

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

Andrew C. Chang and Phillip Li

2015-083



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

Please cite this paper as:

Chang, Andrew C., and Phillip Li (2015). "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not"," Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System, <http://dx.doi.org/10.17016/FEDS.2015.083>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

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

Andrew C. Chang* and Phillip Li†

September 4, 2015

Abstract

We attempt to replicate 67 papers published in 13 well-regarded economics journals using author-provided replication files that include both data and code. Some journals in our sample require data and code replication files, and other journals do not require such files. Aside from 6 papers that use confidential data, we obtain data and code replication files for 29 of 35 papers (83%) that are required to provide such files as a condition of publication, compared to 11 of 26 papers (42%) that are not required to provide data and code replication files. We successfully replicate the key qualitative result of 22 of 67 papers (33%) without contacting the authors. Excluding the 6 papers that use confidential data and the 2 papers that use software we do not possess, we replicate 29 of 59 papers (49%) with assistance from the authors. Because we are able to replicate less than half of the papers in our sample even with help from the authors, we assert that economics research is usually not replicable. We conclude with recommendations on improving replication of economics research.

JEL Codes: B41; C80; C82; C87; C88; E01

Keywords: Data and Code Archives; Gross Domestic Product; GDP; Journals; Macroeconomics; National Income and Product Accounts; Publication; Research; Replication

*Chang: Board of Governors of the Federal Reserve System. 20th St. NW and Constitution Ave., Washington DC 20551 USA. +1 (657) 464-3286. a.christopher.chang@gmail.com. <https://sites.google.com/site/andrewchristopherchang/>.

†Li: Office of the Comptroller of the Currency. phillip.li@occ.treas.gov.

‡The views and opinions expressed here are those of the authors and are not necessarily those of the Board of Governors of the Federal Reserve System, the Department of the Treasury, or the Office of the Comptroller of the Currency. We thank B.D. McCullough, Bo Sun, and Kurt von Tish for helpful comments. We thank Tyler J. Hanson, Erik Larsson, Kim T. Mai, Anthony Marozzi, Shawn M. Martin, Tyler Radler, Adam Scherling, and John Stromme for research assistance. We also thank Felix Galbis-Reig and Spencer C. Li for technical assistance. We are responsible for any errors.

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 top 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. In extreme cases, replication can also facilitate the discovery of scientific fraud, as in the case of Broockman, Kalla, and Aronow (2015)'s investigation of the retracted article by LaCour and Green (2014).

This article is a cross-journal, broad analysis of the state of replication in economics.¹ We attempt to replicate articles using author-provided data and code files from 67 papers published in 13 well-regarded general interest and macroeconomics journals from July 2008 to October 2013. This sampling frame is designed to be more comprehensive across well-regarded economics journals than used by existing research. Previous work has tended to focus on a single journal, such as McCullough, McGeary, and Harrison (2006), who look

¹We follow existing work by, among others, Dewald, Thursby, and Anderson (1986); McCullough and Vinod (2003, 2004); McCullough, McGeary, and Harrison (2006); Hamermesh (2007) and Glandon (2010).

at the *Journal of Money, Credit and Banking* (JMCB); McCullough and Vinod (2003), who attempt to replicate a single issue of the AER (but end up replicating only Shachar and Nalebuff (1999) with multiple software packages); or Glandon (2010), who replicates a selected sample of nine papers only from the AER.

Using the author-provided data and code replication files, we are able to replicate 22 of 67 papers (33%) independently of the authors by following the instructions in the author-provided readme files. The most common reason we are unable to replicate the remaining 45 papers is that the authors do not provide data and code replication files. We find that some authors do not provide data and code replication files even when their article is published in a journal with a policy that requires submission of such files as a condition of publication, indicating that editorial offices do not strictly enforce these policies, although provision of replication files is more common at journals that have such a policy than at journals that do not. Excluding 6 papers that rely on confidential data for all of their results and 2 papers that provide code written for software we do not possess, we successfully replicate 29 of 59 papers (49%) with help from the authors. Because we successfully replicate less than half of the papers in our sample even with assistance from the authors, we conclude that economics research is usually not replicable.²

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

²This result is in line with recent evidence in replicating psychology studies by Open Science Collaboration (2015), where the authors fail to replicate the qualitative result of the majority of their sample of psychology experiments.

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.³ 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.^{4,5} 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

³For example, according to the 2014 social science edition of Thompson Reuters Journal Citation Reports, these journals represent half of the top 10 impact factor journals in economics and, aside from the *Canadian Journal of Economics*, all journals are in the top impact factor quartile for economics.

⁴This definition excludes lectures, symposiums, and literature reviews. We also exclude articles published prior to July 2008 when a comment on the article is published on or after July 2008.

⁵We choose an arbitrary start date of July 2008. We select the end date of October 2013 to match when we began this project.

(GDP), published by the Bureau of Economic Analysis (BEA), in an estimated model.⁶ We choose to focus on GDP because of its status as a standard macroeconomic statistic and its widespread use in research.⁷

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.⁹

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

⁶The authors may calibrate a subset of their model’s parameters and still have their paper fall within our sample. We exclude papers that have only completely calibrated models.

⁷Several of the most cited articles in economics use GDP or its predecessor gross national product (GNP) (Sims, 1980; Kydland and Prescott, 1982; Barro, 1991; Hodrick and Prescott, 1997). Recognizability of GDP extends to the popular press as well. For example, on the HBO late-night talk show “Last Week Tonight,” John Oliver cites GDP as a measure of the importance of the District of Columbia (Last Week Tonight, 2015).

⁸A paper’s primary contribution may be theoretical, but we focus only on the empirical component.

⁹The web appendix details what we interpret as the key result of every paper in our sample and is available on Chang’s website, <https://sites.google.com/site/andrewchristopherchang/research>.

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 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.¹⁰ 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,

¹⁰We email each of the authors sequentially using the email addresses obtained in the following three-step manner, moving down a step when our email was undeliverable: (1) the address listed on the author's personal site, (2) the address listed on the author's current institution's site, and (3) the address listed with the published paper. We wait at least one week between contacting each author. If a corresponding author is listed on the paper, then we start with contacting the corresponding author and work our way through the authors in the order they are presented in the paper. If there is no corresponding author, then we initiate contact with the first author. We stop querying different authors to request data and code after receiving a response from any author, unless we are specifically directed to by an author we had already queried or we fail to receive a response after an author-requested delay. We mark the data and code files as incomplete when we both are unable to locate a complete set of files online and also do not receive a complete set of files in response to our emails after waiting a minimum of one month for each author to respond to our queries. We stopped considering responses as of July 15th, 2015.

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. The authors we query whose papers we believe are subject to a data availability policy yet whose replication files we are unable to locate do not volunteer whether their papers are exempt from the policy, and we do not ask the authors for this information.

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.¹¹

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

¹¹If we already contacted the authors to request data or code but were having difficulty executing the code, then we only queried the authors whom we did not yet contact. We initiate contact with each author a maximum of one time.

that government spending is not merely a transfer or crowding out private investment).¹² We define success using this extremely loose definition to get an upper bound on what the replication success rate could potentially be.¹³ We allow for minimal re-working of the provided files, following the procedure of McCullough, McGeary, and Harrison (2006).¹⁴

One dimension where we are unable to follow the authors exactly is the software version they use. To execute the replications, we make use of the following software version-operating system combinations: Dynare 4.3 and 4.4.2 (Windows), EViews 6 and 7 (Windows), EViews 8 (Linux), Gauss 9.0.2 (Linux), Fortran f90 (Linux), Matlab R2008a and R2012a and R2013a (Windows), Matlab R2010a and R2012a (Linux), OX 6.30 (Windows), Oxmetrics 6.30 (Windows), Stata 11.0 and 13.1 (Windows), Stata 13.0 (Windows and Linux), R 2.15.1 and 3.0.1 and 3.0.2 and 3.0.3 and 3.1.0 (Linux), and RATS 7.10 (Linux).¹⁵ When available in the readme, we attempt to run the software version-operating system combination specified by the authors. When the replication files fail to execute on a given software version-operating system combination, the author readme did not specify a particular software version-operating system combination, and it appeared that the data and code were complete, we email the authors to find out which combination they use.

3 Results

This section presents summary statistics of our replication attempts.¹⁶

¹²This definition corresponds to a rating of three out of five, “minor discrepancies,” or better by Glandon (2010), and “partially successful replication” or better by McCullough, McGeary, and Harrison (2006).

¹³This definition is less stringent than the definition for replication success of McCullough, McGeary, and Harrison (2006, 2008)

¹⁴McCullough, McGeary, and Harrison (2008), in their appendix, suggest that “the author [whose study is being replicated] provides code such that data and code, when placed in the same subdirectory, will execute; and that the output from doing this also will be provided... and produces the results in his paper,” which implies that replication files should contain the data and code that requires no re-working. If the code is clearly missing the ability to replicate results, then we do not attempt to re-code the procedures ourselves.

¹⁵We check the replication results of a small sample of selected papers across different versions of Matlab for Windows and find very minor differences. None of the differences in results across different versions of Matlab are qualitatively significant. The web appendix lists the software version-operating system combination we use for each successfully replicated paper.

¹⁶Interested readers can find detailed results for each paper we successfully replicate in the web appendix on Chang’s website, <https://sites.google.com/site/andrewchristopherchang/research>.

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. The presence of a mandatory data and code policy does not necessarily imply a causal relationship from the policy to successful replication. Authors select which journals to submit papers to, taking into account idiosyncratic journal policies such as mandatory submission of replication data and code. However, we find that it is significantly easier to replicate published research that comes from journals that require authors to submit their data and code.

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%). As we outline in our methodology, for each of these unsuccessful replications we attempt to secure data and code from the authors by visiting their personal websites, visiting the journal websites (when the journal requires authors to submit data or code), and sending email requests. We classify an unsuccessful replication as “missing data or code” when at least one of two events occur: (1) the replication code file(s) are clearly missing necessary author-written functions for a subset or all of the key results or (2) the replication data

file(s) are missing at least one variable. If the replication data has a shorter data sample than reported in the paper, then we still attempt the estimation and do not necessarily classify the paper as “missing data or code.”

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.

We believe we do not have the needed software to run two papers (Senyuz, 2011; Jermann and Quadrini, 2012) because we are unable to locate a necessary packaged function in our versions of the appropriate software, because of significant syntax changes between software versions, or because the authors declared that they use a particular software version and we are aware that our software would not be compatible. However, it is tricky differentiating between an unsuccessful replication due to “incorrect data or code” or due to “missing software.” Because the implementation of packaged functions may differ across software versions even without syntax changes, we believe the number of failed replications we classify as “missing software” is a lower bound. It is possible that a paper we classify as “incorrect data or code” is actually replicable with the appropriate operating system-software combination, so some of the papers that we classify as “incorrect data or code” may belong in the “missing software” category. However, we cannot verify this statement without additional documentation.

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.

4 Conclusion and Recommendations

In this article, we attempt to replicate 67 papers from 13 well-regarded economics journals using author-provided data and code replication files. Improving on existing work evaluating the state of replication in economics, our sampling frame is broader across different journals and covers a large number of original research articles. We replicate 22 of 67 papers (33%) by using only the authors' data and code files, and an additional 7 papers (for a total of 29 papers, 43%) with assistance from the authors. The most common cause of our inability to replicate findings is that authors do not provide files to the journal replication archives, which constitutes approximately half of our failed replication attempts (21 of 38 papers, 55%). Because we are able to replicate less than half of the papers in our sample, we conclude that economics research is generally not replicable.

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.

- *Readme files should indicate the operating system-software version combination used in the analysis.*

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 system-software 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.*

Occasionally, we are presented with replication data and code that requires files to be executed in a particular order to furnish published results. In cases where the execution order is critical but unspecified, we spend a considerable amount of time attempting to determine the proper order of execution and, in some cases, ultimately fail to do so.

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.

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
<i>Panel A: All Journals</i>	29	61	67
<i>Panel B: Journals that Require Data and Code:</i>			
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
<i>Total for Journals that Require Data and Code</i>	23	35	39
<i>Panel C: Journals that Do Not Require Data and Code:</i>			
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
<i>Total for Journals that Do Not Require Data and Code</i>	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: Failed Replication Results, Including Causes of Failure, By Journal Type

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

Table 4: Successful Replication Results By Journal Type

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

References

- Alexopoulos, Michelle, “Read All About it!! What Happens Following a Technology Shock?” *American Economic Review* 101:4 (2011), 1144-1179.
- Alexopoulos, Michelle, and Jon Cohen, “Volumes of Evidence: Examining Technical Change in the Last Century Through a New Lens,” *Canadian Journal of Economics* 44:2 (2011), 413-450.
- Andrews, Donald W.K., “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica* 61:4 (1993), 821-856.
- Auerbach, Alan J., and Yuriy Gorodnichenko, “Measuring the Output Responses to Fiscal Policy,” *American Economic Journal: Economic Policy* 4:2 (2012), 1-27.
- Auerbach, Alan J., and Yuriy Gorodnichenko, “Corrigendum: Measuring the Output Responses to Fiscal Policy,” *American Economic Journal: Economic Policy* 5:3 (2013), 320-322.
- Bansak, Cynthia, Mary E. Graham, and Allan A. Zebede, “Business Cycles and Gender Diversification: An Analysis of Establishment-Level Gender Dissimilarity,” *American Economic Review: Papers & Proceedings* 102:3 (2012), 561-565.
- Barro, Robert J., “Economic Growth in a Cross Section of Countries,” *Quarterly Journal of Economics* 106:2 (1991), 407-443.
- Barro, Robert J., and Charles J. Redlick, “Macroeconomic Effects from Government Purchases and Taxes,” *Quarterly Journal of Economics* 126:1 (2011), 51-102.
- Baumeister, Christiane, and Gert Peersman, “Time-Varying Effects of Oil Supply Shocks on the US Economy,” *American Economic Journal: Macroeconomics* 5:4 (2013), 1-28.

- Bernanke, Ben S., "Editorial Statement," *American Economic Review* 94:1 (2004), 404.
- Broockman, David, Joshua Kalla, and Peter Aronow, "Irregularities in LaCour (2014)," Working Paper (2015).
- Card, David, and Alan B. Krueger, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review* 84:4 (1994), 772-793.
- Card, David, and Alan B. Krueger, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply," *American Economic Review* 90:5 (2000), 1397-1420.
- Canova, Fabio, and Luca Gambetti, "Do Expectations Matter? The Great Moderation Revisited," *American Economic Journal: Macroeconomics* 2:3 (2010), 183-205.
- Carey, Colleen, and Stephen H. Shore, "From the Peaks to the Valleys: Cross-State Evidence on Income Volatility Over the Business Cycle," *Review of Economics and Statistics* 95:2 (2013), 549-562.
- Chang, Andrew C., and Phillip Li, "Measurement Error in Macroeconomic Data and Economics Research: Data Revisions, Gross Domestic Product, and Gross Domestic Income," Working Paper (2015).
- Chen, Han, Vasco Curdia, and Andrea Ferrero, "The Macroeconomic Effects of Large-scale Asset Purchase Programmes," *Economic Journal* 122:564 (2012), F289-F315.
- Clark, Todd E., and Michael W. McCracken, "Averaging Forecasts from VARs with Uncertain Instabilities," *Journal of Applied Econometrics* 25:1 (2010), 5-29.
- Corsetti, Giancarlo, André Meier, and Gernot J. Müller, "Fiscal Stimulus with Spending Reversals," *Review of Economics and Statistics* 94:4 (2012), 878-895.

- Croushore, Dean, and Tom Stark, "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?" *Review of Economics and Statistics* 85:3 (2003), 605-617.
- D'Agostino, Antonello, and Palo Surico, "A Century of Inflation Forecasts," *Review of Economics and Statistics* 94:4 (2012), 1097-1106.
- Den Haan, Wouter J., and Vincent Sterk, "The Myth of Financial Innovation and the Great Moderation," *Economic Journal* 121:553 (2011), 707-739.
- Dewald, William G., Jerry G. Thursby, and Richard G. Anderson, "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project," *American Economic Review* 76:4 (1986), 587-603.
- Evanschitzky, Heiner, Carsten Bamgarth, Raymond Hubbard, and J. Scott Armstrong, "Replication Research in Marketing Revisited: A Note on a Disturbing Trend," *Journal of Business Research* 60:4 (2007), 411-415.
- Favero, Carlo, and Giavazzi, Francesco, "Measuring Tax Multipliers: The Narrative Method in Fiscal VARs," *American Economic Journal: Economic Policy* 4:2 (2012), 69-94.
- Fisher, Jonas D.M., and Ryan Peters, "Using Stock Returns to Identify Government Spending Shocks," *Economic Journal* 120:544 (2010), 414-436.
- Gabaix, Xavier, "The Granular Origins of Aggregate Fluctuations," *Econometrica* 79:3 (2011), 733-772.
- Glandon, Philip, "Report on the American Economic Review Data Availability Compliance Project," Working Paper (2010).
- Gilchrist, Simon, and Egon Zakrajšek, "Credit Spreads and Business Cycle Fluctuations," *American Economic Review* 102:4 (2012), 1692-1720.

- Hall, George J., and Thomas J. Sargent, "Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics," *American Economic Journal: Macroeconomics* 3:3 (2011), 192-214.
- Hamermesh, Daniel S., "Viewpoint: Replication in Economics," *Canadian Journal of Economics* 40:3 (2007), 715-733.
- Hansen, Peter R., Asger Lunde, and James M. Nason, "The Model Confidence Set," *Econometrica* 79:2 (2011), 453-497.
- Herndon, Thomas, Michael Ash, and Robert Pollin, "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff," *Cambridge Journal of Economics* 38:2 (2014), 257-279.
- Heutel, Garth, "How Should Environmental Policy Respond to Business Cycles? Optimal Policy Under Persistent Productivity Shocks," *Review of Economic Dynamics* 15:2 (2012), 244-264.
- Hodrick, Robert J., and Edward C. Prescott, "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking* 29:1 (1997), 1-16.
- Hoxby, Caroline, "Does Competition Among Public Schools Benefit Students and Taxpayers," *American Economic Review* 90:5 (2000), 1209-1238.
- Hoxby, Caroline, "Does Competition Among Public Schools Benefit Students and Taxpayers? A Reply to Rothstein," *American Economic Review* 97:5 (2007), 2038-2055.
- Inoue, Atsushi, and Barbara Rossi, "Identifying the Sources of Instabilities in Macroeconomic Fluctuations," *Review of Economics and Statistics* 93:4 (2011), 1186-1204.
- Ireland, Peter N., "On the Welfare Cost of Inflation and the Recent Behavior of Money Demand," *American Economic Review* 99:3 (2009), 1040-1052.

- Jermann, Urban, and Vincenzo Quadrini, "Macroeconomic Effects of Financial Shocks," *American Economic Review* 102:1 (2012), 238-271.
- Kilian, Lutz, "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review* 99:3 (2009), 1053-1069.
- Kormilitsina, Anna, "Oil Price Shocks and the Optimality of Monetary Policy," *Review of Economic Dynamics* 14:1 (2011), 199-223.
- Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, "The Aggregate Demand for Treasury Debt," *Journal of Political Economy* 120:2 (2012), 233-267.
- Kydland, Finn E., and Edward C. Prescott, "Time to Build and Aggregate Fluctuations," *Econometrica* 50:6 (1982), 1345-1370.
- LaCour, Michael J., and Donald P. Green, "When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality," *Science* 346:6215 (2014), 1366-1369.
- Last Week Tonight with John Oliver, Segment on Washington D.C. Voting Rights, airdate August 2, 2015.
- Levitt, Steven, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review* 87:3 (1997), 270-290.
- Levitt, Steven, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Reply," *American Economic Review* 92:4 (2002), 1244-1250.
- Mavroeidis, Sophocles, "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review* 100:1 (2010), 491-503.
- McCrary, Justin, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment," *American Economic Review* 92:4 (2002), 1236-1243.

- McCullough, B.D., Kerry Anne McGeary, and Teresa D. Harrison, "Lessons from the JMCB Archive," *Journal of Money, Credit and Banking* 38:4 (2006), 1093-1107.
- McCullough, B.D., Kerry Anne McGeary, and Teresa D. Harrison, "Do Economics Journal Archives Promote Replicable Research?" *Canadian Journal of Economics* 41:4 (2008), 1406-1420.
- McCullough, B.D., and H.D. Vinod, "Verifying the Solution from a Nonlinear Solver: A Case Study," *American Economic Review* 93 (2003), 873-892.
- McCullough, B.D., and H.D. Vinod, "Verifying the Solution from a Nonlinear Solver: Reply," *American Economic Review* 94:1 (2004), 391-396.
- Mertens, Karel, and Morten O. Ravn, "Understanding the Aggregate Effects of Anticipated and Unanticipated Tax Policy Shocks," *Review of Economic Dynamics* 14:1 (2011), 27-54.
- Mertens, Karel, and Morten O. Ravn, "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review* 103:4 (2013), 1212-1247.
- Morley, James, and Jeremy Piger, "The Asymmetric Business Cycle," *Review of Economics and Statistics* 94:1 (2012), 208-221.
- Nakov, Anton, and Andrea Pescatori, "Oil and the Great Moderation," *Economic Journal* 120:543 (2010), 131-156.
- Neumark, David, and William Wascher, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment," *American Economic Review* 90:5 (2000), 1362-1396.
- Open Science Collaboration, "Estimating the Reproducibility of Psychological Science," *Science* 349:6251 (2015), aac4716.

- Ramey, Valerie A., "Identifying Government Spending Shocks: It's all in the Timing," *Quarterly Journal of Economics* 126:1 (2011), 1-50.
- Reinhart, Carmen M., and Kenneth S. Rogoff, "Growth in a Time of Debt," *American Economic Review* 100:2 (2010), 573-578.
- Reis, Ricardo, and Mark W. Watson, "Relative Goods' Prices, Pure Inflation, and The Phillips Correlation," *American Economic Journal: Macroeconomics* 2:3 (2010), 128-157.
- Romer, Christina D., and David H. Romer, "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review* 100:3 (2010), 763-801.
- Rothstein, Jesse, "Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)," *American Economic Review* 97:5 (2007), 2026-2037.
- Shachar, Ron, and Barry Nalebuff, "Follow the Leader: Theory and Evidence on Political Participation," *American Economic Review* 83:3 (1999), 525-547.
- Schmitt-Grohé, Stephanie, and Martín Uribe, "Business Cycles with a Common Trend in Neutral and Investment-Specific Productivity," *Review of Economic Dynamics* 14:1 (2011), 122-135.
- Schmitt-Grohé, Stephanie, and Martín Uribe, "What's News in Business Cycles," *Econometrica* 80:6 (2012), 2733-2764.
- Senyuz, Zeynep, "Factor Analysis of Permanent and Transitory Dynamics of the US Economy and the Stock Market," *Journal of Applied Econometrics* 26:6 (2011), 975-998.
- Sims, Christopher A., "Macroeconomics and Reality," *Econometrica* 48:1 (1980), 1-48.
- Taylor, John B., "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* 39 (1993), 195-214.

Thompson Reuters' Web of Science Journal Citation Reports: Social Science Edition.

<https://isiknowledge.com/> (2014).