Critical Finance Review, 2022, 11: 185–206

Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say "Often Not"

Andrew C. Chang[1]
Phillip Li[2]*

[1]*Board of Governors of the Federal Reserve System, USA;*
a.christopher.chang@gmail.com
[2]*Office of Financial Research, U.S. Department of the Treasury, USA;*
phillip.li@ofr.treasury.gov.

ABSTRACT

We attempted to replicate 67 macroeconomic papers published in 13 well-regarded economics journals using author-provided replication files that included both data and code by following a preanalysis plan. Aside from six papers that used confidential data, we obtained data and code replication files for 29 of 35 papers (83%) that were required to provide such files as a condition of publication, compared to 11 of 26 papers (42%) that were not required to provide data and code replication files. Defining replication success as our ability to use the author-provided data and code files to produce the key qualitative conclusions of the original paper, we successfully replicated 22 of 67 papers (33%) without contacting the authors. Excluding the six papers that used confidential data and the two papers that used software we did not possess, we replicated 29 of 59 papers (49%) with assistance from the authors. Because we were able to replicate less than half of the papers in our sample even with help from the authors, we assert that economics research is often not replicable. We conclude with recommendations on improving replication of economics research.

*The views and opinions expressed here are ours and are not necessarily those of the Board of Governors of the Federal Reserve System, the Department of the Treasury, the Office of Comptroller of the Currency, or the Office of Financial Research. The earlier version of this paper was titled "Is

Online Appendix available from:
http://dx.doi.org/10.1561/104.00000053_app
Replication data available from:
http://dx.doi.org/10.1561/104.00000053_supp
ISSN 2164-5744; DOI 10.1561/104.00000053

Keywords: Data and code files, Gross domestic product, GDP, Macroeconomics, National income and product accounts, Preanalysis, Publication bias, Replication

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

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 effect of police on crime; and, more recently, Herndon *et al.* (2014) and Reinhart and Rogoff (2010) on fiscal austerity. In extreme cases, replication can also facilitate the discovery of scientific fraud, as in the case of Broockman *et al.* (2015)'s investigation of the retracted article by LaCour and Green (2014).[1]

Our article is a cross-journal, broad analysis of the state of replication in economics.[2] We attempted to replicate articles using author-provided data and code files from 67 macroeconomic papers in 13 well-regarded journals published from July 2008 to October 2013. Previous research on replication in economics has tended to focus on a single journal, such as McCullough *et al.* (2006), who

Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not.'" This paper was started while Phillip Li was at the Office of the Comptroller of the Currency and before he joined the Office of Financial Research. We thank Sarena F. Goodman, Jan H. Höeffler, Lawrence Katz, Evan F. Koenig, Robert J. MacCoun, B.D. McCullough, Michael Nahas, Min Qi, W. Robert Reed, Bo Sun, Richard H. Thaler, J. Luke Van Cleve, Kurt von Tish, Ivo Welch, Christian Zimmermann, and seminar participants at the 2017 ASSA annual meeting, BEA, the Board, the 2016 CEA annual meeting, CEU, Duke, FRB Cleveland, FRB St. Louis, GWU, OCC, OFR, Penn State, UC - Irvine, and the 2016 WEAI annual meeting for helpful comments. We thank Tyler J. Hanson, Erik Larsson, Dustin Lartigue, Kim T. Mai, Anthony Marcozzi, Shawn M. Martin, Tyler Radler, Adam Scherling, and John Stromme for research assistance. We also thank Sophia Allison, Felix Galbis-Reig, Spencer C. Li, Misty Mirpuri, David Ratner, and Jeffrey Westphal for technical assistance. We are responsible for any errors.

[1] Similarly, see Whitlock and Woodward (2016) on how the Department of Defense reportedly took steps to obstruct replication.

[2] We followed existing work by, among others, Dewald *et al.* (1986), Duvendack *et al.* (2015), Glandon (2010), Hamermesh (2007), McCullough *et al.* (2006), McCullough and Vinod (2003, 2004).

looked at the *Journal of Money, Credit and Banking* (JMCB); McCullough and Vinod (2003), who attempted to replicate a single issue of the AER (but ended up successfully replicating only Shachar and Nalebuff, 1999) or Glandon (2010), who replicated a selected sample of nine papers only from the AER.

We defined a successful replication in which we could use author-provided files to produce the key qualitative results of the paper.[3] Under this definition, we were able to replicate 22 of 67 papers (33%) by following the instructions in the author-provided readme files. The most common reason why we were unable to replicate the remaining 45 papers was that the authors did not provide data and code replication files. We found that some authors did not provide data and code replication files even when their article was published in a journal with a policy that required submission of such files as a condition of publication. This lack of files indicates that editorial offices did not strictly enforce these policies, although provision of replication files was more common at journals that had such a policy than at journals that did not. Excluding six papers that relied on confidential data for all of their results and two papers that provided us code written for versions of Gauss we did not possess, we successfully replicated 29 of 59 papers (49%) with help from the authors.[4]

We investigated how our replication success rate correlated with a paper's characteristics. Our replication success rate for papers published in journals that required replication data and code files as a condition of publication was about three times higher than for journals that did not require such files (59% vs. 21%). Papers that were cited more heavily were more likely to be replicable, with a one natural log point increase in a paper's Google Scholar citation count correlating with about a 10 percentage points increased probability of successful replication. The type of job held by a paper's author (classified as either nonacademic, non-tenured academic, or tenured academic) was uncorrelated with replication success. Papers that acknowledged a source of funding were around twice as likely to be successfully replicated (55% vs. 26%).

Our replication success rate was higher than those reported by existing studies of replication in economics. McCullough *et al.* (2006) found 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 articles' use of non-proprietary data, and lack of assistance from the original article's authors. Adding the condition that the JMCB's archive contained data and code replication files increased their success rate to 14 of 62 papers (23%). Our comparable success rates were 22 of 59 papers (37%), conditioned on our having appropriate software and non-proprietary data, and 22 of 38 papers (58%) when we imposed

[3]There are widespread differences in the definitions of "replication." See, for example, Clemens (2015).

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

the additional requirement of having data and code files. Dewald *et al.* (1986) successfully replicated 7 of 54 papers (13%) from the JMCB, conditioned on the replicators having data and code files, the original articles' use of non-confidential data, help from the original articles' authors, and appropriate software. Our comparable figure was 29 of 38 papers (76%).

2 Methodology and Sample

We followed a preanalysis plan by defining the methodology in this section prior to executing any analysis. This plan had three benefits: (1) we set a uniform standard for analyzing the results of models; (2) we avoided hindsight bias in model selection and analysis; and (3) we avoided pretesting our results.

Our sample included 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 chose papers from these journals because of the relative likelihood that such papers will have a policy effect and also influence future research.[5] We did not select these journals to single out a particular author, methodology, institution, or ideology.

From our sample of journals, we browsed for original research articles published in issues from July 2008 to October 2013.[6,7] Within these issues, we identified all papers with the following three characteristics: (1) an empirical component; (2) model estimation with only US data; and (3) a key empirical result regarding output, measured by US gross domestic product (GDP),[8,9] published by the Bureau of Economic Analysis (BEA), in an estimated model.[10] We chose to focus on output,

[5]For example, according to Thomson Reuters' Web of Science Journal Citation Reports (2014), social science edition, these journals represented half of the top 10 impact factor journals in economics and, aside from the *Canadian Journal of Economics*, all journals were in the top impact factor quartile for economics.

[6]This definition excluded lectures, symposiums, and literature reviews. We also excluded articles published prior to July 2008 when a comment on the article was published on or after July 2008.

[7]We chose an arbitrary start date of July 2008. We selected the end date of October 2013 to match when we began this paper.

[8]We also searched for papers that used gross domestic income (GDI) as a measure of output for use in our related work on measurement error, Chang and Li (2015b), but we did not find any papers that used GDI.

[9]Following submission to the *Critical Finance Review*, we realized that Rudebusch and Wu (2008) used capacity utilization as their measure of the output gap, as opposed to GDP minus potential GDP. Therefore, Rudebusch and Wu (2008) should not have been in our original sample, but because we included Rudebusch and Wu (2008) in our preanalysis plan and did not discover our error until after we finalized our results, our paper treats Rudebusch and Wu (2008) as in sample.

[10]The authors may have calibrated a subset of their model's parameters and still have their paper fall within our sample. We excluded papers that had only completely calibrated models.

measured as GDP, because of GDP's status as a standard macroeconomic statistic and its widespread use in research.[11]

For each paper in this set, we attempted to replicate the key empirical results.[12] Defining a key result was subjective and required judgmental decisions from us. We looked for when the authors referred to output as driving a key result, or when a discussion of output was featured either in the abstract or prominently in the introduction (or both). Once we identified the key results from the text, we attempted to replicate the figures and tables that corresponded to the key results.

We found 67 papers that matched our criteria. Of these papers, six papers used proprietary data for all of the key results, so we did not include them in our replication exercise (Alexopoulos, 2011; Alexopoulos and Cohen, 2011; Bansak *et al.*, 2012; Fisher and Peters, 2010; Gilchrist and Zakrajšek, 2012; Hall and Sargent, 2011). If a subset of the key results could be obtained using non-proprietary data, then we attempted to replicate those results.

For the remaining papers that used public data and were published in journals that maintained data and code archives, we downloaded the replication files provided by the authors through journal online archives. Like McCullough *et al.* (2006) and Vlaeminck and Herrmann (2015), we found that journal data and code archives were incomplete. Of the 35 papers that used public data and that were published in journals that required data and code replication files, we only obtained files for 28 papers (80%) from journal archives.

When we were unable to obtain replication data and code files from journal websites, either because the mandatory files were missing or because the paper was not subject to a data availability policy, we checked the personal websites of each of the authors for replication files. If we were unable to locate replication files online, then we emailed each of the authors individually requesting the replication files.[13]

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

[12] A paper's primary contribution may have been theoretical, but we focused only on the empirical component.

[13] We emailed 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 waited at least 1 week between contacting each author. If a corresponding author was listed on the paper, then we started with corresponding author and worked our way through the authors in the order they were presented in the paper. If there was no corresponding author, then we initiated contact with the first author. We stopped querying different authors to request data and code after receiving a response from any author, unless we were specifically directed to by an author we had already queried or we failed to receive a response after an author-requested delay. This procedure implied that, for example, if we asked the corresponding author for data or code and the corresponding author declined our request, then we did not contact the remaining authors. We marked the data and code files as incomplete when we both were unable to locate a complete set of files online and also did 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 on July 15th, 2015.

Of the seven papers that used public data, were subject to a data and code policy, and did not have replication files on the journal's website, this procedure netted us one additional set of replication files. Therefore, we were unable to locate either data or code replication files for 6 of 35 papers (17%) that were published in journals that required submission of data and code replication files. For papers published in journals without a data and code availability policy and that used public data, we were unable to obtain data or code replication files for 15 of 26 papers (58%).

To determine whether a paper was subject to a data availability policy, we checked the implementation dates of the journal data policies and compared them to the publication and submission dates of the published work. If the journal's website did not allow us to extract this information, then we queried the editorial office as to when their data availability policy became effective. We did not ask the editorial offices whether a particular paper was subject to a data availability policy. Aside from papers with proprietary data, we found that journal data archives did not provide lists of potentially exempt papers. Therefore, we were unable to determine whether a paper was exempt for a reason other than using proprietary data, although we were not aware of reasons why journals would grant a paper a data and code exemption other than for proprietary data. The authors we queried whose papers we believed were subject to a data availability policy yet whose replication files we were unable to locate did not volunteer whether their papers were exempt from the policy, and we did not ask the authors for this information.

For the papers for which we were able to obtain data and code replication files, we attempted to replicate the key results of the papers using only the instructions provided in the author readme files. If the readme files were insufficient or if the replication files were incomplete (or both) and the paper was subject to a replication policy, then we emailed the corresponding author (if no corresponding author, then the first author) for either clarification or to request the missing files. If we did not receive a response within a week, then we queried the second author, and so on, until we contacted all authors.[14]

We defined a successful replication as when the authors or journal provided data and code files that allowed us to qualitatively reproduce the key results of the paper. For example, if the paper estimated 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).[15] We defined success using this

[14]If 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 initiated contact with each author a maximum of one time. We treated authors with multiple papers in our sample as independent observations.

[15]This definition corresponds to replication—verification by Clemens (2015)'s Table 1 with the added condition of the authors providing us data and code files, a rating of three out of five, "minor discrepancies," or better by Glandon (2010), and "partially successful replication" or better by McCullough *et al.* (2006). Interested readers can find all of our replication results, including quantitative estimates, in the Online Appendix.

extremely loose definition to get an upper bound on what the replication success rate could be.[16] We allowed for some re-working of the provided files, following the procedure of McCullough *et al.* (2006).[17]

One dimension where we were unable to follow the authors exactly was the software version they used. To execute the replications, we used the following software version-operating system combinations: Dynare 4.1.1, 4.3.3, and 4.4.2 (Windows), EViews 6, 7, and 8 (Windows), Gauss 9.0.2 (Linux), FAME 10.2 64-bit, 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).[18] When available in the readme, we attempted to run the software version-operating system combination specified by the authors. When the replication files failed 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 emailed the authors to find out which combination they used.

In both the Board of Governors of the Federal Reserve System Finance and Economics Discussion Series version of this paper, Chang and Li (2015a), and in the *American Economic Review: Papers & Proceedings* version, Chang and Li (2017), we only reported the key results of papers that we successfully replicated. For submission to the *Critical Finance Review* and at the request of the editor, the appendix lists the key results and dependent variables (where applicable) for all papers in our sample, including those papers that we failed to replicate.

3 Summary of Results

This section presents a summary of our replication results, using the statuses of our replication attempts as of when we published the working paper version of this paper, Chang and Li (2015a).[19] This distinction matters because in Chang and Li (2015a) we mistakenly thought that we lacked the files to replicate Shore (2010). However, after preparing our replication datasets for the *Critical Finance*

[16]This definition was less stringent than the definition for replication success of McCullough *et al.* (2006, 2008).

[17]McCullough *et al.* (2008), in their appendix, suggested 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 was clearly missing the ability to replicate results, then we did not attempt to re-code the procedures ourselves.

[18]We checked the replication results of a small sample of selected papers across different versions of Matlab for Windows and found very minor differences. None of the differences in results across different versions of Matlab were qualitatively significant.

[19]Interested readers can find detailed results for each paper we have results for in the appendix on Chang's website, https://sites.google.com/site/andrewchristopherchang/research.

Review we realized that the archive that we had for Shore (2010), which came directly from Stephen Shore, was complete.

Figure 1 displays a Sankey diagram, breaking down our replication success and failure rates into the causes of success and failure. From our initial sample of 67 papers in the leftmost column, moving to the top branch shows that our unconditional replication success rate was 29 of 67 papers (43%). Breaking down that success rate, we were able to replicate 22 of these papers independently of the authors, and another seven papers after soliciting help from the authors.

The bottom branch of Figure 1 shows the causes for our failure to replicate the remaining 38 of 67 papers (57%).

The most common cause of our failure to replicate articles, shown in the top leaf of the bottom branch, was that both authors and journals did not provide us at least some of the code or data files. This lack of files accounted for 21 of our failed replications. As we outlined in our methodology, for each paper we attempted to secure data and code from the authors by visiting their personal websites, visiting the journal websites (when the journal required authors to submit data or code), and sending email requests. For these 21 papers at least one of two events occurred: (1) the replication code file(s) were clearly missing necessary author-written functions for a subset or all of the key results or (2) the replication data file(s) were missing at least one variable. To be clear: some of these authors did provide us with a set of files, but those files were incomplete.[20]

Moving down to the next leaf, the second most common cause of our failure to replicate papers was "incorrect data or code," which accounted for nine of our failed replications. For this group, we were provided code and data where all variables were present in the dataset and the authors self-identified code for each of the key figures and tables, yet we were unable to replicate the published results. For this group, the author-provided code may have finished executing but gave different qualitative results than published or the code may not have finished executing.

As shown in the third leaf, six papers used proprietary data for all of the key results, so we were unable to replicate them.

Finally, as shown in the bottom leaf, we believe we did not have the needed software to replicate two papers. However, it was 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 classified as "missing software" was a lower bound. It is possible that a paper we classified as "incorrect data or code" is actually replicable with the appropriate operating system-software combination, so some of the papers that we

[20] If the replication data had a shorter data sample than reported in the paper, then we still attempted the estimation and did not necessarily classify the paper as a failed attempt due to missing data or code.

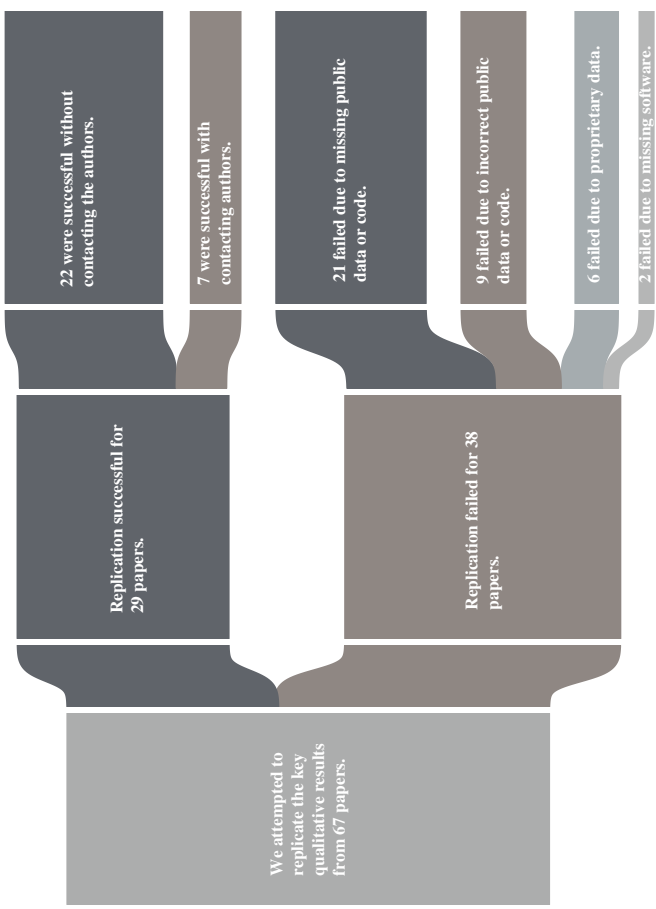

Figure 1: Causes of Replication Success or Failure.

Description: This figure is a Sankey diagram that separates our replication successes and failures by cause.

Interpretation: Our overall replication success rate was a bit below fifty percent. The most common cause of a failed replication was that authors did not provide us with either their data or their code files.

classified as "incorrect data or code" may belong in the "missing software" category. However, we could not verify this statement without additional documentation.

4 What About Trying Harder to Obtain Replication Files?

The procedure that we specified in our preanalysis plan for soliciting assistance from authors was to initiate contact with each author a maximum of once. Following submission to the *Critical Finance Review* and at the behest of the editor, Ivo Welch, to see whether additional effort spent on obtaining files would lead to more successfully replicated papers, we attempted to re-contact all of the authors that ignored our initial requests for data or code replication files.

Our procedure for re-contacting these authors was identical to our original procedure from our preanalysis plan, effectively doubling the number of emails we sent to the authors of the following papers: Bai and Wang (2011), Clements and Galvão (2009, 2013), Fève and Guay (2010), Gordon (2010), Ilbas (2012), Milani (2011), Perotti (2012), and Piazzesi and Schneider (2010). This second round of emails netted us one additional set of code, for Clements and Galvão (2013). With help from Ana Beatriz Galvão, we were able to replicate most of the key results from Tables 4 and 5 of Clements and Galvão (2013).[21] We did not re-contact authors that denied our initial requests for data and code.

5 Heterogeneity in Paper Characteristics and Replication Success

5.1 Did Journal Policies on Replication Files Matter?

Yes.

Figure 2 displays a Sankey diagram breaking down our replication success and failures into whether the journal had a policy that required authors to submit data and code files as a condition of publication.

From the top branch of Figure 2, we successfully replicated 23 of 39 papers (59%) from journals that required these files, shown in the top leaf. This rate compares to a success rate of 6 of 28 papers (21%) from journals that did not require such files, shown in the bottom leaf. These replication rates by journal type were similar when we only considered papers with publicly available data: we successfully replicated 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 that did not have 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 selected which journals to submit papers to, taking into account idiosyncratic journal policies such as mandatory submission of replication data and code. However, we

[21] See the Appendix for results.

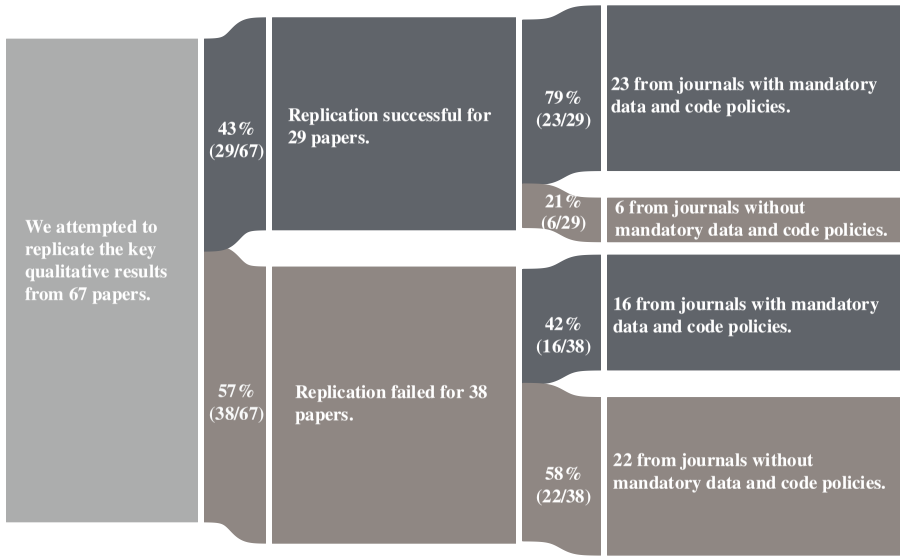

Figure 2: Replication Success or Failure by Journal Type.

Description: This figure is a Sankey diagram that separates our replication successes and failures by whether the journal that the paper was published in required authors to submit their data and code as a condition of publication.

Interpretation: We found that it was easier to replicate papers from journals that required authors to submit their data and code files.

found that it was significantly easier to replicate published research that came from journals that required authors to submit their data and code.

We specified looking at journal replication file policies before submitting to *Critical Finance Review*. For the next three subsections on citation counts, academic rank, and grants we added the analysis after establishing our preanalysis plan and our initial submission to the *Critical Finance Review*. Therefore, the following subsections on citation counts, academic rank, and grants treat Shore (2010) and Clements and Galvão (2013) as successful replications.

5.2 Were More Heavily Cited Papers Easier to Replicate?

Yes.

Using our entire sample of 67 papers, Figure 3 plots a replication success dummy variable against the natural log of Google Scholar citation counts.[22]

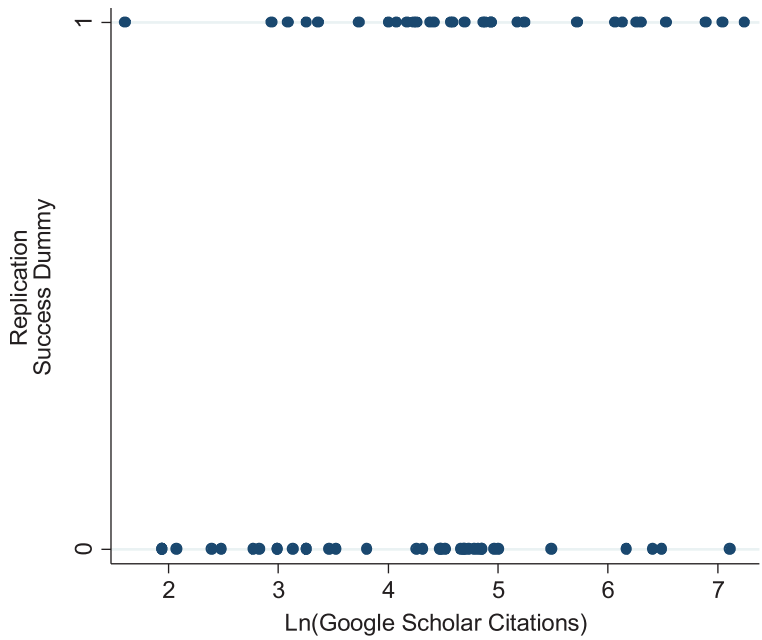

Figure 3: Replication Success vs. Citations.

Description: This figure plots a replication success dummy against the paper's Google Scholar citations as of July 12th, 2016.

Interpretation: We found that more heavily cited papers were easier to replicate.

[22]The replication success variable is equal to 1 for a successful replication and 0 otherwise.

	(1)	(2)	(3)	(4)
ln(Citations)	9.8		11.9	
	(3.8)*		(3.7)*	
AboveMedianCites*ln(Citations)		12.5		14.9
		(5.1)*		(4.8)*
BelowMedianCites*ln(Citations)		14.6		18.1
		(7.4)*		(7.1)*
Time dummies	No	No	Yes	Yes
Pseudo-R^2	0.05	0.06	0.14	0.15

Table 1: Probit Regressions of Replication Success Dummy on Natural Log of Citations.

Description: This table shows probit regressions of a replication success dummy on citations, with average marginal effects expressed as the percentage increase in the probability of successful replication for a one natural log point increase in Google Scholar citation counts, recorded on July 12th, 2016. Columns (1) and (3) estimate one average marginal effect across all papers. Columns (2) and (4) show separate average marginal effects for papers that were above and below the median citation count. Standard errors computed using the delta method in parentheses. *: significant at the 5% level or lower.

Interpretation: We found that more heavily cited papers were easier to replicate.

We collected the Google Scholar citation counts for the papers in our sample on July 12th, 2016. From visual inspection of Figure 3, it appears that more heavily cited papers were easier to replicate, although we still failed to replicate a few heavily cited papers including a few that had over 500 citations.

Table 1 shows average marginal effects from probit regressions of the replication success dummy on citation counts. From column (1), we estimated that a one natural log point increase in citation counts correlated with a 9.8 percentage point increase in the probability of successful replication. Splitting into separate effects for papers with above or below the median citation count of the sample, shown in column (2), showed relative homogeneity in effect size over this dimension.

A paper's observed citation counts at a given date is a right-censored observation of a paper's lifetime citations. To address censoring and potential cohort effects in citation count, columns (3) and (4) show our estimated probit model with a full set of year dummies. The results are largely similar to using raw citations in columns (1) and (2): we found that more heavily cited papers were easier to replicate.

5.3 Did Academic Rank Matter for Replication Success?

No.

Again using our entire sample of 67 papers combined with online curriculum vitaes, we classified each author's primary position at the time of their paper's publication into one of three categories: (1) nonacademic; (2) nontenured academic

or student; and (3) tenured academic.[23] We treated each coauthor on a paper as an independent observation, because it was unclear who the representative coauthor should have been in a multi-authored paper. Probit regressions of the replication success dummy on a full set of category dummies for primary position showed no statistically significant average marginal effects.[24]

5.4 Were Papers that Disclosed Funding Easier to Replicate?

Yes.

To check for funding, we browsed the front and back matter to check whether authors disclosed funding. From these self-disclosures, 16 of 29 papers (55%) that we successfully replicated acknowledged a source of funding, compared to 10 of 38 papers (26%) where we were unsuccessful.[25]

6 Conclusion and Recommendations

In this article, we attempted 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 research, we designed a broad sampling frame that spanned different journals and covered a large number of original research articles. We replicated 22 of 67 papers (33%) by using only the authors' data and code files, and an additional seven papers (for a total of 29 papers) with assistance from the authors. The most common cause of our inability to replicate findings was that authors did not provide files to the journal replication archives, which constitutes approximately half of our failed replication attempts (21 of 38 papers, 55%). Because we were able to replicate less than half of the papers in our sample, we conclude that economics research is often 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 *et al.* (2006).

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

[23] We assumed that Associate Professors (North American system) and Senior Lecturers (European system) and up had tenure, unless the vitae specified a lack of tenure.

[24] With nonacademic jobs as the omitted category, the estimated average marginal effects (standard errors) were −3.4 (13.3) percentage points for nontenured academics and 4.5 (11.5) percentage points for tenured academics, with a pseudo R^2 of less than one basis point.

[25] Checking the front and back matter of papers for funding sources does not say too much about whether funding was used for those particular articles due to two factors: (1) the practice of writing grants for completed papers but instead using the grant money for in-progress work; and (2) front and back matter data are self-reported by authors.

Our replication success rate was significantly higher when we attempted to replicate papers from journals that had a mandatory replication data and code policy, as opposed to a data-only or no submission policy.

- *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.*

This entry 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.*

Although it was not a focus of our paper, we noticed minor discrepancies for a selected subset of papers when running programs on different versions of Matlab.

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

Many estimation routines can take a considerable amount of 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.[26]

- *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, exact replication requires the same set of numbers

[26]We are elated to report that the working paper version of this article Chang and Li (2015a) caused the *Review of Economic Dynamics* to change its data and code policy to include several of our recommendations, very soon after the paper was posted to the Board of Governors of the Federal Reserve System Finance and Economics Discussion Series on October 5th, 2015. The *Review of Economic Studies* also changed its data and code policy to include our recommendations sometime between October 5th, 2015 and February 21st, 2016, but we are unsure whether they changed their policy because of our working paper.

that are generated in the published article. We hope that the conclusions of a given article would be robust to the random number generator and seed used. For purposes of exactly replicating an author's result, the random number generator and seed would be required.

- *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 replicate published results, raw data facilitate extensions of research.[27] 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 (2015b).

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

Some of these replication files we obtained, instead of estimating the models, took 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.

References

Alexopoulos, M. 2011. "Read All About It!! What Happens Following a Technology Shock?" *American Economic Review*. 101(4): 1144–1179.

Alexopoulos, M. and J. Cohen. 2011. "Volumes of Evidence: Examining Technical Change in the Last Century Through a New Lens". *Canadian Journal of Economics*. 44(2): 413–450.

Andrews, D. W. K. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point". *Econometrica*. 61(4): 821–856.

Ang, A., J. Boivin, S. Dong, and R. Loo-Kung. 2011. "Monetary Policy Shifts and the Term Structure". *Review of Economic Studies*. 78(2): 429–457.

Arestis, P., G. Chortareas, and J. D. Tsoukalas. 2010. "Money and Information in a New Neoclassical Synthesis Framework". *Economic Journal*. 120(542): F101–F128.

Auerbach, A. J. and Y. Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy". *American Economic Journal: Economic Policy*. 4(2): 1–27.

[27] This point is also mentioned by Duvendack *et al.* (2015). The raw data would also have pedagogical uses, as outlined by Höffler (2014).

Auerbach, A. J. and Y. Gorodnichenko. 2013. "Corrigendum: Measuring the Output Responses to Fiscal Policy". *American Economic Journal: Economic Policy*. 5(3): 320–322.

Bai, J. and P. Wang. 2011. "Conditional Markov Chain and Its Application in Economic Time Series Analysis". *Journal of Applied Econometrics*. 26(5): 715–734.

Bansak, C., M. E. Graham, and A. A. Zebedee. 2012. "Business Cycles and Gender Diversification: An Analysis of Establishment-Level Gender Dissimilarity". *American Economic Review: Papers & Proceedings*. 102(3): 561–565.

Barro, R. J. 1991. "Economic Growth in a Cross Section of Countries". *Quarterly Journal of Economics*. 106(2): 407–443.

Barro, R. J. and C. J. Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes". *Quarterly Journal of Economics*. 126(1): 51–102.

Barsky, R. B. and E. R. Sims. 2012. "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence". *American Economic Review*. 102(4): 1343–1377.

Baumeister, C. and G. Peersman. 2013. "Time-Varying Effects of Oil Supply Shocks on the US Economy". *American Economic Journal: Macroeconomics*. 5(4): 1–28.

Bernanke, B. S. 2004. "Editorial Statement". *American Economic Review*. 94(1): 404.

Bhattarai, S., J. W. Lee, and W. Y. Park. 2012. "Monetary-Fiscal Policy Interactions and Indeterminacy in Postwar US Data". *American Economic Review: Papers & Proceedings*. 102(3): 173–178.

Bianchi, F. 2012. "Evolving Monetary/Fiscal Policy Mix in the United States". *American Economic Review: Papers & Proceedings*. 102(3): 167–172.

Bianchi, F. 2013. "Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics". *Review of Economic Studies*. 80(2): 463–490.

Bilbiie, F. O. and R. Straub. 2013. "Asset Market Participation, Monetary Policy Rules, and the Great Inflation". *Review of Economics and Statistics*. 95(2): 377–392.

Broockman, D., J. Kalla, and P. Aronow. 2015. "Irregularities in LaCour (2014)". *Working Paper*.

Card, D. and A. B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania". *American Economic Review*. 84(4): 772–793.

Card, D. and A. B. Krueger. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply". *American Economic Review*. 90(5): 1397–1420.

Canova, F. and L. Gambetti. 2010. "Do Expectations Matter? The Great Moderation Revisited". *American Economic Journal: Macroeconomics*. 2(3): 183–205.

Carey, C. and S. H. Shore. 2013. "From the Peaks to the Valleys: Cross-State Evidence on Income Volatility Over the Business Cycle". *Review of Economics and Statistics*. 95(2): 549–562.

Castelnuovo, E. and P. Surico. 2010. "Monetary Policy, Inflation Expectations and the Price Puzzle". *Economic Journal*. 120(549): 1262–1283.

Chang, A. C. and P. Li. 2015a. "Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say 'Usually Not'". *Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System*: 2015–083.

Chang, A. C. and P. Li. 2015b. "Measurement Error in Macroeconomic Data and Economics Research: Data Revisions, Gross Domestic Product, and Gross Domestic Income". *Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System*: 2015–102.

Chang, A. C. and P. Li. 2017. "A Preanalysis Plan to Replicate Sixty Economics Papers That Worked Half of the Time". *American Economic Review*. 107(5): 60–64.

Chen, H., V. Curdia, and A. Ferrero. 2012. "The Macroeconomic Effects of Large-Scale Asset Purchase Programmes". *Economic Journal*. 122(564): F289–F315.

Clark, T. E. and M. W. McCracken. 2010. "Averaging Forecasts from VARs with Uncertain Instabilities". *Journal of Applied Econometrics*. 25(1): 5–29.

Clemens, M. A. 2015. "The Meaning of Failed Replications: A Review and Proposal". *Working Paper 399*. Center for Global Development.

Clements, M. P. and A. B. Galvão. 2009. "Forecasting US Output Growth Using Leading Indicators: An Appraisal Using MIDAS Models". *Journal of Applied Econometrics*. 24(7): 1187–1206.

Clements, M. P. and A. B. Galvão. 2013. "Real-Time Forecasting of Inflation and Output Growth With Autoregressive Models in the Presence of Data Revisions". *Journal of Applied Econometrics*. 28(3): 458–477.

Corsetti, G., A. Meier, and G. J. Maller. 2012. "Fiscal Stimulus with Spending Reversals". *Review of Economics and Statistics*. 94(4): 878–895.

Croushore, D. and T. Stark. 2003. "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?" *Review of Economics and Statistics*. 85(3): 605–617.

D'Agostino, A. and P. Surico. 2012. "A Century of Inflation Forecasts". *Review of Economics and Statistics*. 94(4): 1097–1106.

Del Negro, M. and F. Schorfheide. 2009. "Monetary Policy Analysis With Potentially Misspecified Models". *American Economic Review*. 99(4): 1415–1450.

Den Haan, W. J. and V. Sterk. 2011. "The Myth of Financial Innovation and the Great Moderation". *Economic Journal*. 121(553): 707–739.

Dewald, W. G., J. G. Thursby, and R. G. Anderson. 1986. "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project". *American Economic Review*. 76(4): 587–603.

Duvendack, M., R. W. Palmer-Jones, and W. R. Reed. 2015. "Replication in Economics: A Progress Report". *Econ Journal Watch*. 12(2): 164–191.

Favero, C. and F. Giavazzi. 2012. "Measuring Tax Multipliers: The Narrative Method in Fiscal VARs". *American Economic Journal: Economic Policy*. 4(2): 69–94.

Fève, P. and A. Guay. 2010. "Identification of Technology Shocks in Structural VARs". *Economic Journal*. 120(549): 1284–1318.
Fisher, J. D. M. and R. Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks". *Economic Journal*. 120(544): 414–436.
Gabaix, X. 2011. "The Granular Origins of Aggregate Fluctuations". *Econometrica*. 79(3): 733–772.
Glandon, P. 2010. "Report on the American Economic Review Data Availability Compliance Project". *Working Paper*.
Gilchrist, S. and E. Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations". *American Economic Review*. 102(4): 1692–1720.
Gordon, R. J. 2010. "Okun's Law and Productivity Innovations". *American Economic Review: Papers & Proceedings*. 100(2): 11–15.
Hall, G. J. and T. J. Sargent. 2011. "Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics". *American Economic Journal: Macroeconomics*. 3(3): 192–214.
Hamermesh, D. S. 2007. "Viewpoint: Replication in Economics". *Canadian Journal of Economics*. 40(3): 715–733.
Hansen, P. R., A. Lunde, and J. M. Nason. 2011. "The Model Confidence Set". *Econometrica*. 79(2): 453–497.
Herndon, T., M. Ash, and R. Pollin. 2014. "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff". *Cambridge Journal of Economics*. 38(2): 257–279.
Heutel, G. 2012. "How Should Environmental Policy Respond to Business Cycles? Optimal Policy Under Persistent Productivity Shocks". *Review of Economic Dynamics*. 15(2): 244–264.
Hodrick, R. J. and E. C. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation". *Journal of Money, Credit and Banking*. 29(1): 1–16.
Hoxby, C. 2000. "Does Competition Among Public Schools Benefit Students and Taxpayers". *American Economic Review*. 90(5): 1209–1238.
Hoxby, C. 2007. "Does Competition Among Public Schools Benefit Students and Taxpayers? A Reply to Rothstein". *American Economic Review*. 97(5): 2038–2055.
Höffler, J. H. 2014. "Teaching Replication in Quantitative Empirical Economics". *Working Paper for the 68th European Meeting of the Econometric Society*.
Ilbas, P. 2012. "Revealing the Preferences of the US Federal Reserve". *Journal of Applied Econometrics*. 27(3): 440–473.
Inoue, A. and B. Rossi. 2011. "Identifying the Sources of Instabilities in Macroeconomic Fluctuations". *Review of Economics and Statistics*. 93(4): 1186–1204.
Ireland, P. N. 2009. "On the Welfare Cost of Inflation and the Recent Behavior of Money Demand". *American Economic Review*. 99(3): 1040–1052.
Jermann, U. and V. Quadrini. 2012. "Macroeconomic Effects of Financial Shocks". *American Economic Review*. 102(1): 238–271.

Jore, A. S., J. Mitchell, and S. P. Vahey. 2010. "Combining Forecast Densities from VARs with Uncertain Instabilities". *Journal of Applied Econometrics*. 25(4): 621–634.

Justiniano, A., G. E. Primiceri, and A. Tambalotti. 2011. "Investment Shocks and the Relative Price of Investment". *Review of Economic Dynamics*. 14(1): 102–121.

Keen, B. D. 2009. "Output, Inflation, and Interest Rates in an Estimated Optimizing Model of Monetary Policy". *Review of Economic Dynamics*. 12: 327–343.

Kilian, L. 2009. "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market". *American Economic Review*. 99(3): 1053–1069.

Koenig, E. F., S. Dolmas, and J. Piger. 2003. "The Use and Abuse of Real-Time Data in Economic Forecasting". *Review of Economics and Statistics*. 85(3): 618–628.

Kormilitsina, A. 2011. "Oil Price Shocks and the Optimality of Monetary Policy". *Review of Economic Dynamics*. 14(1): 199–223.

Krishnamurthy, A. and A. Vissing-Jorgensen. 2012. "The Aggregate Demand for Treasury Debt". *Journal of Political Economy*. 120(2): 233–267.

Kydland, F. E. and E. C. Prescott. 1982. "Time to Build and Aggregate Fluctuations". *Econometrica*. 50(6): 1345–1370.

LaCour, M. J. and D. P. Green. 2014. "When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality". *Science*. 346(6215): 1366–1369.

Last Week Tonight, with John Oliver, Segment on Washington D.C. Voting Rights, airdate August 2. 2015.

Levine, P., J. Pearlman, G. Perendia, and B. Yang. 2012. "Endogenous Persistence in an Estimated DSGE Model Under Imperfect Information". *Economic Journal*. 122(565): 1287–1312.

Levitt, S. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime". *American Economic Review*. 87(3): 270–290.

Levitt, S. 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Reply". *American Economic Review*. 92(4): 1244–1250.

Maheu, J. M. and S. Gordon. 2008. "Learning, Forecasting, and Structural Breaks". *Journal of Applied Econometrics*. 23(5): 553–583.

Mavroeidis, S. 2010. "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence". *American Economic Review*. 100(1): 491–503.

McCrary, J. 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment". *American Economic Review*. 92(4): 1236–1243.

McCullough, B. D., K. A. McGeary, and T. D. Harrison. 2006. "Lessons From the JMCB Archive". *Journal of Money, Credit and Banking*. 38(4): 1093–1107.

McCullough, B. D., K. A. McGeary, and T. D. Harrison. 2008. "Do Economics Journal Archives Promote Replicable Research?" *Canadian Journal of Economics*. 41(4): 1406–1420.

McCullough, B. D. and H. D. Vinod. 2003. "Verifying the Solution From a Nonlinear Solver: A Case Study". *American Economic Review*. 93: 873–892.

McCullough, B. D. and H. D. Vinod. 2004. "Verifying the Solution From a Nonlinear Solver: Reply". *American Economic Review*. 94(1): 391–396.

Mertens, K. and M. O. Ravn. 2010. "Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach". *Economic Journal*. 120(544): 393–413.

Mertens, K. and M. O. Ravn. 2011. "Understanding the Aggregate Effects of Anticipated and Unanticipated Tax Policy Shocks". *Review of Economic Dynamics*. 14(1): 27–54.

Mertens, K. and M. O. Ravn. 2012. "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks". *American Economic Journal: Economic Policy*. 4(2): 145–181.

Mertens, K. and M. O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States". *American Economic Review*. 103(4): 1212–1247.

Milani, F. 2011. "Expectation Shocks and Learning As Drivers of the Business Cycle". *Economic Journal*. 121(552): 379–401.

Morley, J. and J. Piger. 2012. "The Asymmetric Business Cycle". *Review of Economics and Statistics*. 94(1): 208–221.

Mountford, A. and H. Uhlig. 2009. "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics*. 24(6): 960–992.

Nakov, A. and A. Pescatori. 2010. "Oil and the Great Moderation". *Economic Journal*. 120(543): 131–156.

Neumark, D. and W. Wascher. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment". *American Economic Review*. 90(5): 1362–1396.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science". *Science*. 349(6251): aac4716.

Perotti, R. 2012. "The Effects of Tax Shocks on Output: Not So Large, But Not Small Either". *American Economic Journal: Economic Policy*. 4(2): 214–237.

Piazzesi, M. and M. Schneider. 2010. "Interest Rate Risk in Credit Markets". *American Economic Review: Papers & Proceedings*. 100(2): 579–584.

Polito, V. and M. Wickens. 2012. "Optimal Monetary Policy Using an Unrestricted VAR". *Journal of Applied Econometrics*. 27(4): 525–553.

Ramey, V. A. 2011. "Identifying Government Spending Shocks: It's All in the Timing". *Quarterly Journal of Economics*. 126(1): 1–50.

Reinhart, C. M. and K. S. Rogoff. 2010. "Growth in a Time of Debt". *American Economic Review*. 100(2): 573–578.

Reis, R. and M. W. Watson. 2010. "Relative Goods' Prices, Pure Inflation, and The Phillips Correlation". *American Economic Journal: Macroeconomics*. 2(3): 128–157.

Romer, C. D. and D. H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks". *American Economic Review*. 100(3): 763–801.

Rothstein, J. 2007. "Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)". *American Economic Review*. 97(5): 2026–2037.

Rudebusch, G. D. and T. Wu. 2008. "A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy". *Economic Journal*. 118(530): 906–926.

Shachar, R. and B. Nalebuff. 1999. "Follow the Leader: Theory and Evidence on Political Participation". *American Economic Review*. 83(3): 525–547.

Stephanie, S.-G. and M. Uribe. 2011. "Business Cycles with a Common Trend in Neutral and Investment-Specific Productivity". *Review of Economic Dynamics*. 14(1): 122–135.

Schmitt-Groche, S. and M. Uribe. 2012. "What's News in Business Cycles". *Econometrica*. 80(6): 2733–2764.

Senyuz, Z. 2011. "Factor Analysis of Permanent and Transitory Dynamics of the US Economy and the Stock Market". *Journal of Applied Econometrics*. 26(6): 975–998.

Shore, S. H. 2010. "For Better, For Worse: Intrahousehold Risk-Sharing Over the Business Cycle". *Review of Economics and Statistics*. 92(3): 536–548.

Sims, C. A. 1980. "Macroeconomics and Reality". *Econometrica*. 48(1): 1–48.

Šustek, R. 2011. "Monetary Business Cycle Accounting". *Review of Economic Dynamics*. 14(4): 592–612.

Taylor, J. B. 1993. "Discretion Versus Policy Rules in Practice". *Carnegie-Rochester Conference Series on Public Policy*. 39: 195–214.

Thomson Reuters'. 2014. "Web of Science Journal Citation Reports: Social Science Edition". https://isiknowledge.com/.

Vlaeminck, S. and L.-K. Herrmann. 2015. "Data policies and data archives: A new paradigm for academic publishing in economic sciences?" In: *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science*. Edited by B. Schmidt and M. Dobreva. DOI:10.3233/-61499-562-3-145.

Whitlock, C. and B. Woodward. 2016. "Pentagon Buries Evidence of $125 Billion in Bureaucratic Waste". *The Washington Post*: 76–9. Accessed December 6. URL: https://www.washingtonpost.com/investigations/pentagon-buries-evidence-of-125-billion-in-bureaucratic-waste/2016/12/05/e0668c76-9af6-11e6-a0ed-ab0774c1eaa5_story.html.