

NBER WORKING PAPER SERIES

INCENTIVES FOR REPLICATION IN ECONOMICS

Sebastian Galiani
Paul Gertler
Mauricio Romero

Working Paper 23576
<http://www.nber.org/papers/w23576>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2017

We gratefully acknowledge funding for this research from the Berkeley Initiative for Transparency in the Social Sciences, a program of the Center for Effective Global Action (CEGA), with support from the Laura and John Arnold Foundation. The paper has also benefited from comments by Abhijit Banerjee, Annette Brown, Rob Jensen, Temina Madon, Ted Miguel, Don Moore, Emily Oster, Jennifer Sturdy, Sarah White, Benjamin Wood and participants in the 2016 BITSS annual meeting. Ada Kwan and Alexandra Wall provided excellent research assistance. The authors have no material or financial interests in the results of the paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Sebastian Galiani, Paul Gertler, and Mauricio Romero. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Incentives for Replication in Economics
Sebastian Galiani, Paul Gertler, and Mauricio Romero
NBER Working Paper No. 23576
July 2017
JEL No. A1,A12

ABSTRACT

Replication is a critical component of scientific credibility as it increases our confidence in the reliability of the knowledge generated by original research. Yet, replication is the exception rather than the rule in economics. In this paper, we examine why replication is so rare and propose changes to the incentives to replicate. Our study focuses on software code replication, which seeks to replicate the results in the original paper using the same data as the original study and verifying that the analysis code is correct. We analyse the effectiveness of the current model for code replication in the context of three desirable characteristics: unbiasedness, fairness and efficiency. We find substantial evidence of “overturn bias” that likely leads to many false positives in terms of “finding” or claiming mistakes in the original analysis. Overturn bias comes from the fact that replications that overturn original results are much easier to publish than those that confirm original results. In a survey of editors, almost all responded they would in principle publish a replication study that overturned the results of the original study, but only 29% responded that they would consider publishing a replication study that confirmed the original study results. We also find that most replication effort is devoted to so called important papers and that the cost of replication is high in that posited data and software are very hard to use. We outline a new model for the journals to take over replication post acceptance and prepublication that would solve the incentive problems raised in this paper.

Sebastian Galiani
Department of Economics
University of Maryland
3105 Tydings Hall
College Park, MD 20742
and NBER
galiani@econ.umd.edu

Mauricio Romero
Department of Economics
University of California
San Diego CA
mtromero@ucsd.edu

Paul Gertler
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720
and NBER
gertler@haas.berkeley.edu

Replication is a critical component of scientific credibility as it increases our confidence in the reliability of the knowledge generated by original research.¹ Yet, replication is the exception rather than the rule in economics.^{2,3} In this paper, we examine why replication is so rare and propose changes to the incentives to replicate.

Our study focuses on software code replication, which seeks to replicate the results in the original paper using the same data as the original study and verifying that the analysis code is correct. Other forms of replication include reanalysis and study replication. Reanalysis examines the original study data to assess whether the conclusions of the original study are robust to different assumptions about variable construction, sample, identification strategy, and statistical methods. A study replication uses different data to investigate the external validity of the conclusions.

Code replication or “verification”^{3,4} is a two-step process; first reconstruct the sample and variables used in the analysis from the raw data, second confirm that the analysis software code that fits the statistical models reproduce the reported results.

We analyze the effectiveness of the current model for code replication in the context of three desirable characteristics:

1. ***Unbiasedness***: there is no “overturn bias;” i.e., the model does not create incentives to “find” or claim mistakes in the original analysis.
2. ***Fairness***: all papers have the same (perhaps conditional) positive probability of being replicated and is independent of author, topic, and results.
3. ***Efficiency***: the model should provide the right incentives at minimum cost.

Replication needs to be low cost for researchers to undertake it, fair so that all studies, maybe within the same category, face the same probability of being replicated, and unbiased so that the original authors have reason to participate and the profession believes the replication results. These characteristics are necessary to establish a credible threat of valid replication that authors take seriously enough to modify behavior. We document that the current model for code replication does not have these characteristics, and then outline a new model that solves many of the actual problems.

Incentives for Replication

While it is hard to know how many replications have been started, few have been published. We searched for “replication studies” of any type among articles and comments published in 11 of the top-tier journals in economics since 2011, and found eleven, all of which claimed to overturn the original results. Table S1 in online supplemental materials lists the journals searched and Table S2 lists the replication studies found. This suggests two problems: First, it is hard to publish replication studies and thus the expected professional return to replication is low, and second that there are substantial incentives to “overturn” the original results in order to get a replication study published.

There appears to be substantial “overturn bias” among journal editors. We surveyed 204 editors and co-editors from 11 top journals in economics. Table S3 lists the Journals surveyed. Overall the response rate was 43%, with at least one editor from every journal answering our survey. While all editors responded they would in principle publish a replication study that overturned the results of the original study, only 29% responded that they would consider publishing a replication study that confirmed the original study results.

More evidence of possible “overturn bias” comes from the experiences of the International Initiative for Impact Evaluation’s (3ie) replication program. While the 3ie replication program more generally sponsored all types of replication, their experience is extremely valuable because it is a rare case where we have a sample of replication studies started as opposed to published. 3ie selected “important” or influential papers to be replicated and then held an open competition for replication of these studies awarding approximately \$25,000 to each study.⁵ 3ie set up a process that consisted of peer review and offered the original authors the opportunity to review and comment on the replication studies. Of the 27 studies commissioned, 20 were completed, and 7 (35%) overturned some of the original results; i.e., claimed to be not able to fully replicate the original article. Only 1 was published in a peer reviewed journal and it overturned the results from the original paper.

Despite the best efforts of 3ie, an adversarial relationship between original and replication researchers can be inferred from the responses of the original authors to the 3ie

replication reports. Indeed, we take insightful quotes from the original authors' responses and associated blog posts in 5 of the 7 replication studies that claimed to overturn the original results. For example, the 3ie sponsored replication⁷ of a highly cited paper on deworming⁶ resulted in a heated public debate.⁸ Several independent scholars questioned the assumptions made by the replicators, claiming that many of these lacked scientific justification and may have been made to maximize the likelihood of overturning the original results.^{9,10}

In one response to the 3ie replication of their paper, the original authors explicitly address overturn bias: "The incentives of the replicators, particularly in terms of publication, are to "overturn" the original results, and could lead to overstatement of the magnitude of criticism."¹¹ Several of the original authors' replies to other 3ie replication studies include: " [Despite replicating all the results in the paper], ... we disagree with the unnecessarily aggressive tone of some statements in the replication report particularly in the abstract ...",¹² " ... the statement that our original conclusions were robust was buried in the text with no mention of this in either the abstract or conclusion; instead, emphasis was placed on the statement that our findings on agricultural extension were not robust,"¹³ and "[Despite having] informed the replicators about this, we find this added comment in the abstract of the replication report inaccurate, inappropriate and, arguably, misleading to the readers of their report, something we had hoped to correct with this added section to our original response note [to the replicators]."¹⁴

Data access

One of the biggest costs of replication is access to original data and analysis code. In the past, replicators had to depend on the original authors, who may have little incentive to cooperate post-publication. The economics profession has recently made great strides towards lowering the cost of replication by requiring that data and code used in published papers be posted. In this section, we assess how well this policy is working.

We surveyed journal websites for their policies regarding publicly sharing data and computer code before publication (See

Table 1). We surveyed 11 top-tier and 23 mid-tier empirical economics journals. We also surveyed the ten top journals in the other social science disciplines and the general science journals *Nature*, *Science* and the *Proceedings of the National Academy of Sciences* to benchmark economics. Table S4 lists the journals searched.

In the sample, economics and political science journals are more likely to have policies requiring authors to submit their code and data before publication. While the journals in economics have an explicit policy regarding raw de-identified data where raw data refers to the original data files used in the study. In contrast, estimation data refers to the final estimation data set after data cleaning and variable manipulation. This is not an explicit requirement in other disciplines. Most journals that require data posting, except for some political science journals, do not verify that the code and data submitted by the authors are easily executable and actually replicate the original.

Table 1: Journal Policies on Posting Data and Code

(See Table S4 for the specific journals assessed)

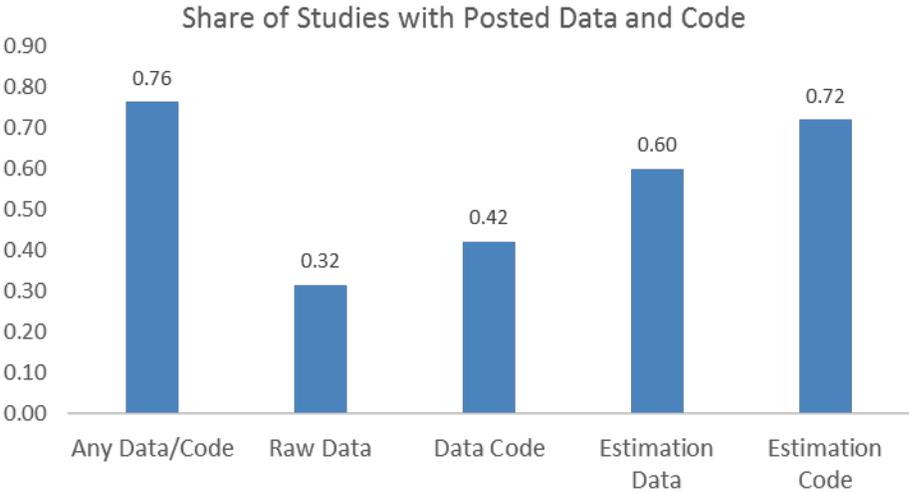
	Economics (Top tier)	Economics (Mid tier)	Political Science	Sociology	Psychology	General Science
Journals analyzed	11	23	10	10	10	3
Code/Data required before publication	10	8	8	2	1	3
Code/Data optional/encouraged	1	9	0	2	2	0
Raw data must be submitted	10	7	0	0	0	0
Code/Data verified before publication	0	0	3	0	0	0

As a result, much of the data and code are not easily usable to replicate the original results. Despite these posting requirements, compliance with journal data transparency policies is low in economics. We attempted to replicate the tables and figures of a paper using the code and data provided by the author explicitly for those purposes. We surveyed the last three issues as of May 2016 of nine leading economics journals. Table S5 lists the journals used in this exercise. In total, 415 articles were published in these journals, of which, 266 (64%) are “non-structural” empirical papers and 63 of those used restricted or proprietary data. The remaining 203 articles were included in our main sample.

Among those 203 articles, we first checked to see whether the following files were posted and downloadable: i) the raw data used in the study, ii) the final estimation data set after data cleaning and variable manipulation were performed, iii) the data manipulation code used to convert the raw data to the estimation data, and iv) the estimation code used to produce the final tables and figures. Overall, we found that only 76% of studies published at least one of the four files.

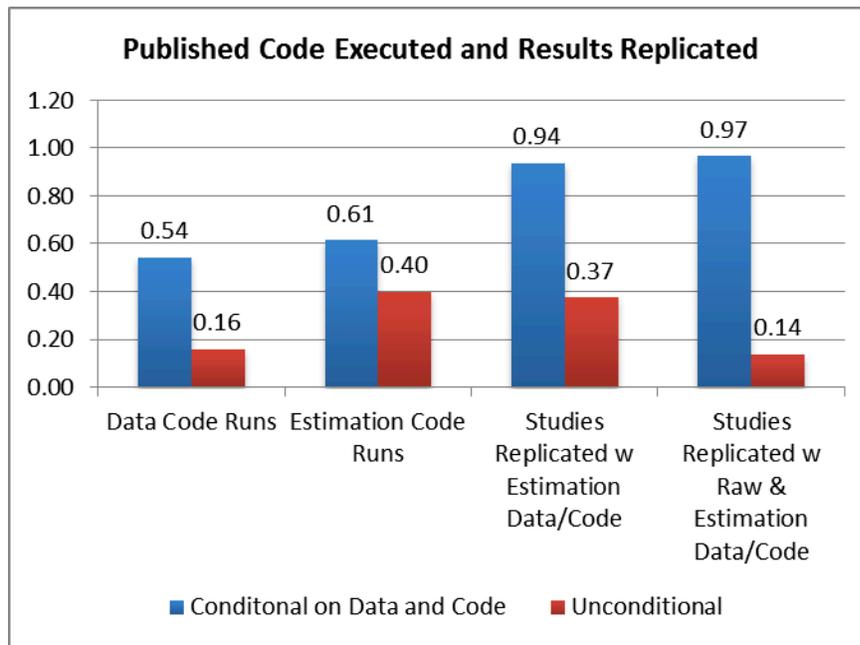
The raw data and data manipulation code were posted in about one-third of the cases, while the final estimation data and code were posted in about two-thirds of the cases.

Figure 1



We then tried to replicate the tables and figures for the papers that posted data and code. Conditional on having the data and code available, only 54% of articles had “data manipulation code” that did not require major modifications (e.g., changing folder directories and installing additional packages) and only 61% of articles had “estimation code” that did not require major modifications. In short, only 14% of the articles in our sample of 203 were fully replicable (i.e. from raw data, to final tables and figures) and only 37% were partially replicable (i.e. from the estimation data to final tables and figures).

Figure 2



Our results align with previous findings in the literature.² A study of the articles published in the *Journal of Money, Credit and Banking* found that only 37% of articles met data archive requirements, and only 20% of studies could be replicated using the information from the archive.¹⁶ Another study attempted to replicate 67 papers published in 13 well-regarded general interest and macroeconomics journals and were only able to replicate 29 of them¹⁷ This problem is not confined to economics. In 2013 only 18 of 120

political science journals had replication policies¹⁸ and a more recent study found that only 58% of articles in top political science journals publish their data and code.¹⁹

A new model for code replication

We outline a new simple model that would reduce overturn bias, increase fairness and reduce the cost of replication, and thereby increase the prevalence and effectiveness of replication. The core of the model is to have journals take responsibility for overseeing the replication exercise post-acceptance but pre-publication. Specifically, authors would submit their data and code after a conditional acceptance. Journals would then verify that all raw data and code (i.e. sample and variable construction, as well as estimation code) are included and executable. They would then commission research associates perform a “push button exercise” that verifies that the code executes and reproduces the tables and figures in the article. If the code does not execute or reported results are different, editors could either ask authors to correct their errors or choose to re-review the paper.

Finally, for a random sample of papers the journal would attempt to re-construct the code from scratch or search the executable code for errors. This would be an iterative process until authors and editors are able to reach agreement. If the results change, the editors could then allow the authors to revise the paper or choose to re-review the paper.

This simple procedure has three desirable properties. First, it is unbiased since there are no overturn bias incentives for the parties involved (editors/researchers). Second, it is fair because all papers have an equal probability of being replicated. Third, it is low-cost: there is little cost associated with having a research associate perform “push button exercises,” authors have strong incentives to cooperate pre-publication, and there are fewer adversarial feelings. However, it would increase journal costs that could be recovered through increased subscription fees, submission fees or publication fees. Initially, it may also slow down time from acceptance to publication for some papers. However, over time, authors will internalize the incentives provided and will submit the materials and analysis in a form that the study replication will be done very efficiently, at low cost, and very fast. Thus, such a mechanism would create a strong incentive not to misreport findings and to

ensure that code is free of errors thereby instilling confidence in the credibility of the science.

Bibliography

1. Nosek, B. A., et al., et al. Promoting an open research culture. *Science*, Vol. 348, pp. 1422-1425 (2015)
2. Christensen, Garret S and Miguel, Edward. Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature* (Forthcoming).
3. Berry, James, Lucas C. Coffman, Douglas Hanley, Rania Gihleb, and Alistair J. Wilson. Assessing the Rate of Replication in Economics. *American Economic Review*, 107(5) pp. 27-31 (2017)
4. Clemens, Michael A. The meaning of failed replications: A review and a proposal. *Journal of Economic Surveys*, Vol. 31, pp. 326-342 (2017)
5. The International Initiative for Impact Evaluation. 3ie Replication Programme. [Online] 2017. [Cited: May 30, 2017.] http://www.3ieimpact.org/media/filer_public/2016/11/22/3ie-replication-program-document.pdf
6. Miguel, Edward and Kremer, Michael. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, Vol. 72, pp. 159-217 (2004).
7. Davey, Calum, et al., et al. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial *International Journal of Epidemiology*, Vol. 44, p. 1581 (2015).
8. Evans, David. Worm Wars: The Anthology. [Online] August 04, 2015. [Cited: May 16, 2017.] <https://blogs.worldbank.org/impactevaluations/worm-wars-anthology>
9. Blattman, Chris. Dear journalists and policymakers: What you need to know about the Worm Wars. [Online] July 23, 2015. [Cited: May 16, 2017.] <http://chrisblattman.com/2015/07/23/dear-journalists-and-policymakers-what-you-need-to-know-about-the-worm-wars/>
10. Ozler, Berk. Worm Wars: A Review of the Reanalysis of Miguel and Kremer's Deworming Study. [Online] July 24, 2015. [Cited: May 16, 2017.] <http://blogs.worldbank.org/impactevaluations/worm-wars-review-reanalysis-miguel-and-kremer-s-deworming-study>

11. Jensen, Robert and Oster, Emily. TV, Female Empowerment and Fertility Decline in Rural India: Response to Iversen and Palmer-Jones. [Online] 2014. [Cited: May 30, 2017.] http://www.3ieimpact.org/media/filer_public/2014/06/07/jensen_oster_response.pdf
12. Galiani, Sebastian and Schargrodsky, Ernesto. Response to Replication Report for "Property Rights for the Poor: Effects of Land Titling". [Online] 2015. [Cited: May 2017, 30.] http://www.3ieimpact.org/media/filer_public/2015/11/03/rps9-original-author-response.pdf
13. Dercon, Stefan, et al., et al. The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages: Response to William Bowser. [Online] 2015. [Cited: May 30, 2017.] http://www.3ieimpact.org/media/filer_public/2015/02/06/original_author_response_rps_4.pdf
14. Cattaneo, Matias, et al., et al. Second Response to Replication Report for "Housing, Health and Happiness". [Online] 2015. [Cited: May 30, 2017.] http://www.3ieimpact.org/media/filer_public/2015/08/17/original_author_response_to_basurto_replication-2007.pdf
15. Wood, Benjamin DK. Reflections on replication research: a conversation with Paul Winters. [Online] November 3, 2015. [Cited: May 16, 2017.] <http://blogs.3ieimpact.org/reflections-on-replication-research-a-conversation-with-paul-winters/>
16. McCullough, Bruce D, McGeary, Kerry Anne and Harrison, Teresa D. Lessons from the JMCB Archive. , Journal of Money, Credit, and Banking, Vol. 38, pp. 1093-1107 (2006).
17. Chang, Andrew C, Li, Phillip and others. Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say " Usually Not". [Online] 2015. [Cited: May 30, 2017.] <http://www.federalreserve.gov/econresdata/feds/2015/files/2015083pap.pdf>
18. Gherghina, Sergiu and Katsanidou, Alexia. Data availability in political science journals. European Political Science, Vol. 12, pp. 333-349 (2013)
19. Key, Ellen M How Are We Doing? Data Access and Replication in Political Science. Political Science & Politics, Vol. 49, pp. 268-272 (2016).
20. Open Science Collaboration. Estimating the reproducibility of psychological science. Science 349, no. 6251 (2015).

Supplementary Material

Table S1: Journals Searched for Published Replication Studies

- | | |
|-----|------------------------------------|
| 1. | American Economic Review |
| 2. | AEJ: Economic Policy |
| 3. | AEJ: Applied Economics |
| 4. | Quarterly Journal of Economics |
| 5. | Econometrica |
| 6. | The Review of Economic Studies |
| 7. | Review of Economics and Statistics |
| 8. | Journal of Labor Economics |
| 9. | Journal of Public Economics |
| 10. | Journal of Political Economy |
| 11. | Journal of Development Economics |

Table S2: Replication studies published

Original				Replication				
	Title	Authors	Journal	Year	Title	Authors	Journal	Year
1.	Heterogeneity and aggregation: Implications for labor-market fluctuations	Chang, Yongsung, and Kim, Sun-Bin	AER	2007	Heterogeneity and Aggregation: Implications for Labor-Market Fluctuations: Comment	Shuhei Takahashi	AER	2014
2.	Stock Prices, News, and Economic Fluctuations	Paul Beaudry and Franck Portier	AER	2006	Stock Prices, News, and Economic Fluctuations: Comment	André Kurmann and Elmar Mertens	AER	2014
3.	Intergenerational occupational mobility in Great Britain and the United States since 1850	Jason Long and Joseph Ferrie	AER	2013	Intergenerational occupational mobility in Great Britain and the United States since 1850: Comment	Yu Xie and Alexandra Killewald	AER	2013
4.	Intergenerational occupational mobility in Great Britain and the United States since 1850	Jason Long and Joseph Ferrie	AER	2013	Intergenerational occupational mobility in Great Britain and the United States since 1850: Comment	Michael Hout and Avery M. Guest	AER	2013
5.	The colonial origins of comparative development: An empirical investigation	Daron Acemoglu, Simon Johnson and James A. Robinson	AER	2001	The colonial origins of comparative development: an empirical investigation: comment	David Y. Albouy	AER	2012
6.	Taxes, cigarette consumption, and smoking intensity	Jérôme Adda and Francesca Cornaglia	AER	2006	Taxes, cigarette consumption, and smoking intensity: comment	Jason Abrevaya and Laura Puzello	AER	2012
7.	Growth dynamics: the myth of economic recovery	Valerie Cerra and Sweta Chaman Saxena	AER	2008	Growth dynamics: the myth of economic recovery: comment	Hannes Mueller	AER	2012
8.	The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather	Olivier Deschênes and Michael Greenstone	AER	2007	The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment	Anthony C. Fisher, W. Michael Hanemann, Michael J. Roberts and Wolfram Schlenker	AER	2012
9.	Economic shocks and civil conflict: An instrumental variables approach	Edward Miguel, Shanker Satyanath and Ernest Sergenti	JPE	2004	Economic shocks and civil conflict: A comment	Antonio Ciccone	AEJ: Applied	2011
10.	Natural resource abundance and economic growth	Sachs and Warner	Working Paper, CGD	1997	Replicating Sachs and Warner's working papers on the resource curse	Davis	JDE	2013
11.	Institutions, and economic performance: the legacy of colonial land tenure systems in India	Banerjee and Iyer	AER	2005	On the colonial origins of agricultural development in India: a re-examination of Banerjee and Iyer, "History, institutions and economic performance"	Iversen, Palmer-Jones, and Sen	JDE	2013

Table S3: Journals from Which Editors and Co-Editors Surveyed

Journal	Discipline
1. American Economic Review	Economics
2. AEJ: Economic Policy	Economics
3. AEJ: Applied Economics	Economics
4. Quarterly Journal of Economics	Economics
5. Econometrica	Economics
6. The Review of Economic Studies	Economics
7. Review of Economics and Statistics	Economics
8. Journal of Labor Economics	Economics
9. Journal of Public Economics	Economics
10. Journal of Political Economy	Economics
11. Journal of Development Economics	Economics

Table S4: Journals Reviewed for Policies on Posting Code and Data

Journal	Discipline	Journal	Discipline
1. American Economic Review	Economics	2. American Sociological Review	Sociology
3. AEJ: Economic Policy	Economics	4. American Journal of Sociology	Sociology
5. AEJ: Applied Economics	Economics	6. Social Forces	Sociology
7. Quarterly Journal of Economics	Economics	8. Annual Review of Sociology	Sociology
9. Econometrica	Economics	10. Sociological Methods & Research	Sociology
11. The Review of Economic Studies	Economics	12. Theory & Society	Sociology
13. Review of Economics and Statistics	Economics	14. Social Networks	Sociology
15. Journal of Labor Economics	Economics	16. Sociological Theory	Sociology
17. Journal of Public Economics	Economics	18. Gender & Society	Sociology
19. Journal of Political Economy	Economics	20. Work & Occupations	Sociology
21. Journal of Development Economics	Economics	22. American J of Political Science	Political Science
23. Journal of Economic Perspectives	Economics	24. American Political Science Review	Political Science
25. Journal of Economic Literature	Economics	26. Journal of Politics	Political Science
27. AEJ: Macroeconomics	Economics	28. Quarterly J of Political Science	Political Science
29. AEJ: Microeconomics	Economics	30. Political Analysis	Political Science
31. Economic Journal	Economics	32. Comparative political Studies	Political Science
33. Journal of Economics Growth	Economics	34. World Politics	Political Science
35. International Economic Review	Economics	36. British Journal of Political Science	Political Science
37. The Rand Journal of Economics	Economics	38. International Organization	Political Science
39. Journal of Health Economics	Economics	40. International Security	Political Science
41. European Economics Review	Economics	42. Psychological Science	Psychology
43. Journal of Human Resources	Economics	44. J of Personality and Social Psych	Psychology
45. Journal of Industrial Economics	Economics	46. Journal of Experimental Psych	Psychology
47. Journal of Applied Econometrics	Economics	48. Journal of Applied Psychology	Psychology
49. Journal of Monetary Economics	Economics	50. Cognitive Psychology	Psychology
51. Journal of International Economics	Economics	52. Org Behavior & Human Decision	Psychology
53. Journal of Law and Economics	Economics	54. Social Psych and Personality Sci	Psychology
55. Journal of Business & Economic Stat	Economics	56. J of Experimental Social Psych	Psychology
57. Journal of Finance	Economics	58. Journal of Personality	Psychology
59. Journal of Law, Economics & Org	Economics	60. Personality & Social Psych Bull	Psychology
61. International Journal of Industrial Org	Economics	62. PNAS	General Science
63. Journal of Economic Behavior & Org	Economics	64. Nature	General Science
65. The Scandinavian Journal of Economics	Economics	66. Science	General Science
67. Oxford Economic Papers	Economics		

Table S5: Journals Included in Verification Studies

1. American Economic Review
2. AEJ: Economic Policy
3. AEJ: Applied Economics
4. Econometrica
5. The Review of Economic Studies
6. Review of Economics and Statistics
7. Journal of Labor Economics
8. Journal of Political Economy
9. Journal of Development Economics