
A Call for Replication Studies

Public Finance Review
38(6) 787-793
© The Author(s) 2010
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1091142110385210
<http://pfr.sagepub.com>


Leonard E. Burman,¹ W. Robert Reed² and James Alm³

“It is our belief that journals should publish the results of replication attempts—favorable or unfavorable.”

—Dewald, Thursby, and Anderson (1988)

“Econometric software has bugs.”

—McCullough and Vinod (1999)

“... [R]eplicable economic research is the exception and not the rule.”

—Anderson et al. (2005)

With this issue, *Public Finance Review* issues an open call for papers that report the results of attempts to replicate significant empirical research in public economics, published in this journal or elsewhere.

The Scientific Need for Replication Studies

A basic requirement for scientific integrity is the ability to replicate the results of research, and yet, with some occasional historical exceptions,

¹ Syracuse University, Syracuse, NY

² University of Canterbury Christchurch, New Zealand

³ Tulane University New Orleans, LA

replication has never been an important part of economic research. The absence of replication studies is particularly problematic because empirical economic research is often prone to error (Dewald, Thursby, and Anderson 1986; Anderson et al. 2005). The errors can arise from inadvertent and innocent mistakes by researchers or from bugs in computer programs but also from carelessness or even dishonesty. Several researchers, such as Lovell and Selover (1994) and McCullough and Vinod (1999), have even found that different packaged software programs produce very different results for relatively straightforward statistical techniques applied to identical data sets. Stokes (2004) found that virtually all the standard econometric software programs failed to recognize that a probit maximum likelihood problem that was originally posed by Maddala (1992) did not in fact have a unique solution.¹ Even as identified errors are corrected, the increasing complexity of canned software raises the likelihood that numerical errors can materially affect empirical results. In addition, authors can misconstrue the results of packaged software because of the many different ways in which software can code the estimators.

One thing standing in the way of replications is that there is little professional reward for doing them (Anderson et al. 2005). *The Journal of Political Economy* added a section devoted to validation of articles published in *JPE*, but “[i]nvariably, this section contained papers employing either new data sets or alternative statistical techniques; little attention was paid to replication” (Dewald, Thursby, and Anderson 1986).² Indeed, the article by Dewald, Thursby, and Anderson (1986) is a rare example of an article explicitly devoted to replication using original data and methods that was published in a major journal. The article was published because of its broader point about the importance of replication and the appalling results it found in trying to validate research published in a major journal, the *Journal of Money, Credit, and Banking*. The *Journal of Human Resources* also has a policy inviting replication, although the last published replication experiment appears to date back to 1991 (Moffitt and Rangarajan 1991), reporting a failed replication.

Public Finance Review proposes to subject empirical public finance research to the scientific standard of replicability, by providing an outlet for the publication of replication studies. Of course, this journal and others have always published articles that refute the results of earlier research, and we will continue to do so when the findings are significant. A difference is that we, for the first time, will also report findings that validate, as well as those that invalidate, previous research, a practice that is common in the natural sciences. We encourage all researchers, especially graduate students in

economics, to attempt to replicate significant research findings, and after our standard peer-review process, *Public Finance Review* will publish the results of these replication studies.

Public Finance Review envisions three kinds of replications:

- Positive (or validating) replications: These are studies where the replicating author shows the original article's findings are robust to substantial extensions over time, explanatory variables, and/or alternative estimation procedures.
- Negative replications (negative—Type 1): These are studies where the replicating author is unable to reproduce the original article's results using the same data, the same specification, and the same econometric software. In these cases, supplementary correspondence with the editor should provide evidence that substantial efforts were made by the researcher to work with the original author to reproduce the original results.
- Negative replications (negative—Type 2): These are studies where the replicating author is able to reproduce the original article's results, but he or she finds that the original results are not robust to substantial extensions over time, data sets, explanatory variables, functional forms, software, and/or alternative estimation procedures.

Public Finance Review will publish all three kinds of replication studies, those that validate and those that invalidate previous research.

Some Ground Rules and Principles for Replication Studies

First, it is our expectation that these replication studies will be standard, full-length manuscripts, although shorter manuscripts will also be considered.

Second, in most cases the original research must have been published in this or another peer-reviewed economics journal; however, widely cited articles in conference volumes or books or even unpublished working papers may also be considered, depending on the importance and visibility of their results. Given the focus of *Public Finance Review*, the articles should be broadly in the area of public economics. Researchers are welcome to ask the editor for guidance in advance on the potential suitability of a particular study for replication. Replication papers should give some evidence of the original article's influence.

Third, the researcher conducting the replication experiment must be independent of the original authors; that is, the researcher should not be a graduate student under the supervision of any of the original authors or a current or recent coauthor.

Fourth, whenever possible, the replicating researcher must first attempt to replicate exactly the original findings by starting with the same data, the same specification, and the same econometric software, before testing the robustness of the original research, say, using a different data set, adding or subtracting years or observations, adding or subtracting variables, trying different functional forms, applying different estimation techniques, using different software, and so on. If the original results cannot be replicated, then the replicating researcher should attempt to reconcile the differences by communicating with the original author. The original author is encouraged to cooperate with those conducting the replication experiments to the extent practicable. A researcher who reports failed replication experiments must submit to the editor copies of all correspondence with the original author. (The correspondence will not be published.)

Fifth, the resulting replication paper should contain a detailed exposition describing the efforts to replicate the results of the original article. The exposition should be sufficiently detailed so that the original author (and any others who wish to replicate these results) will understand that the replication was done correctly. The paper should also attempt to explain the reasons for any differences.

Sixth, any submitted paper will be subject to the standard peer-reviewed referee process. One of the referees will normally be the original author.

Seventh, any submitted paper should be clearly identified in the submission letter to the editor as a "Replication Study."

Eighth, if the submitted paper is accepted, then the replicating researcher will be asked to submit a brief (approximately 1,000 words) summary of the results. This summary will be published in *Public Finance Review* in the Replications section, along with a link to a Web site (provided by the journal) at which the full-length paper will be posted.

Ninth, the author of the original article will be given the opportunity to respond to the replication. He or she can also choose to submit a brief (approximately 1,000 words) summary of their response, which will be published in *Public Finance Review* alongside the summary of the replication study, with a Web site link to the same site where the full replication study will be posted.

Tenth, it will not be the practice of this journal to publish clarifications of data, programs, or procedures from the original author that could have been supplied to the researcher attempting replication but were not.³

Eleventh, Public Finance Review will consider for publication multiple replications of the same original article, if done by different replicating authors.

Twelfth, Public Finance Review will not attempt to adjudicate disputes between the replicating and original authors. The responsibility of the editor is merely to ensure that the replicating work has been done to a high standard of competence.

Conclusions

We recognize that conducting research so that it could be replicated is not easy, but we also believe that the benefits of replication in economics are well worth the costs. The fact that the few researchers who have conducted replication experiments have failed more often than not is deeply disturbing, especially in a field such as public economics, whose *raison d'être* is to inform public policy.

We understand that authors may be concerned about their time because they believe that the person attempting the replication is incompetent. We expect that young researchers who conduct replication experiments will do it under the supervision of academic advisors who will be cognizant of this risk and will take steps to minimize it. We will take seriously the evidence presented by authors that someone attempting a replication is failing because of incompetence rather than any problems with the original research. That evidence will be deemed most persuasive when the authors have adhered to appropriate standards for archiving programs, data, and methodology, therefore, making replication relatively straightforward for a competent researcher.

Those concerns notwithstanding, we strongly urge academic public finance economists and applied econometricians and statisticians to encourage promising graduate students to undertake replication experiments. Ultimately, our goal is for every empirical study published in the *Public Finance Review* to be replicable by a researcher who is independent of the original study's author. If these replication studies are done according to the standards set forth here, then the researchers can expect publication to result in *Public Finance Review*. Moreover, the process of replication itself is a valuable method for teaching young scholars appropriate research methodology.⁴ And we remain convinced that the benefits of replication far exceed the costs. Indeed, as emphasized by McCullough and Vinod (2003): Research that cannot be replicated is not science, and

cannot be trusted either as part of the profession's accumulated body of knowledge or as a basis for policy.

Notes

1. McCullough and Vinod (2003) drew more sweeping conclusions: "Either intentionally or unintentionally, it is fairly easy to trick a solver [of nonlinear optimization problems] into falsely reporting an extremum—whether a maximum for likelihood estimation, or a minimum for least-squares estimation."
2. *The Journal of Political Economy* had a "Confirmations and Contradictions" section from 1976 to 1999. Mirowski and Sklivas (1991) reported that 5 of 36 notes appearing in this section from 1976 to 1987 included replications, of which only 1 was successful in actually replicating the original results. Anderson et al. (2005) counted 13 more notes through 1999, of which only 1 included a replication and wrote, "Apparently *JPE* has allowed the section to die an ignominious death befitting the section's true relation to replication: It has been inactive since 1999" (Anderson et al. 2005).
3. "[J]ournals provide inappropriate incentives when they publish clarifying comments by authors who have failed to respond to requests for clarification prior to publication of negative results" (Dewald, Thursby, and Anderson 1988).
4. For example, Auerbach, Hassett, and Oliner (1994) reexamined a pair of cross-country empirical studies that measured large excess social returns to equipment investment. Those excess returns turned out to depend entirely on one country, Botswana, whose equipment was used to mine diamonds and which had experienced exceptional economic growth over the time period studied. The surprising result disappeared when Botswana was excluded from the data set.

Acknowledgments

We thank Alan Auerbach, James Poterba, and John Karl Scholz for helpful discussions.

References

- Anderson, R. G., W. H. Greene, B. D. McCullough, and H. D. Vinod. 2005. The role of data and program code archives in the future of economic research. Working Paper 2005-014C, Federal Reserve Bank of St. Louis, St. Louis, MO.
- Auerbach, A. J., K. A. Hassett, and S. D. Oliner. 1994. Reassessing the social returns to equipment investment. *Quarterly Journal of Economics* 109:789-802.
- Dewald, W. G., J. G. Thursby, and R. G. Anderson. 1986. Replication in empirical economics: The *Journal of Money, Credit, and Banking* project. *The American Economic Review* 76:587-603.

- Dewald, W. G., Thursby, J. G., and Anderson, R. G. 1988. Replication in empirical economics: The *Journal of Money, Credit, and Banking* project: Reply. *The American Economic Review* 78:1162-1163.
- Lovell, M. C., and D. D. Selover. 1994. Econometric software accidents. *Economic Journal* 104:713-725.
- Maddala, G. S. 1992. *Introduction to econometrics* 2nd ed. New York: MacMillan Publishers.
- McCullough, B. D., and H. D. Vinod. 1999. The numerical reliability of econometric software. *The Journal of Economic Literature* 37:633-665.
- . 2003. Verifying the solution from a nonlinear solver: A case study. *The American Economic Review* 93:873-892.
- Mirowski, Philip E. and Steven Sklivas. 1991. Why econometricians don't replicate (although they do reproduce). *Review of Political Economy* 3(2): 146-163.
- Moffitt, R., and A. Rangarajan. 1991. The work incentives of AFDC tax rates: Reconciling different estimates. *The Journal of Human Resources* 26:165-179.
- Stokes, H. H. 2004. On the advantage of using two or more econometric software systems to solve the same problem. *Journal of Economic and Social Measurement* 29:307-320.