Comments, etc., reflecting a change that began by the 1980s. The editors have shifted away from cluttering the *Review*'s pages with Comments, etc. This change may have reflected their desire to maximize its impact on the profession in light of their conscious or subconscious realization that pages devoted to this type of exercise do not generate the same attention from other authors' scholarly work as does journal space devoted to original articles (Whaples, 2006). We have had replications or approximations thereof in the past, but the market for scholarship—as indicated by impact—has exhibited little interest in them.

Perhaps this market-based evidence showing a decline in publishing of replication-like work is irrelevant, either because it is out of date or simply not on point? If one believes this, a possible method for encouraging replication would be a statement by the AEA urging its editors, and those of other journals, explicitly to encourage replications among their submissions—to announce and follow through with replication-friendly editorial policies that actually lead to publishing replications in major journals. Why editors in a competitive publishing market would be willing to do this, even editors who are supposed to be creatures of the organization suggesting this approach, is unclear; but perhaps moral suasion would effect changes in editorial policies.

⁴The *Journal of Political Economy* published “Confirmations and Contradictions” for a while also.

Another approach would be the creation of a *Journal of Economic Extensions and Replications*, publishing articles whose sole purposes would be to replicate and extend other articles (presumably those previously published in major journals). Given the lags in publishing today, by the time an article appears in a major journal it has been circulating in something near its final form for at least two years, and often four or more. Even with a fairly quick turn-around, articles in the *JEER* would be attempting to replicate research that had been known to most interested scholars for three years or more. Who would wish to publish in such a journal? Who would bother subscribing to or reading it? The current nature of economics publishing—the ridiculously long publishing lags in this profession—would seem to doom such a journal to obscurity and early death.

Since the market for major articles in applied microeconomics demonstrates that they are usually replicated, and appropriately so, in other “laboratories” (other times or economies) than the original one, perhaps there is a place for direct replication aimed mainly at catching errors. Such a scheme would go well beyond the current practice at AEA-funded journals of checking for consistency between code and data sets (in those studies that do not use proprietary data). Consider the following change in the publication process at major journals: Once an empirical article is accepted for publication it, along with all the data and code that underlay it (and that are currently deposited with the journal), would be sent to one of a cadre of Replicators that the journal has established. Members of this cadre would have agreed to take no more than three months to examine the study, including using the data to replicate and perhaps expand upon (using additional specifications) the original article. Their completed replication—written up in three to five manuscript pages—would then be guaranteed publication in the journal just behind the original article.

The benefit from this scheme is clear—it would discover errors of calculation and transcription that referees rarely have the time, inclination or material to catch, and it would allow some extension of the original research. Given current publication lags, an extra three-month lag is unimportant—although with those lags and, more important, requirements for depositing data and code that were instituted a decade ago and that disadvantaged empirical relative to theoretical work, this makes the proposal only marginally unattractive.

The suggestion has, however, more serious problems. Who would wish to become a Replicator? Only younger scholars or extremely senior scholars are likely to be interested; and those younger economists at more influential institutions would probably not see their careers advanced by publishing several pages of replication once per year, even in a top journal, and would be unlikely to agree to participate. The program might only attract scholars outside the most highly-regarded institutions. They would catch some errors, and some interesting new findings might develop from articles that these Replicators examine. But would their work be credible to the profession—*quis custodiet ipsos custodes?*

III. Whither Replication in Empirical Microeconomics?

This examination of the reality of replication in empirical work suggests that research which the community of scholars implicitly deems important is replicated, including both on other data and on the data that are now typically required to be deposited with the journal. The more important type of replication is not like that of “hard-scientific” research, but rather in the only sensible way for a **social** science—by testing the fundamental idea or construct in a different **social** context. Important mistakes do get caught, and important ideas initially tested on only one set of data must survive tests on other data. The market appears to work well, and for various reasons the alternatives that I have suggested do not seem practicable or necessary. Taking these

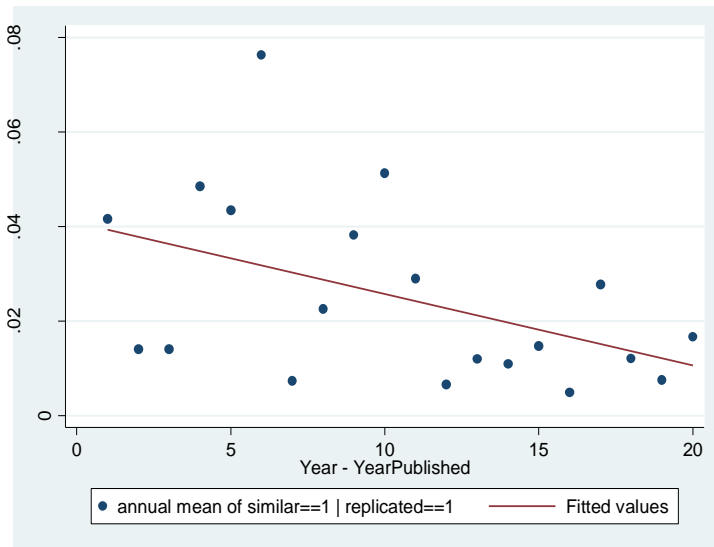
considerations together, *Mad's* attitude, "What—Me Worry?" is apropos the existing role of replication in empirical economics.

REFERENCES

- Ashenfelter, Orley C., and Alan B. Krueger.** 1994. "Estimates of the Economic Return to Schooling from a New Sample of Twins." *American Economic Review* 94 (5): 1157-73.
- Borjas, George J.** 1995. "Ethnicity, Neighborhoods and Human-Capital Externalities." *American Economic Review* 85 (3): 365-90.
- Card, David E., and Alan B. Krueger.** 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772-93.
- Currie, Janet M., and Duncan Thomas.** 1995. "Does Head Start Make a Difference?" *American Economic Review* 85 (3): 341-64.
- Davis, Steven J., and John C. Haltiwanger.** 1992. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation." *Quarterly Journal of Economics* 107 (3): 819-63.
- Fourcade, Marion, Etienne Ollion, and Yann Algan.** 2015. "The Superiority of Economists." *Journal of Economic Perspectives* 29 (1): 89-114.
- Hamermesh, Daniel S.** 2007. "Viewpoint: Replication in Economics." *Canadian Journal of Economics* 40 (3): 715-33.
- Hamermesh, Daniel S.** 2017. "Citations in Economics: Measurement, Impacts and Uses." *Journal of Economic Literature*, 55, forthcoming.
- Hamermesh, Daniel S., and Jeff E. Biddle.** 1994. "Beauty and the Labor Market." *American Economic Review* 94 (5): 1174-94.
- Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan.** 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83 (4): 685-709.
- Lazear, Edward P.** 1990. "Job Security Provisions and Employment." *Quarterly Journal of Economics* 105 (3): 699-726.
- Meyer, Bruce D.** 1990. "Unemployment Insurance and Unemployment Spells." *Econometrica* 58 (4): 757-82.
- Neal, Derek A., and William R. Johnson.** 1995. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104 (5): 869-95.
- Reinhart, Carmen M., and Kenneth S. Rogoff.** 2010. "Growth in a Time of Debt." *American Economic Review* 100 (2): 573-8.
- Whaples, Robert M.** 2006. "The Costs of Critical Commentary in Economics Journals." *Econ Journal Watch* 3(2): 275-82.

TABLE 1. TOTAL CITATIONS AND RANK OF STUDIES SUMMARIZED

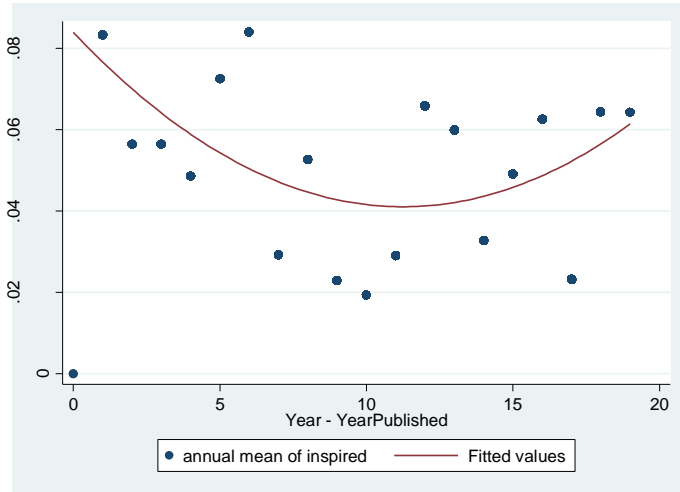
Study, and Rank in Volume	GS Citations	WoS Citations	Percent Only Related
Ashenfelter and Krueger (1994) 12/182	1284	303	91.7
Borjas (1995) 10/175	1099	269	94.1
Card and Krueger (1994) 6/182	2053	460	91.7
Currie and Thomas (1995) 9/175	1026	254	94.5
Davis and Haltiwanger (1992) 5/56	1736	399	87.7
Hamermesh and Biddle (1994) 10/182	1289	322	91.6
Jacobson <i>et al</i> (1993) 3/176	1785	457	95.0
Lazear (1990) 3/54	1423	305	94.4
Meyer (1990) 5/68	2068	549	92.9
Neal and Johnson (1996) 2/48	1365	397	95.5



$$\text{pr}(\text{More than inspired}) = 0.0409 - 0.00151\text{PostPubYear}; \text{Adj. } R^2 = 0.174 \text{ (N = 20)}$$

(0.0081) (0.00068)

FIGURE 1. PROBABILITY OF CITING ARTICLE BEING SIMILAR OR REPLICATION



$$\text{pr}(\text{Inspired}) = 0.0808 - 0.00679\text{PostPubYear} + 0.00029\text{PostPubYear}^2; \text{Adj. } R^2 = 0.154 \text{ (N = 20)}$$

(0.0134)
(0.00294)
(0.00014)

FIGURE 2. PROBABILITY OF CITING ARTICLE BEING INSPIRED