Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not"

Andrew C. Chang and Phillip Li
Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.
2015

James J. Heckman

THE UNIVERSITY OF
CHICAGO

Econ 312, Spring 2021

1. Introduction

- In response to McCullough and Vinod (2003)'s failed replication attempt of several articles in the *American Economic Review (AER)*, then-editor of the *AER* Ben Bernanke strengthened the *AER's* data and code availability policy to allow for successful replication of published results by requiring authors to submit to the *AER* data and code replication files (Bernanke, 2004).

- Since the *AER* strengthened its policy, many of the other to journals in economics, such as *Econometrica* and the *Journal of Political Economy*, also started requiring data and code replication files.

- There are two main goals of these replication files:

(1) to bring economics more in line with the natural sciences by embracing the scientific method's power to verify published results, and
(2) to help improve and extend existing research, which presumes the original research is replicable.

- These benefits are illustrated by the policy-relevant debates between Card and Krueger (1994, 2000) and Neumark and Wascher (2000) on minimum wages and employment; Hoxby (2000, 2007) and Rothstein (2007) on school choice; Levitt (1997, 2002) and McCrary (2002) on the causal impact of police on crime; and, more recently, Reinhart and Rogoff (2010) and Herndon, Ash, and Pollin (2014) on fiscal austerity.

- Despite our finding that economics research is usually not replicable, our replication success rates are still notably higher than those reported by existing studies of replication in economics.

- McCullough, McGeary, and Harrison (2006) find a replication success rate for articles published in the JMCB of 14 of 186 papers (8%), conditioned on the replicators' access to appropriate software, the original article's use of non-proprietary data, and without assistance from the original article's authors.

- Adding a requirement that the JMCB archive contain data and code replication files the paper increases their success rate to 14 of 62 papers (23%). Our comparable success rates are 22 of 59 papers (37%), conditioned on our having appropriate software and non-proprietary data, and 22 of 38 papers (58%) when we impose the additional requirement of having data and code files. Dewald, Thursby, and Anderson (1986) successfully replicate 7 of 54 papers (13%) from the JMCB, conditioned on the replicators having data and code files, the original article's use of non-confidential data, help from the original article's authors, and appropriate software. Our comparable figure is 29 of 38 papers (76%).

2. Methodology and Sampling Frame

- Our sampling frame includes papers from 13 well-regarded macroeconomics and general interest economics journals:

American Economic Journal: Economic Policy, American Economic Journal: Macroeconomics, American Economic Review, American Economic Review: Papers and Proceedings (P&P), Canadian Journal of Economics, Econometrica, Economic Journal, Journal of Applied Econometrics, Journal of Political Economy, Review of Economic Dynamics, Review of Economic Studies, Review of Economics and Statistics, and Quarterly Journal of Economics.

- We choose papers from these journals because of the relative likelihood that such papers will have a policy effect and also influence future research.

- We do not select these journals to single out a particular author, methodology, institution, or ideology.

- From our sample of journals, we browse for original research articles published in issues from July 2008 to October 2013.4; Within these issues, we identify all papers with the following three characteristics:

1. An empirical component,
2. Model estimation with only US data, and
3. A key empirical result produced by inclusion of US gross domestic product (GDP), published by the Bureau of Economic Analysis (BEA), in an estimated model.

- We choose to focus on GDP because of its status as a standard macroeconomic statistic and its widespread use in research.

- For each paper in this set, we attempt to replicate the key empirical results.8 We focus on the key empirical results for two reasons: (1) replicating only the key results allows us to expand the sample to more papers, and (2) the key result of the paper is presumably what drove the paper's publication; robustness checks merely serve as confirming evidence.

- Defining a key result is subjective and requires judgmental decisions on our part.

- We attribute a key result of the paper to GDP when the authors themselves refer to GDP as driving a key result, or when a discussion of GDP is featured either in the abstract or prominently in the introduction of their work (or both).

- We also take key results as those that appear in figures and tables.

- We find 67 papers that fit these criteria.

- Of these papers, 6 papers use proprietary data for all of the key results, so we do not include them in our replication exercise (Fisher and Peters, 2010; Alexopoulos, 2011; Alexopoulos and Cohen, 2011; Hall and Sargent, 2011; Bansak, Graham, and Zebedee, 2012; Gilchrist and Zakrajšek, 2012).

- If a subset of the key results could be obtained using non-proprietary data, then we attempt to replicate those results.

- For the remaining papers that use public data and are published in journals that maintain data and code archives, we download the replication files provided by the authors through the online archives provided by the journals.

- Unlike prior work by McCullough, McGeary, and Harrison (2006), who found difficulty in accessing the archives of selected journals, we had no trouble doing so through the Board of Governors of the Federal Reserve System or Office of the Comptroller of the Currency subscriptions.

- However, consistent with McCullough, McGeary, and Harrison (2006), we find that journal data and code archives are incomplete.

- Of the 35 papers that use public data and are published in journals that require data and code replication files, we obtain files for 28 papers (80%) from journal archives.

- For papers where we are unable to obtain replication data and code files from journal archive sites, either because the mandatory files are is missing or because the paper is not subject to a data availability policy, we check the personal websites of each of the authors for replication files.

- If we are unable to locate replication files online, then we email each of the authors individually requesting the replication files.

- Of the 7 papers that use public data, are subject to a data and code policy, and do not have replication files on the journal's archive site, this procedure nets us one additional set of replication files.

- Therefore, we are unable to locate replication files for 6 of 35 papers (17%) that are published in journals that require submission of data and code replication files.

- For papers published in journals without a data and code availability policy and that use public data, we are unable to obtain data and code replication files for 15 of 26 papers (58%).

- We do not single out any paper or author that fails to comply with a journal's mandatory data and code policy.

- We therefore only report these summary statistics of compliance with data availability policies and only cite papers that we either successfully replicate, that use proprietary data, or where we have what appears to be a complete set of replication files in a software we do not possess.

- Our intention is to highlight the general state of replication files for published economics research, not to berate any given author, methodology, institution, or ideology.

- To determine whether a paper was subject to a data availability policy, we check the implementation dates of the journal data policies and compare them to the publication and submission dates of the published work.

- If the journal's website does not allow us to extract this information, then we query the editorial office as to when their data availability policy became effective.

- We do not ask the editorial offices whether a particular paper was subject to a data availability policy.

- Aside from papers with proprietary data, we find that journal data archives do not provide lists of potentially exempt papers.

- Therefore, we are unable to determine whether a paper is exempt for a reason other than using proprietary data, although we are not aware of reasons why journals would grant a paper a data and code exemption other than for proprietary data.

- For the papers for which we are able to obtain data and code replication files, we attempt to replicate the key results of the paper using only the instructions provided in the author readme files.

- If the readme files are insufficient or if the replication files are incomplete (or both) and the paper is subject to a replication policy, then we email the corresponding author (if no corresponding author, then the first author) for either clarification or to request the missing files.

- If we do not receive a response within a week, then we query the second author, and so on, until all authors on the paper had been contacted.

- We define a successful replication as when the authors or journal provide data and code files that allow us to qualitatively reproduce the key results of the paper.

- For example, if the paper estimates a fiscal multiplier for GDP of 2.0, then any multiplier greater than 1.0 would produce the same qualitative result (i.e., there is a positive multiplier effect and that government spending is not merely a transfer or crowding out private investment).

- We define success using this extremely loose definition to get an upper bound on what the replication success rate could potentially be.

- We allow for minimal re-working of the provided files, following the procedure of McCullough, McGeary, and Harrison (2006).

3. Results

- **Table 1** lists the papers we successfully replicate.

- **Table 2** breaks down our replication results by journal type.

- Panel A of Table 2 shows that our overall replication success rate is 29 of 67 papers (43%).

- Table 2, Panel B shows that we successfully replicate 23 of 39 papers (59%) from journals that require data and code replication files.

- This rate compares to 6 of 28 (21%) of the papers from journals that do not require such files, shown in Table 2, Panel C.

- These replication rates are similar when we only consider papers with publicly available data: we successfully replicate 23 of 35 (66%) of the papers from journals with mandatory data and code policies and 6 of 26 (23%) of the papers from journals without such policies.

Table 1: Successfully Replicated Papers

Auerbach and Gorodnichenko (2012)
Barro and Redlick (2011)
Baumeister and Peersman (2013)
Canova and Gambetti (2010)
Carey and Shore (2013)
Chen, Curdia, and Ferrero (2012)
Clark and McCracken (2010)
Corsetti, Meier, and Müller (2012)
D'Agostino and Surico (2012)
Den Haan and Sterk (2011)
Favero and Giavazzi (2012)
Gabaix (2011)
Hansen, Lunde, and Nason (2011)
Heutel (2012)
Inoue and Rossi (2011)
Ireland (2009)
Kilian (2009)
Kormilitsina (2011)
Krishnamurthy and Vissing-Jorgensen (2012)
Mavroeidis (2010)
Mertens and Ravn (2011)
Mertens and Ravn (2013)
Morley and Piger (2012)
Nakov and Pescatori (2010)
Ramey (2011)
Reis and Watson (2010)
Romer and Romer (2010)
Schmitt-Grohé and Uribe (2011)
Schmitt-Grohé and Uribe (2012)

We replicate the corrected results of Auerbach and Gorodnichenko (2012) found in Auerbach and Gorodnichenko (2013).

Table 2: Replication Sample and Results By Journal

	Papers Replicated Successfully	Papers With Public Data	Total Papers
Panel A: All Journals	29	61	67
Panel B: Journals that Require Data and Code:			
American Economic Journal: Economic Policy	2	4	4
American Economic Journal: Macroeconomics	3	3	4
American Economic Review	5	8	10
Canadian Journal of Economics	0	0	1
Econometrica	3	3	3
Journal of Political Economy	1	1	1
Review of Economic Dynamics	4	7	7
Review of Economic Studies	0	2	2
Review of Economics and Statistics	5	7	7
Total for Journals that Require Data and Code	23	35	39
Panel C: Journals that Do Not Require Data and Code:			
American Economic Review: P&P	0	4	5
Economic Journal	3	10	11
Journal of Applied Econometrics	1	10	10
Quarterly Journal of Economics	2	2	2
Total for Journals that Do Not Require Data and Code	6	26	28

Journal of Applied Econometrics requires data only. Economic Journal currently requires data and code, but the papers in our sample were not subject to a data and code policy according to the Economic Journal's editorial office.

- **Table 3**, Panel A provides explanations for why we are unable to replicate papers according to four broad classifications: "missing public data or code," "incorrect public data or code," "missing software," or "proprietary data."

- Panel B provides the breakdown for journals that require data and code. Panel C shows the results for journals that do not require data and code.

- From Table 3, Panel A we find that we are unable to replicate 21 papers because of "missing data or code," which constitutes the majority of our failed replications (55%).

- We are unable to replicate 9 papers (24% of failed replications) because of "incorrect data or code."

- We classify an unsuccessful replication as "incorrect data or code" when all variables are present in the dataset and the authors self-identify code for each of the key figures and tables we attempt to replicate.

- The author-provided code may finish executing and give different results or the code may not finish executing and still fall into this category.

Table 3: Failed Replication Results, Including Causes of Failure, By Journal Type

	Paper Count	Percentage of Sample
Panel A: All Journals		
Replication Unsuccessful	38	100
Unsuccessful Because of:		
Missing Public Data or Code	21	55
Incorrect Public Data or Code	9	24
Missing Software	2	5
Proprietary Data	6	16
Panel B: Journals With Mandatory Data and Code Policies		
Replication Unsuccessful	16	100
Unsuccessful Because of:		
Missing Public Data or Code	6	38
Incorrect Public Data or Code	5	31
Missing Software	1	6
Proprietary Data	4	25
Panel C: Journals Without Mandatory Data and Code Policies		
Replication Unsuccessful	22	100
Unsuccessful Because of:		
Missing Public Data or Code	15	68
Incorrect Public Data or Code	4	18
Missing Software	1	5
Proprietary Data	2	9

- **Table 4** shows our summary statistics for successful replications independent of the authors versus replications that were successful with the author's help.

- Overall, we find that contacting the authors marginally improves our success rate for replication.

- Of the 29 successful replications, we complete 22 without any help from the authors.

Table 4: Successful Replication Results By Journal Type

	Paper Count	Percentage of Sample
Panel A: All Journals		
Replication Successful	29	100
Successful:		
With Contacting Authors	7	24
Without Contacting Authors	22	76
Panel B: Journals With Mandatory Data and Code Policies		
Replication Successful	23	100
Successful:		
With Contacting Authors	3	13
Without Contacting Authors	20	87
Panel C: Journals Without Mandatory Data and Code Policies		
Replication Successful	6	100
Successful:		
With Contacting Authors	4	67
Without Contacting Authors	2	33

4. Conclusion and Recommendations

- We now turn to some recommendations that we feel would improve the ability for researchers to replicate and extend published articles, largely echoing the recommendations of McCullough, McGeary, and Harrison (2006).

  - *Mandatory data and code files should be a condition of publication.*

- Our replication success rate is significantly higher when we attempt to replicate papers from journals that have a mandatory replication data and code submission policy.

- We believe that replication files need to encompass both data and code.

- As shown in Table 2, the data-only archives at Economic Journal and Journal of Applied Econometrics only allow for replication of 4 of 20 papers (20%) that use non-confidential data, compared to the replication success rate of 23 of 35 papers (66%) that use non-confidential data from journals that require both data and code.

  - *An entry in the journal's data and code archive should indicate whether a paper without replication files in the journal's archive is exempt from the journal's replication policy.*

- Among papers that we believe were subject to a mandatory data and code policy, we are unable to acquire replication files for 6 of 35 papers (15%) even after emailing, and often receiving a response from, the authors.

- However, we are unsure whether these six papers are exempt from their respective journal's mandatory data and code policies, and the authors did not volunteer whether their papers are exempt in response to our requests for replication files.

- Therefore, we suggest that journals include an exemption entry in their replication archives.

- This note in the replication archives would have four virtues: (1) it is low-cost for the journal, (2) it would save authors who are exempt from submitting replication files from needing to respond to queries about replication files, (3) it would save would-be-replicators from searching for replication files for papers that are exempt from the journal's policy, and (4) it would identify those authors who are not compliant with the journal's mandatory data and code policy.

- We attempt to use the operating system-software version combination reported by the authors in their readme files, but we notice that very few readmes include the operating systemsoftware version combination used to conduct their analysis.

- When we ask authors about the operating system or software version they use to run their models, most authors do not recall this information.

- Although it is not a focus of our paper, we notice minor discrepancies for a selected subset of papers when running programs on different versions of Matlab (although the discrepancies are not large enough to change the key qualitative results).

  - *Readme files should contain an expected model estimation time.*

- Many macroeconomic models are estimated with Bayesian (i.e., Markov Chain Monte Carlo) methods, which can take a considerable amount of processing time to execute even under the best of circumstances.

- We encountered a few instances where we believed an estimation was executing, only to find out weeks later that the programs were stuck in an infinite loop and were supposed to run in much less time.

- In addition, frequently programs are not written to optimize computation time and also frequently written without a progress bar, so there is no way to track the expected completion time of estimation.

- A low-cost alternative to a progress bar is simply writing the expected estimation time in the readme file.

  - *Code that relies on random number generators should set seeds and specify the random number generator.*

- Optimization algorithms often rely on a set of initial conditions, which are commonly specified through a random number generator.

- For any research that relies on a random number generator, replication requires the same set of numbers that are generated in the published article.

  - *Readme files should clearly delineate which files should be executed in what order to produce desired results.*

- We now turn to two recommendations that will improve the ability of researchers to extend published work, in addition to merely replicating it.

  - *Authors should provide raw data in addition to transformed series.*

- While only the transformed data are needed to conduct replication of published results, raw data facilitate potential extensions of research. For example, raw data allow for the investigation of the effect that revisions to macroeconomic data have on previously published research, as in Croushore and Stark (2003) and Chang and Li (2015).

    - *Programs that replicate estimation results should carry out the estimation.*

- We notice that the replication files for a few papers run smoothly and exactly furnish the results of the tables and figures as published. However, oftentimes the results in tables and figures depend on a model's parameters being estimated.

- Some of these replication files, instead of estimating the models, take the relevant parameters as given to produce results in tables and figures. For verification of published results, and particularly for purposes of extending research, we assert that code that actually estimates the relevant models would be far more useful.