

DEPARTMENT OF ECONOMICS AND FINANCE
COLLEGE OF BUSINESS AND ECONOMICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND

Replications in Economics: A Progress Report

Maren Duvendack
Richard W. Palmer-Jones
W. Robert Reed

WORKING PAPER

No. 26/2014

Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 26/2014

Replications in Economics: A Progress Report

Maren Duvendack
Richard W. Palmer-Jones
W. Robert Reed

December 3, 2014

Abstract: This study reports on various aspects of replication research in economics. It includes (i) a brief history of data sharing and replication; (ii) the results of the authors' survey administered to the editors of all 333 "Economics" journals listed in Web of Science in December 2013; (iii) an analysis of 155 replication studies that have been published in peer-reviewed economics journals from 1977-2014; (iv) a discussion of the future of replication research in economics, and (v) observations on how replications can be better integrated into research efforts to address problems associated with publication bias and other Type I error phenomena.

Keywords: Replication, data sharing, publication bias

JEL Classifications: A1, B4

Acknowledgements: The authors wish to thank Richard Anderson and Bruce McCullough for comments on an earlier draft of this paper. We also wish to thank numerous journal editors who worked with us to ensure that our reporting of their replication policies was accurate. Cathy Kang, Akmal Fazleen, Alfred Zhao, and Sonja Marzi provided excellent research assistance. Financial support from the College of Business & Law at the University of Canterbury is gratefully acknowledged.

¹ School of International Development, University of East Anglia, United Kingdom.

² Department of Economics and Finance, University of Canterbury, New Zealand

* Corresponding author: W. Robert Reed, Email: bob.reed@canterbury.ac.nz

I. INTRODUCTION

This study provides a progress report on the use of replications in economics. At least since the seminal study by Dewald et al. (1986), there has been a recognition in the economics profession that many of the empirical results in economics are not reproducible; and/or not generalizable to alternative empirical specifications, econometric procedures, extensions of the data, and other modifications to the original study. A survey of the current literature reveals that addressing this state of affairs has not been an easy task. While there have been substantial improvements in the sharing of data and code, publication in peer-reviewed journals of studies that replicate previous research is still a rare event.

The concern that a substantial portion of empirical research is not reproducible and/or generalizable is not restricted to economics and the social sciences. A recent issue of *Science* was devoted to Data Replication and Reproducibility in the so-called “hard sciences.”¹ The concern with replication in science has become sufficiently widespread that it has crossed over to popular media. The *Economist*², the *New Yorker*³, the *Atlantic*⁴, *BBC Radio*⁵, and the *Los Angeles Times*⁶ are just a few of the popular media outlets that have recently reported on widespread concern with reproducibility in scientific research. And while popular interest tends to focus on academic fraud, others have pointed out that the nature of statistical analysis as it is practiced in applied disciplines such as economics is inclined to produce a

¹ *Science*, 2 December 2011.

² <http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong>, accessed 10 September 2014.

³ <http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off>, accessed 10 September 2014.

⁴ <http://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/>, accessed 10 September 2014.

⁵ <http://www.bbc.co.uk/programmes/b04f9r4k>, accessed on 10 September 2014.

⁶ <http://articles.latimes.com/2013/oct/27/business/la-fi-hiltzik-20131027>, accessed 10 September 2014.

disproportionate rate of false positives (Maniadis et al., 2014, Ioannidis, 2005, Ioannidis and Doucouliagos, 2013, Camfield et al., 2014).

Replication by itself is not a panacea for the problems facing scientific verifiability. However, it can provide a useful check on the spread of incorrect results. Therefore, the use of replications should be of interest to many economists, even those not directly involved in the production of empirical research.

Our “progress report” proceeds as follows. Section II provides a brief history of replication and data sharing in economics journals. Section III reports the results of a recent survey of replication policies at all 333 economics journals listed in *Web of Science*. Section IV analyses a collection of 155 replication studies published in peer-reviewed economics journals. Section V summarizes our findings and concludes with (i) some closing thoughts about the future of replication in economics, (ii) a report of some recent replication initiatives, and (iii) some suggestions about how replication analysis can be more effectively employed.

II. A BRIEF HISTORY OF REPLICATIONS IN ECONOMICS

Replication and data sharing. From the early days of applied economics it has been acknowledged that sharing of data is desirable. The introductory editorial to the new journal *Econometrica* stated that the policy would be

“In statistical and other numerical work presented in *ECONOMETRICA* the original raw data will, as a rule, be published, unless their volume is excessive. This is important in order to stimulate criticism, control, and further studies. The aim will be to present this kind of paper in a condensed form. Brief, precise descriptions of (1) the theoretical setting, (2) the data, (3) the method, and (4) the results, are the essentials” (Frisch, 1933, p.3).

It is not clear to what extent these precepts were practiced, although it is unlikely that data sets were widely shared outside research groups. Restricting access to data has generally been legitimised by reference to the heavy investment of primary researchers in data production

and the long lead times from collection to publication of analyses, as well as issues of anonymity and protection of subjects. However, availability of data and code are crucial to the practice of replication in economics and the social sciences more widely. Thus, the issues raised by Frisch remain front and center up to the present time.

Concerns about the quality of data, and the questionable validity of social and economic statistical analysis were present in the post-World War II period (Morgenstern, 1951; Tullock, 1959). Tullock drew attention to the problem that significance testing that led to publishing only significant results was likely to lead to errors, because this practice was vulnerable to Type 1 errors. This is due to what is now known as the file drawer problem (Rosenthal, 1979) where negative findings are likely to be filed, while results which are statistically significant get published. Tullock (1959) also advocated replication:

“The moral of these considerations would appear to be clear. The tradition of independent repetition of experiments should be transferred from physics and chemistry to the areas where it is now a rarity” (p. 593).

The *Journal of Human Resources (JHR)* was an early fore-runner in the publication of replications. Early on, articles in *JHR* included replication as part of their analysis. For example, Smith (1968) reported that “The reader may note that the results in Tables 1 and 2 should replicate some of the results shown in Table 3.23.2 (p. 306) of the report. This is the case” (1968, p.386, fn 1). Replication, in the sense of repeating a prior analysis, was promoted; for example “these findings must be regarded as relatively weak tendencies requiring further study and badly in need of replication in independent data” (Gallaway and Dykman, 1970, p.199⁷). Later, in the same journal, authors reported that their results replicated or were consistent (or not) with the results of others (e.g. Winkler, 1975, p. 202). Others reported replicating, or at least re-estimations portions of other papers.⁸

⁷ It appears that Gallaway and Dykman (1970) see their paper as in part a replication of a report.

⁸ See Link and Ratledge (1975) on which Akin and Kneiser (1976) comment. See also Link and Ratledge (1976).

JHR continued to publish papers that contained replications through the 1970s and 1980s, and published a commitment to this in its Winter issue in 1990 (25(1)):

“The JHR Detailed Policy on Replication and Data Availability: 1. Manuscripts submitted to the JHR will be judged in part by whether they have reconciled their empirical results with already published work on the same topic. [...] 2. Authors of accepted manuscripts will be asked to [...] make the data available to others at reasonable cost from a date six months after JHR publication data and for a period three years thereafter. [...] 3. The JHR may grant a waiver of the replication policy if the data meet these criteria: 1) There is any method at all by which other researchers may obtain the data, and 2) The authors commit to providing guidance about obtaining the data. [...]”⁹.

This policy was reaffirmed and modified in 2012.

In the mid-1970s the *Journal of Political Economy* (*JPE*) responding to the assertion “that current journal editorial policy bearing on empirical literature puts an inordinate premium on the attainment of “statistically significant results,” with the effect of contaminating our published literature with a proliferation of Type 1 errors“ (Feige, 1975, p.1291), initiated a “Confirmations and Contradictions” section. This section existed from 1976 to 1999.

Confirmations could come from using new data, while contradictions would “be most powerful when based upon the same data” (cf. the *JPE* Editors commenting in Feige, 1975, p.1296). However, Mirowski and Sklivas (1991, p.159) reported that only 5 of 36 notes appearing in this section from 1976 to 1987 included replications, of which only 1 was “successful” in actually replicating the original results. Anderson et al. (2008) counted 13 more notes through 1999, of which only 1 included a replication. This led them to conclude, “Apparently the *JPE* has allowed the section to die an ignominious death befitting the section’s true relation to replication: It has been inactive since 1999” (Anderson et al., 2008, p.108).

⁹ More details are available here: http://uwpress.wisc.edu/journals/journals/jhr_replication.html, accessed 29 August 2014.

In the 1980s few major economics journals had a data sharing or replication policies in place, even though some economists recognised the need for replication (Mayer, 1980). A notable exception at the time was the *Journal of Money, Credit and Banking (JMCB)* which requested authors to make data and code available upon submission of their articles (Dewald et al., 1986). Subsequently an increasing number of journals adopted data sharing policies, either requiring authors to (i) provide data and code upon request or (ii) deposit their data and code in journal-managed data archives upon submission of their article. McCullough et al. (2006) argue that the former are ineffective because most authors and editors ignore them.

Substantial progress has been made in the last two decades with respect to the publishing of replications, in general; and in the mandatory provision of data and code in particular. Several economic journals now have official data sharing/archiving or replication policies, e.g. the *American Economic Review (AER)*, *Econometrica*, the *Journal of Applied Econometrics (JAE)*, and a number of others.

The *AER*'s policy statement, adopted in 2004 following the critical paper of McCullough and Vinod (2003), has served as a model for other journals:

“For econometric and simulation papers, the minimum requirement should include the data set(s) and programs used to run the final models, plus a description of how previous intermediate data sets and programs were employed to create the final data set(s).”¹⁰

More recently, the *AER* has tightened its policy by undertaking checks that submitted code and data do indeed produce the results published.

While the *AER*'s current policy is “an important step towards a more transparent and credible applied economic research” (Palmer-Jones and Camfield, 2013:1610), it should be noted that there is an important limitation. The *AER* only requires authors to include the data set(s) and programs necessary to run the “final models,” along with a “description of how previous intermediate data sets and programs were employed to create the final data set(s).”

¹⁰ <http://www.aeaweb.org/aer/data.php>, accessed 5 August 2014.

However, much data manipulation can take place between original and final data sets that is not carefully documented, hindering the ability of would-be replicators to obtain the final results from the raw data (Palmer-Jones and Camfield, 2013¹¹).

The mandatory submission of raw data sets, along with the programs that produce the final data sets, would enable researchers to identify how the data were “cleansed”, and help identify coding errors embodied in the final data set. These issues are little discussed in the replication literature (cf. Glandon, 2011 for an exception). Accordingly, Anderson et al. (2008) assert that much remains to be done “before empirical economics ceases to be a ‘dismal science’ when judged by the replicability of its published results” (p.99).

The replication policy adopted by *Econometrica*¹² is similar to the one by *AER* but less specific. It distinguishes between empirical analysis, experiments and simulation studies with an emphasis on experimental papers where authors are required to provide more detailed information. Like the *JMCB*, the *JAE* also has a data archive¹³, and a replication section¹⁴. *JAE* clearly specifies the format in which data sets and computer code should be made available. Making data and code available is mandatory for all papers published in *JAE*.

The requirement of making available the author’s data and code is a necessary, but by no means sufficient condition, to enable replicators to confirm the original study’s results. In some cases the policies are not strictly enforced.¹⁵ Even the *AER* has been faulted for issues of non-compliance¹⁶ Subsequently, “As a result of this project, graduate students at the University of Pittsburgh have been checking all submitted files for the last two years for

¹¹ For example Iversen and Palmer-Jones (2014) identify errors in data management which invalidate one of the two analyses in Jensen and Oster (2009) (but see Jensen and Oster, 2014).

¹² <http://www.econometricsociety.org/submissioninstructions.asp#replication>, accessed 5 August 2014.

¹³ <http://qed.econ.queensu.ca/jae/author-instructions.html>, accessed 5 August 2014.

¹⁴ <http://dmmsclick.wileyurope.com/share.asp?m=mddmkmqe668g3r5fpf6z>, accessed 5 August 2014.

¹⁵ The next section discusses a survey about replication policies that was administered to a large number of economic journals. In several cases, journal editors were surprised to discover that their requirement that data and code be provided and posted on the journal’s website was not being enforced.

¹⁶ See Glandon (2011) for an assessment of *AER*’s policy.

completeness and general compliance with the policy”, (Moffitt, 2011, p.686). In other cases, there is nominal compliance by authors – that is, they provide at least some data and or code – but the data and code are poorly documented, incomplete, or do not produce the tabulated results.

Incentives for replication. Many economics journals have adopted replication/data sharing policies over recent years but despite these positive developments replication activities have only marginally increased. To understand why this is the case we need to engage with the incentives for replication. The economics literature on replications has long recognized the important role that incentives, or lack of incentives, play. Dewald et al (1986) remarked that the incentives to undertake replication are low “however valuable in the search for knowledge”.¹⁷ This they attribute, following Kuhn¹⁸, to replication “not fit[ting] within the “puzzle solving paradigm which defines the reward structure in scientific research. Scientific and professional laurels are not awarded for replicating another scientist’s findings”; replicators are “lacking imagination” and “unable to allocate [...] time wisely”; replications may be seen “as reflecting a lack of trust in another scientist’s integrity and ability” or “personal dispute between [the replicating and replicated] researchers”.

Most discussion of incentives for replication include three actors¹⁹ – replicators, journal editors and original authors. More recently one might take account of social commentators, journalists, and the like. Replicators take account of the time to undertake the replication and the likelihood of being published. They may be concerned about the implication of lack of originality, or of gaining a reputation of having an unfavourable

¹⁷ All quotes in this paragraph are from p.587.

¹⁸ Originally published in 1970, but we refer to 1996.

¹⁹ This paragraph draws on Dewald et al (1986), Mirowski and Sklivas (1991) and Feigenbaum and Levy (1993).

personality, or being driven by cheap motives to advance themselves at the expense of more established authors.

Months of effort may yield results which cannot be conclusive about the validity of the original study in part because failure to replicate may have been due to errors in the original research or in the replication. These and similar arguments have been widely repeated to the present, and are also found in debates about replication in other disciplines. A recent example is the heated debate among social psychologists over the replication by Johnson et al. (2014) of Schnall et al. (2008)²⁰.

To further add to the disincentives to replication, many data sets are not made available for replication due to their proprietary value and or confidential nature (see the *JAE* “Instructions for Authors” for how to deal with proprietary data²¹). Furthermore, many researchers are reluctant to share data sets as they have not yet fully explored them for their own further research (Dewald et al., 1986).

In this respect it should be noted that there is a trend among official social science research funders, including the UK Economic and Social Research Council and the USA National Science Foundation, to insist that data sets that are produced with research that they have funded should be publicly available^{22,23,24}. In the non-profit aid research funding sector,

²⁰ See Mukunth (2014) for an overview, Schnall (2014) and <http://www.psychol.cam.ac.uk/cece/blog>, accessed 29 August 2014 for more details on this exchange.

²¹ “we ask authors to provide a readme file that provides a reasonable amount of information about the data. In particular, the source of the data must be described in enough detail so that other researchers can apply to obtain access to them ... It is highly desirable, as in the case of the second example, to provide the programs that were used to extract the data from the original source files” at <http://qed.econ.queensu.ca/jae/author-instructions.html>, accessed 29 August 2014.

²² “National Science Foundation is committed to the principle that the various forms of data collected with public funds belong in the public domain. Therefore, the Division of Social and Economic Sciences has formulated a policy to facilitate the process of making data that has been collected with NSF support available to other researchers.” (<http://www.nsf.gov/sbe/ses/common/archive.jsp>, accessed 6 August 2014).

²³ “The ESRC Research data policy states that research data created as a result of ESRC-funded research should be openly available to the scientific community to the maximum extent possible, through long-term preservation and high quality data management” (see <http://ukdataservice.ac.uk/manage-data/plan/dmp-esrc/esrc-data-policy>, accessed 6 August 2014).

The Gates Foundation has policies on data access²⁵. 3ie has also adopted a policy that all data produced by activities which it funds should be archived (although at the time of writing no precise protocols for this could be found²⁶).

Journal editors are concerned with the (per page) citations of replication compared to original papers. Page hits and citations are thought to be lower than for original studies. Also, the editorial costs of allowing replication may be heavy where controversy with between original and replicating authors ensues. They may also be concerned about alienation of established authors of original highly cited papers, and implications for their own reputations of publishing papers with errors.

Original authors are concerned about the costs of compiling data and code into usable forms. They may expect that the benefit of providing well documented, easily usable code are small or even negative. There may be little credit given if the replicating authors are easily able to confirm the original results, while the damage to reputation may be large if the original results cannot be confirmed. Further, replication of one paper may lead to further replications of the same paper, or to attempted replications of others (see Schnall, 2014).

To address the costs that arise from interactions between replicating and original authors, attention has focused on the establishment of protocols between replicators and replicatees to mitigate the possibilities of errors or misunderstandings in replications, and the associated potential damage to reputations, or indeed to substantive findings. However, this is unlikely to avoid harms due to malicious replicators (who can be characterised as “bullies” or

From April 2011, the “ESRC requires a data management plan for all research award applications where new data are being created. Such a plan will promote a structured approach to managing data throughout the research lifecycle, with research data ready for depositing and sharing afterwards “ (see <http://ukdataservice.ac.uk/manage-data/plan/dmp-esrc/esrc-data-policy>, accessed 6 August 2014). See also http://www.esrc.ac.uk/images/Research_Data_Policy_2010_tcm8-4595.pdf, accessed 6 August 2014.

²⁴ Gary King’s has a list of further funding agencies with data sharing and archiving policies in place: <http://gking.harvard.edu/pages/data-sharing-and-replication>, accessed 6 August 2014.

²⁵ <http://www.gatesfoundation.org/global-health/Documents/faq.pdf>, accessed 6 August 2014.

²⁶ <http://www.3ieimpact.org/en/funding/data-preparation-and-release-window/>, accessed 6 August 2014.

“data detectives”, Schnall, 2014, or belligerent replicates, such as Hoxby, 2007, or Acemoglu et al., 2012²⁷).

These considerations imply that incentives for replication are complicated by reputational issues which have become more prominent in recent years. The cases of Reinhart and Rogoff (O’Brien, 2013) and Piketty²⁸ have received much media attention, “naming and shaming” leading academics as well as putting pressure on the profession as a whole to account for their work. In one prominent case, a Japanese stem cell researcher committed suicide²⁹ as he was so deeply ashamed that two studies conducted within his laboratory and to which he put his name could not be replicated. One of his papers had to be retracted from Nature³⁰.

The growth of the internet has introduced large scale uncertainties into this process, where potentially erroneous or harmful information or views can spread quickly and extensively, while rebuttals or more considered views do not attract much attention. The skills necessary to navigate successfully in social media may be orthogonal to scientific merit.

Several authors have suggested that the “push for replication” could entail perverse effects (Bissell, 2013; Gelman, 2013³¹). The danger is that authors could become more cautious and direct their efforts away from controversial or difficult topics (Schnall, 2014).

²⁷ These characterisations are based on reading the working papers and media commentary involved in the replication-reply exchanges between these authors and their respective replicators, Rothstein (2007) and Albouy (2012).

²⁸ <http://www.ft.com/cms/s/2/e1f343ca-e281-11e3-89fd-00144feabdc0.html#axzz39aqj3dMn>, accessed 6 August 2014.

²⁹ <http://www.bbc.co.uk/news/science-environment-28658269>, accessed 29 August 2014.

³⁰ <http://www.bbc.co.uk/news/health-28124749>, accessed 29 August 2014.

³¹ <http://andrewgelman.com/2013/12/17/replication-backlash/>, accessed 29 August 2014.

Kahneman (2014) acknowledges this danger and offers suggestions to mitigate these effects through the establishment of “A New Etiquette for Replication”³².

III. CURRENT REPLICATION POLICIES AT *WEB OF SCIENCE ECONOMICS JOURNALS*

This section describes the results of a study of current replication policies of economics journals. We began by identifying all the journals that were categorized as “Economics journals” by Thomson Reuters’ *Journal Citation Reports*; a total of 333 journals.³³ Each journal was researched with respect to two questions: (i) Does the journal regularly publish data and code for its empirical research articles?³⁴, and (ii) Does the journal’s website explicitly mention that it publishes data and code.

The two issues have related, but distinct, consequences for replication studies. In order for replication to provide meaningful feedback about the reliability of previously published research, it is vital that the replication effort ensures that it is capable of reproducing the original results. Without this knowledge, one cannot be certain that the inability to reproduce another author’s findings is due to an incomplete understanding of their research data and methodology, as opposed to some inadequacy in the original study.

Without accompanying data and code, it can be very difficult to guess all the multifarious decisions that an author must make in progressing data to estimated results. Publication of data and code that allow other authors to reproduce the original study is necessary if researchers are to be confident they have correctly understood the original research. As noted above, thirty years ago, it was very difficult to obtain authors’ original

³² See also 3ie (2012) at http://www.3ieimpact.org/media/filer_public/2012/09/30/3ie_replication_contracts_notification_and_communication_policy_2.pdf, accessed 29 August 2014.

³³ Journals were identified from the online 2012 JCR Social Science Edition, retrieved September 19th, 2013, and included all journals that were categorized as “Economics” in the Subject Category Selection dialog box of the webpage.

³⁴ “Regularly” was defined as at least 50 percent of recent issues of the journal attached data and code to the online version of the article.

data and “code”. However, there has been considerable progress, led by the American Economic Association, in making this standard practice, at least at some journals. This greatly facilitates the task of replicating previous studies, lowering the cost of undertaking this research, and thereby making the benefit-cost ratio more favourable for researchers considering this type of activity. This motivates the first of our two questions.

Related to the earlier discussion on incentives, access to data and code will likely be inadequate to sufficiently incentivize replication efforts if researchers are unable to publish the results of their analyses. Whether from the narrow perspective of personal professional advancement, or the desire to have one’s research findings impact the larger community of scholars, having outlets to publish one’s research is a key determinant of whether replication research will remain anything more than a small niche activity. If all economics journals made their data and code available, but none were willing to publish replication studies, then it is unlikely that any more than a few such studies would be undertaken. While personal websites, social media, and other outlets allow replication studies some access to a larger community of scholars, the absence of professional review would make it difficult for any but the most prominent replication studies to achieve notice in the profession.

Our study was conducted as follows. The first step consisted of consulting the online website for Thomson Reuters’ *Journal Citation Reports (JCR)* and identifying those journals which *JCR* categorized under “Economics.” From there, the website of each journal was investigated. With respect to determining whether the journal “regularly published data and code for empirical research articles”, we read through recent online issues of the journal and counted up the number of empirical research articles that were published. If more than 50% of empirical articles had attached data and code, it was classified as “regularly publishes data and code”. With respect to determining whether the website explicitly mentioned that they published replications, we read through the “About,” “Aims and Scopes”, and related

sections of journal websites for some explicit mention that the journal invited submissions of replication studies or published replication studies.

After compiling our results, we then individually emailed the managing editors of all 333 journals, reporting to them what we found and asking them to correct any mistakes or omissions in our records. Editors were also invited to make any comments they wished to have associated with the results of our study. The response rate to the survey was approximately 20 percent.³⁵

TABLE 1 reports the results concerning availability of data and code for empirical articles. 24 of 333 journals regularly publish this information. To be fair, a number of journals exclusively publish theoretical content, so the absence of data and code should not be inferred as lack of support for the general policy of making this information available.

Other journals, such as the *Journal of Agricultural and Resource Economics* and the *Journal of Human Resources*, while not posting data and code through the journal's website, state that authors are required to make their data "available" for replication purposes. We did not inquire whether these journals attempted to monitor the extent to which published authors followed through on this responsibility. Nor what actions a journal might take if it was reported that an author was uncooperative in providing this information.

Even if such policies were duly monitored and enforced, there are advantages with moving the locus of responsibility of data and code availability to the journal. Foremost is that the journal can standardize the formatting of data and code so as to ease their use. Again, our survey did not go so far as to inquire whether journals had policies about the format and structure of data and code files. An unscientific sampling of files suggests that authors are largely uninstructed in this area. We also did not inquire whether journals had

³⁵ 66 journals responded to the survey including one journal whose editor wrote to inform us that the journal (*Pacific Economic Bulletin*) was no longer being published. The corresponding response rate is $66/333=0.198$.

internal processes for ensuring that the results of the published study were easily replicated with the files provided. This is another way that journals could lower the cost of replication.

As things currently stand, there is little personal incentive for published authors to ensure their data and code files can be easily understood by another researcher. The time costs of organising files and making them sufficiently transparent so as to be profitably used by others can be quite substantial. In the best case scenario, the replicating researcher will confirm the publishing author's results. In the worst case scenario, the replicating researcher will identify fatal mistakes in the data and code that undermine the original study's findings. Many researchers, weighing the benefit and costs of providing transparent data and code files, would find little personal incentive to do so. Journals can solve this public goods problem by requiring it as a condition of publication.

TABLE 2 lists the journals that explicitly mention that they invite submission of replications, or publish replications.³⁶ Again, to be fair, some journals publish replications without explicitly stating that they do so. If journals are willing to publish replications, it is important that they explicitly state this in a public place where potential authors can easily learn that fact. Many authors submit articles to journals they are not deeply familiar with. If journals are willing to publish replications, but do so infrequently, it may not be apparent to a casual observer of the journal that editors would be open to receiving these kinds of submissions. This will discourage submission of a replication study to that journal, especially if a submission keeps a paper out of circulation for three, six, or more months before it ends up being rejected. Further, by leaving a potential replicating researcher unaware of the possibility of publishing in that journal, it narrows the pool of potential outlets in which a researcher thinks he/she can publish their work. Therefore, a publicly stated policy

³⁶ In at least two cases, journal editors modified their journal websites after we told them that our classification system required explicit mention of this policy on the journal website.

on the journal website identifying the journal's willingness to consider replication studies is very important if authors are to perceive a benefit for undertaking this research.

Of the 333 "economics" journals listed in the *Journal of Citation Reports*, only 7 explicitly state that they publish replication studies. Further, some of these are specialty journals that only publish studies in a particularly area, such as the journal *Experimental Economics*; whereas others, such as the *Journal of Applied Econometrics*, only publish replications where the original article was published in one of a few elite journals.³⁷ Thus, as a practical matter, there may only be one or two journals that will publish a replicating author's research. The lack of publishing outlets is perhaps the most serious obstacle to researchers interested in undertaking replication research.

IV. AN ANALYSIS OF PUBLISHED REPLICATIONS

This section analyses a collection of published replication studies found in refereed economics journals. To be categorized as a "replication study," the article had to (i) have been published in a peer-reviewed journal and (ii) have as its main purpose the replication of another research article.³⁸ The replication studies were identified from a number of sources: (i) keyword searches in Google Scholar and Web of Science; (ii) the "Replication in Economics!" wiki³⁹; and (iii) authors' own collections. Subsequent to that, we also did a more systematic search that targeted the top 50 economics journals based on impact factors.⁴⁰

³⁷ The journals are: *Econometrica*, *American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economics and Statistics*, *Review of Economic Studies*, *Journal of Econometrics*, *Journal of Business and Economic Statistics*, and the *Economic Journal*.

³⁸ We did not include articles that had been published online as "early access". One of the characteristics we wanted to record was whether the journal published a "reply/response" to the replication study. It was not possible to determine this from "early access" articles. We also did not include "replies" or "responses" to replication studies, or replies or responses to replies/responses. We judged that the motivation underlying these was likely to be different, as these were colored by the incentive to defend the author's earlier research.

³⁹ http://replication.uni-goettingen.de/wiki/index.php/Main_Page

⁴⁰ The impact factors were taken from here: https://ideas.repec.org/top/top_journals_simple.html. We actually referenced 51 journals, since two journals had identical impact factors. The journals are: *American Economic Journal: Microeconomics*, *American Economic Journal: Macroeconomics*, *American Economic Review*, *Econometric Reviews*, *Econometrica*, *Econometrics Journal*, *Economic Journal*, *European Economic Review*,

Each of these journals was searched using the terms (i) “replicat*” and (ii) “Replicate*”. This generated 13,261 potentially relevant papers. Not having the time or financial resources to screen all of these, we randomly sampled approximately 10% of them (1601 articles), reviewing the full text to determine if the article satisfied our criteria to be classified as a “replication study.” Of these 1601 studies, most did not actually undertake a formal replication exercise; or the replication was not the main focus of the paper; or the paper styled itself as an empirical or conceptual extension of the original paper without attempting to confirm or disconfirm the original study. In the end, our search produced 155 replication studies.

FIGURE 1 presents a plot of replication studies by year. The first article that we can identify whose main focus was to replicate a previous study dates to 1977. It is a replication of a minimum wage study published in *Economic Inquiry* (Siskind, 1977). Over the next fourteen years (through 1991), fifteen more replication studies were published, the great majority of which (eleven) were published in the *Journal of Human Resources*. Other journals publishing replication studies were *Applied Economics* (1983, 1985), the *Quarterly Journal of Economics* (1984), and the *Journal of Applied Econometrics* (1990).

Beginning in 1992, a number of other journals started to publish replication studies: *Review of Economics and Statistics* (1992), *Journal of Development Studies* (1993), *Marketing Letters* (1994), *Labour Economics* (1995), *Empirical Economics* (1997), *Public*

Experimental Economics, Games and Economic Behavior, International Economic Review, International Journal of Central Banking, Journal of Accounting and Economics, Journal of Applied Econometrics, Journal of Development Economics, Journal of Econometrics, Journal of Economic Dynamics and Control, Journal of Economic Growth, Journal of Economic Perspectives, Journal of Economic Surveys, Journal of Economic Theory, Journal of Empirical Finance, Journal of Environmental Economics and Management, Journal of Finance, Journal of Financial and Quantitative Analysis, Journal of Financial Economics, Journal of Financial Intermediation, Journal of Financial Markets, Journal of Health Economics, Journal of Human Resources, Journal of International Business Studies, Journal of International Economics, Journal of International Money and Finance, Journal of Labor Economics, Journal of Law and Economics, Journal of Monetary Economics, Journal of Political Economy, Journal of Population Economics, Journal of Public Economics, Journal of Risk and Uncertainty, Journal of the European Economic Association, Journal of Urban Economics, Labour Economics, Mathematical Finance, Oxford Bulletin of Economics and Statistics, Quarterly Journal of Economics, RAND Journal of Economics, Review of Economic Dynamics, Review of Economic Studies, Review of Financial Studies, and World Bank Economic Review.

Choice (1998), *Journal of Political Economy* (1998), *Experimental Economics* (2000), *Journal of International Development* (2000), *Quarterly Journal of Business and Economics* (2000), *Journal of Development Economics* (2001), and the *Journal of Law and Economics* (2001). Interestingly, some of these journals never published another replication study. The *American Economic Review* published its first replication study in 2002 (McCrary, 2002); though its earlier publication of the landmark study by Dewald, Thursby, and Anderson (1986) did much to illuminate the need for replications in the discipline.

A major development in the publication of replications occurred in January 2003 when the *Journal of Applied Econometrics (JAE)* began a replication section under the editorship of Badi Baltagi. From that time on, the *JAE* has become the most prolific publisher of replication studies amongst economics journals. Another notable journal event was the first-time publication of replication studies by *Econ Journal Watch* in 2004. As FIGURE 1 makes clear, since the early 2000s, journals have published replication studies with increasing frequency.

TABLE 3 provides a detailed listing of the journals that publish replications studies. As has already been stated, the *JAE* is the most frequent publisher of replication studies. It accounts for about one-fifth of all replication studies published in peer-reviewed economics journals. The next most frequent publishers are the *Journal of Human Resources*, *American Economic Review*, *Econ Journal Watch*, *Experimental Economics*, and the *Journal of Development Studies*. These six journals account for almost 60% of all replication studies. Only ten economics journals have ever published more than 3 replication studies.

The remainder of this section identifies some general characteristics of the published replication studies. The studies were coded on six dimensions:

1. **Summary?** Was the published article a full study, or did it only summarize the key results from a study?
2. **Exact?** Did the replication study attempt to exactly reproduce the original findings?

3. ***Extension?*** Did the replication study go beyond attempting to reproduce the original results by extending the analysis to different types of subjects, time periods, or test additional hypotheses?
4. ***Original Results?*** Did the replication study report the findings of the original study in a way that facilitated comparison of results without having to access the original study?
5. ***Negative? Mixed? Positive?*** Did the replication study confirm or disconfirm the original study, or were the results mixed?
6. ***Reply?*** Did the journal publish a reply or response from the original authors?

Each of the categories are described in more detail in TABLE 4.

TABLE 5 reports the results. The numbers in the table are averages of the 0-1 values of the respective dummy variables. As these numbers report population rather than sample values, hypothesis testing is not applicable. The first category (*Summary?*) is largely driven by a practice of the *JAE* to sometimes publish paragraph-length summaries of replication studies. An example is Drukker and Guan (2003):

We are able to reproduce the results in Tables I and II of Baltagi and Khanti-Akom (1990) using STATA programs. With respect to Table III, we obtain a different estimate of $\hat{\sigma}_\alpha^2$ than Baltagi and Khanti-Akom. This changed the estimates slightly. The programs and results are available from ddrukker@stata.com on request.

That short paragraph comprises the entirety of the published “replication study.” For this reason, the subsequent analysis separates out *JAE* studies from other journals’ replication studies, as roughly a fifth of all *JAE* studies consist of short summaries (though usually not this short). We also separate out experimental replication studies, because they are inherently not exactly reproducible, using different subjects, and often subjects from different countries. This raises interpretation issues associated with reproducibility. Accordingly, we report the characteristics of replication studies for four categories of journals: (i) Studies from all journals (138), (ii) *JAE* studies (30), (iii) experimental studies (11), and (iv) non-*JAE*/non-experimental studies (97).

With respect to the *Exact?* category, TABLE 5 reports that a little less than two-thirds of all published replication studies attempt to exactly reproduce the original findings. The number is slightly higher for the *JAE*. A major reason for not attempting to exactly reproduce an original study's findings consists of a replication attempting to confirm an original study using a different data base. An example is a replication study by Iversen and Palmer-Jones (2008) that tested a result obtained by Basu et al. (2002) using more recent data and data from a different country.

The next category, *Extension?*, is designed to determine whether it is sufficient for replication studies to merely reproduce an original study's findings, or must they have some independent novelty or innovation (different data, additional hypotheses). On this dimension, there is wide variation across journal categories. Studies published in the *JAE* are more likely to consist entirely of attempts to confirm the original study's findings, without performing analyses that go beyond the original study. Only about a fourth of *JAE* replication studies perform extensions of the original study. In contrast, over 80 percent of experimental studies go beyond the original study's analysis, often to explore additional hypotheses. Unfortunately, our analysis is unable to distinguish between "demand" and "supply" factors. That is, we cannot tell if the difference in replication studies between, say, the *JAE* and experimental studies, is driven by the preferences of journal editors or the preferences of replicating authors.

The next category (*Original Results?*) is helpful in knowing whether journals require replicating studies to report original results in a way that facilitates comparison with the original study. A large portion of replication studies do not do this. This may be due to limitations associated with scarce journal space. This is sometimes more than a minor inconvenience, as a replication study may refer qualitatively to results from an original study,

but without precisely identifying the table or regression number from which this result(s) comes.

The next three categories report the ability of replicating studies to confirm findings from the original study. Across all categories of journals/studies, approximately two-thirds of replicating studies disconfirm major findings from the original study. Interpretation of this number is difficult. On the one hand, one might take this as an upper bound on the unreliability of previous research, because researchers who confirm the results of original studies may anticipate difficulty in getting their results published since they have nothing “new” to report.

On the other hand, the upper bound may be higher than indicated by the numbers in TABLE 5. This would be the case if journal editors are loathe to offend influential researchers, or the editors at journals that published the original study. The *Journal of Economic & Social Measurement* and *Econ Journal Watch* apparently encourage/allow replicating authors to detail their difficulties in getting their negative results published. These accounts detail first-hand the reticence of some journal editors to publish disconfirming replication studies (Davis, 2007a; Jong-A-Pin & De Haan, 2008).

The last category, *Reply?*, indicates how frequently journals publish a response by the original authors to the replication study. It is clear that – given how we have defined this category -- replies are generally infrequent.⁴¹ Approximately one in five replication studies are responded to by the original authors in the same issue. Not surprisingly, replies are most likely to occur when the replicating study disconfirms the original study. Of the 31 replication studies that elicited a published response from the original authors, all but one were in response to the replicating study disconfirming the original results (Muños, 2012; Findlay and Santos, 2012).

⁴¹ We did not include replies that were published in later issues of the journal. This would bias finding replies/responses to older replication studies, since there was more time for the reply/response to appear.

What can we learn from our analysis of replication studies? Most importantly, and perhaps not too surprisingly, the main takeaway is that, conditional on getting published, there is a high rate of disconfirmation. Over the full set of replication studies, approximately two out of every three studies disconfirmed the original findings. Another 12 percent disconfirmed at least one major finding of the original study, while confirming others (*Mixed?*). In other words, approximately 80 percent of the replication studies found major flaws in the original research.

Could this be an overestimate of the true rate of Type II errors in original studies? While the question is impossible to answer conclusively with our sample, there is some indication that this rate overstates the unreliability of original studies. The *JAE* is noteworthy in that it publishes many replications that consist of little more than the statement “we were able to reproduce the original study’s findings” (see Drukker and Guan, 2003; quoted above). This suggests that the *JAE* does not discriminate on the basis of whether the replication study confirms or disconfirms the original study. This contrasts with the *American Economic Review*, which has never published a replication that merely confirmed the original study. If we take the *JAE*’s findings as representative, we see that the rate of (*Negative?* + *Mixed?*) falls to 67 percent (0.467+0.200). By any account, this is still a large number. It raises serious concerns about the reliability of published empirical research in economics. We shall have more to say about the results of TABLE 5 in the conclusion.

V. SUMMARY AND CLOSING THOUGHTS

Summary. The importance of making data available to researchers in order to enable replication has long been noted, going back at least to 1933 and Ragnar Frisch in the first issue of *Econometrica*. Over the years, the importance of replication has been oft-repeated, accompanied by concern about the lack of replicability of reported economic findings (Dewald et al., 1986). Even so, it has taken a long time for replication studies to gain a

foothold in peer-reviewed journals. The first study that we identify whose main focus was to replicate a previous study was published in 1977. As FIGURE 1 illustrates, the publication of replication studies has been increasing slowly since then, albeit at an increasing rate.

Even so, there are relatively few journals that publish replication studies. Six journals account for approximately 60% of all published replication studies in economics. Only ten journals have ever published more than 3 replication studies (cf. TABLE 3).

There are noteworthy differences in the type of replication studies that different journals publish (cf. TABLE 5). The *Journal of Applied Econometrics* often publishes short summaries of replication research in which the authors note whether they were able to successfully replicate previous research. Replications of experimental studies often include extensions to the original research that go beyond a strict confirmation/disconfirmation of the original study. Some journals, like the *Journal of Human Resources*, have routinely published responses by the original authors when replications fail to confirm the original findings. Others, such as *Experimental Economics*, do so only rarely.

Overall, we find that it is quite common to find that major results from empirical research in economics journals cannot be confirmed. Over three-fourths of replication studies report failing to confirm one or more major findings from the original research. It is difficult to identify whether this is because (i) most published economic research is unreliable, or because (ii) journals choose to disproportionately publish negative studies.

One possible identification strategy is to focus on a journal whose editorial policy is to publish replication studies without regard to whether they confirm or disconfirm the original study. The *Journal of Applied Econometrics* seemingly comes close to this ideal. We find that approximately two-thirds of the replication studies published by the *JAE* fail to confirm one or more major findings from the studies they attempt to replicate. While this is less than the rate for all journals, it still suggests that a large portion of empirical research in

economics journals is unreliable. The task of identifying which results are reliable, and which are not, should be an important priority for the economics discipline.

The future of replications. The fields of science and political science have been very active in calling for an increase in replication activities. For example, the Center for Open Science received 1.3 Million USD to start the Reproducibility Initiative^{42, 43}, which aims to independently verify the results of major scientific experiments. There have also been renewed calls for replication in the political sciences, e.g. Gary King's website⁴⁴ is a good resource, the political science replication blog⁴⁵ is another. More recently the Berkeley Initiative for Transparency in the Social Sciences (BITSS)⁴⁶ was started with the objective to make empirical social science research more transparent which includes promoting replications.

The area of economics has seen some but relatively few replication initiatives, one is the "Replication in Economics" project at Goettingen University which is funded by the Institute for New Economic Thinking and which has compiled a wiki⁴⁷ containing an extensive number of replication studies published in economic journals. Another replication initiative in the field of development economics has been launched by 3ie⁴⁸.

Will replication research expand from its current foothold in the journals? We can examine this from the perspective of supply and demand for replication research. On the supply side are producers/researchers. They weigh the costs and benefits of producing replication studies relative to other types of research outputs. The increasing availability of

⁴² <http://validation.scienceexchange.com/#/reproducibility-initiative>, accessed 29 August 2014.

⁴³ <https://osf.io/ezcuj/wiki/home/>, accessed 29 August 2014.

⁴⁴ <http://gking.harvard.edu/pages/data-sharing-and-replication>, accessed 29 August 2014.

⁴⁵ <http://politicalsciencereplication.wordpress.com/>, accessed 29 August 2014.

⁴⁶ <http://bitss.org/>, accessed 29 August 2014.

⁴⁷ http://replication.uni-goettingen.de/wiki/index.php/Main_Page, accessed 29 August 2014.

⁴⁸ <http://www.3ieimpact.org/evaluation/impact-evaluation-replication-programme/>, accessed 29 August 2014.

data and code reduces the cost of undertaking replication research. This is one possible explanation for the observed increase in the number of published replication studies over time (cf. FIGURE 1). Further availability of data and code should result in more resources being devoted to replication research.

Moderating this are the expected benefits to the researcher of producing this type of research. This is a function of the professional rewards associated with publishing replication research, which is related to the probability of publication in a respected peer-reviewed journal. Liebowitz reports that quality of journal in which an author's work appears is the single most important criterion for promotion.⁴⁹ If this is the case, then unless there is an increase in the frequency with which top journals publish replication studies, it will be difficult for a published replication study to produce the same benefit to a researcher of publishing "original research."⁵⁰ However, given that very few journals currently publish replication research, even a small numerical increase in their number could have a significant impact on expected benefits by increasing the probability that a replication study will get published.

On the demand side of the market are the journals. The recent increase in the number of top journals requiring authors to make available their data and code generates dynamic externalities that should make it easier for lower-ranked journals to follow suit: Authors have a variety of outlets to which they can submit their research. Requiring an author to make available their data and code reduces the attractiveness of submitting to that journal, which serves as a disincentive for journals to impose this restriction in the first place. As more journals adopt this requirement, this lowers the cost for other journals to follow suit. Therefore, we expect the trend towards requiring authors to make available their data and

⁴⁹ <http://www.voxeu.org/article/our-uneconomic-methods-measuring-economic-research> accessed 11 September 2014.

⁵⁰ Balanced against this is a recent study by Gibson et al. (2014) which finds that membership in the club of "top journals" may be wider than is commonly asserted.

code to continue. As noted above, this should stimulate more replication studies, *ceteris paribus*.

An important determinant of journal demand for research articles is the extent to which research in that journal is likely to get cited. Evidence of the power of citations is the rising influence of “impact factors” in ranking journals (Wilhite and Fong, 2012). Thus, a key factor in journal demand is the extent to which replication studies are cited by other researchers.

We expect that elite journals will likely continue to find little benefit to publishing replication studies, as they receive high quality, original research with much citation potential. However, journals of lesser quality may find that replications of widely-cited papers can be expected to produce more citations than original research submitted to those journals. If that is the case, the pursuit of citations may help replication studies to establish a niche within the hierarchy of economics journals.

Technological innovation also affects journal demand. The *Journal of Applied Econometrics*’ practice of publishing summaries of replications allows it to allocate less journal space for a replication study relative to an original research study. The increasing sophistication of online publishing also creates opportunities for journals to use their scarce journal space more efficiently. *Public Finance Review* publishes a summary version of a replication study in its print edition, but attaches the full-length manuscript as “Supplemental material” that can be accessed at the journal’s online website. These innovations increase the ratio of citations/journal page, and hence can shift the demand for replication studies relative to original studies at some journals.

Finally, widespread attention directed towards the replicability of scientific research may affect journal editors’ and researchers’ “tastes” for replication studies. This also

generates dynamic externalities that simultaneously increases the demand and supply of replication studies.

In summary, recent developments in the market for replication studies suggests that the key forces that have driven the expansion of replication research in recent decades – the increasing availability of data and code, technological innovations in the allocation of journal space, and societal factors that affects “tastes” for replication research – are likely to expand the use of replications in the future.

Replication and publication bias. Implicit in the previous discussion (in relation to the file drawer problem mentioned above), but not explicitly addressed, is the potential for replication research to mitigate the problem of publication bias. Publication bias is the phenomenon that empirical results that are large and statistically significant are more likely to be published. In a recent study, Franco et al. (2014) report that “Strong results are 40 percentage points more likely to be published than null results, and 60 percentage points more likely to be written up.” They identify the locus of publication bias residing, not with the journals, but with researchers who choose not to write up and submit empirical findings that are insignificant. Evidence of publication bias in economics has been reported by Card and Krueger (1995), Ashenfelter et al. (1999), and Doucouliagos (2005), among others.

Closely related to the phenomenon of publication bias is “HARKing”, “Hypothesizing After the Results are Known” (Kerr, 1998). This is effectively data-mining, where researchers stumble upon significant results in their regression runs and then work backwards deductively to identify hypotheses consistent with those results.

Replication holds promise to address the problems caused by publication bias and HARKing. If published results reflect Type I errors, replication research can uncover this by, among other things, modifying model specifications and sample periods. Spurious results will have difficulty being sustained when different variable combinations and unusual

observations are investigated. In principle, spurious results of this sort could be caught in the refereeing process. In practice, they often are obscured because reviewers do not have access to the original authors' data in the review process. Further, knowledge that an article's data and code will be made available at publication may cause researchers to take additional precautionary steps to ensure that their results are robust, lest their research be caught out in subsequent replication research.

Using replications more effectively. The discussion of publication bias leads to mention of another tool for assessing the reliability and validity of economic research. Meta-analysis/meta-regression is a procedure for aggregating estimated effects across many studies. It has long been used in medical, education, and psychology research. Over the last decade, it has become increasingly employed in economics (Ringquist, 2013; Stanley and Doucouliagos, 2012).

To date, replication and meta-analysis have largely lived parallel lives. One future development we would like to see is increased use of these procedures in tandem. Meta-regression can be used to identify study characteristics that "explain" why different studies reach different conclusions. Replication studies can then take the results of meta-analyses and investigate whether changing the empirical design of a study has the effect predicted by meta-analysis. Conversely, replication studies may identify study characteristics that meta-analyses can incorporate in subsequent meta-regression research.

While replication is no panacea, it is a useful and, in our opinion, underutilized tool for assessing the reliability and validity of empirical results. It may be too much to expect replications to establish an entirely solid foundation upon which further research can build, however, they can help to substantially firm up the existing base. Given the current state of economic research, that would constitute a valuable contribution. It is our hope that this progress report on replications in economics will further this development.

REFERENCES

- Acemoglu, D., Johnson, S., Robinson, J.A. (2012). Hither Thou Shalt Come, But No Further: Reply to “The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. *American Economic Review*, 102, pp.3007–3110.
- Akin, J.S. and Kniesner, T.J. (1976). Proxies for Observations on Individuals Sampled from a Population. *The Journal of Human Resources*, 11, pp.411–413.
- Albouy, D.Y. (2012). The Colonial Origins of Comparative Development: an Empirical Investigation: Comment. *American Economic Review*, 102, pp.3059–3076.
- Anderson, R. G., Greene, W. H., McCullough, B D, and Vinod, H D. (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology*, 15(1), pp.99-119.
- Ashenfelter, O., Harmon, C., and Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics* 6, pp. 453–470.
- Basu, K., Narayan, A. and Ravallion, M. (2002) Is literacy shared within households? Theory and evidence from Bangladesh, *Labour Economics*, 8(6), pp. 649–665.
- Bernanke, B.S. (2004). Editorial Statement. *American Economic Review*, 94(3), p.404.
- Bissell, M. (2013). Reproducibility: The risks of the replication drive. *Nature*, 503, pp.333–334.
- Burman, L.E., Reed, W.R. and Alm, J. (2010). A call for replication studies. *Public Finance Review*, 38(6), pp.787-793.
- Cable, V. (2003). The Political Context – Does Evidence Matter? ODI Briefing Note. Available at: <http://www.odi.org/sites/odi.org.uk/files/odi-assets/events-documents/2609.pdf>.
- Camfield, L., Duvendack, M. and Palmer-Jones, R.W. (2014). All you wanted to know about bias in Impact Evaluation but never dared to ask. *IDS Bulletin forthcoming*.
- Card, D., and Krueger, A.B. (1995). Time-series minimum wage studies: a meta-analysis. *American Economic Review* 85, pp. 238–243.
- Collins, H. M. (1985). *Changing Order: Replication and Induction to Scientific Practice*. Chicago: University of Chicago Press.
- Davis, Graham A. (2007). Reflections of a first-time replicator. *Journal of Economic and Social Measurement*, 32(4), 263–266.
- Dawson, A. (1985). Comment upon a new-classical model of the postwar UK. *Applied Economics*, 17(2), 257-262.

- Dawson, A. (1983). The performance of three wage equations in postwar Britain. *Applied Economics*, 15(1), 91-105.
- Dewald, W. G., Thursby, J. G., and Anderson, R. G. (1986). Replication in Empirical Economics: the Journal of Money, Credit and Banking Project. *American Economic Review*, 76, pp.587-603.
- Doucouliaagos, C. (2005) Publication bias in the economic freedom and economic growth literature. *Journal of Economic Surveys* 19, pp. 367–387.
- Drukker, D. M., & Guan, W. (2003). Replicating the results in ‘on efficient estimation with panel data: an empirical comparison of instrumental variables estimators’. *Journal of Applied Econometrics*, 18(1), 119-119.
- Duvendack, M. and Palmer-Jones, R. (2013). Replication of quantitative work in development studies: Experience and suggestions. *Progress in Development Studies*, 13(4), pp.307-322.
- ESRC (2014). ESRC research funding guide. Available at: http://www.esrc.ac.uk/images/Research-Funding-Guide_tcm8-2323.pdf, accessed 29 August 2014.
- Feige, E. L. (1975). The Consequences of Journal Editorial Policies and a Suggestion for Revision. *Journal of Political Economy*, 83(6), pp.1291-1296.
- Feigenbaum, S. and Levy, D.M. (1993). The market for (ir)reproducible econometrics. *Social Epistemology*, 7(3), pp.215–232.
- Findlay, D. W., & Santos, J. M. (2012). Race, ethnicity, and baseball card prices: A replication, correction, and extension of Hewitt, Muñoz, Oliver, and Regoli. *Econ Journal Watch*, 9(2), 122-140.
- Franco, A., Malhotra, N., and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science* 19(September), pp. 1502-1505.
- Frisch, R. (1933). Editor’s Note. *Econometrica*, 1(1), pp.1-4 CR.
- Gallaway, L.E. and Dyckman, Z. (1970). The Full Employment-Unemployment Rate: 1953-1980. *The Journal of Human Resources*, 5, pp.487–510.
- Gibson J., Anderson, D., and Tressler, J. (2014), “Which Journal Rankings Best Explain Academic Salaries? Evidence from the University of California”, *Economic Inquiry*, 52(4), pp. 1322-1340.
- Glandon, P. (2011). Report on the American economic review data availability compliance project. *American Economic Review*, 101(3), pp.695–699.
- Gordon, P., & Wang, L. (2004). Does economic performance correlate with big government. *Econ Journal Watch*, 1(2), 192-221.

- Hallsworth, M. and Rutter, J. (2011). Making Policy Better: Improving Whitehall's Core Business. Institute for Government. Available at: <http://www.instituteforgovernment.org.uk/sites/default/files/publications/Making%20Policy%20Better.pdf>.
- Hoxby, C.M. (2007). Does Competitions am Public Schools Benefit Students and Taxpayers? A Reply. *American Economic Review*, 97, pp.2038–2055.
- Ioannidis, J.P.A. (2005). Why Most Published Research Findings Are False. *PLoS Med*, 2, p.e124.
- Ioannidis, J. and Doucouliagos, C. (2013). What's to Know About the Credibility of Empirical Economics? Scientific Credibility of Economics. *Journal of Economic Surveys*, 27(5), pp.997-1004.
- Iversen, V., Palmer-Jones, R.W. (2008). Literacy sharing, assortative mating, or what? Labour market advantages and proximate illiteracy revisited. *Journal of Development Studies* 44(6), 797–838.
- Iversen, V. and Palmer-Jones, R.W. (2014). TV, female empowerment and fertility decline demographic transition in rural India. *3ie replication Paper No 2, New Delhi*. Available at: <http://www.3ieimpact.org/publications/3ie-replication-paper-series/3ie-replication-paper-2/>, accessed 29 August 2014.
- Jasny, B.R. Chin, G., Chong, L. and Vigneri, S. (2011). Data replication & reproducibility. Again, and again, and again Introduction. *Science*, 334(6060), p.1225.
- Jensen, R. and Oster, E. (2009). The Power of TV: Cable Television and Women's Status in Rural India, *Quarterly Journal of Economics*, 124(3), pp.1057-94.
- Jensen, R. and Oster, E. (2014). TV, Female Empowerment and Fertility Decline in Rural India: Response to Iversen and Palmer-Jones. *3ie, New Delhi*. Available at: <http://www.3ieimpact.org/publications/3ie-replication-paper-series/3ie-replication-paper-2/>, accessed 29 August 2014.
- Johnson, D.J., Cheung, F. and Donnellan, M.B. (2014). Does Cleanliness Influence Moral Judgments? A Direct Replication of. *Social Psychology*, 45, pp.209–215.
- Jong-A-Pin, R., & De Haan, J. (2008). Growth accelerations and regime changes: A correction. *Econ Journal Watch*, 5(1), 51-58.
- Kahneman, D. (2014). A New Etiquette for Replication. Available at: <http://www.scribd.com/doc/225285909/Kahneman-Commentary>, accessed 29 August 2014.
- Kerr, N.L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review* 2(3), pp. 196–217.
- Kuhn, T.S. (1996). *The Structure of Scientific Revolutions*. New ed of 3rd revised ed. ed. Chicago: University of Chicago Press.

- Link, C.R. and Ratledge, E.C. (1975). Social Returns to Quantity and Quality of Education: A Further Statement. *The Journal of Human Resources*, 10, pp.78–89.
- Link, C.R. and Ratledge, E.C. (1976). Proxies for Observations on Individuals Sampled from a Population: Reply. *The Journal of Human Resources*, 11, pp.413–419.
- Lovell, M.C. and Selover, D.D. (1994). Econometric Software Accidents. *The Economic Journal*, 104, pp.713–725.
- Maniadis, Z., Tufano, F. and List, J.A. (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American Economic Review*, 104, pp.277–290.
- Mayer, T. (1980). Economics as a Hard Science: Realistic Goal or Hard Science. *Economic Inquiry*, 18, p.165-178.
- McCrary, J. (2002). Using electoral cycles in police hiring to estimate the effect of police on crime: Comment. *The American Economic Review* 92(4), 1236–1243.
- McCullough, B.D. (2009). Open Access Economics Journals and the Market for Reproducible Economic Research. *Economic Analysis and Policy*, 39, pp.117–126.
- McCullough, B. D., and Vinod, H. D. (2003). Verifying the Solution from a Nonlinear Solver: A Case Study. *The American Economic Review*, 93(3), pp.873-892.
- McCullough, B., McGeary, K. A. and Harrison, T D. (2006). Lessons from the JMCB Archive. *Journal of Money, Credit and Banking*, 38(4), pp.1093-1107.
- Mirowski, P. and Sklivas, S. (1991). Why econometricians don't replicate (although they do reproduce). *Review of Political Economy*, 3(2), pp.146–163.
- Moffitt, R.A. (Ed.) (2011). Report of the Editor: American Economic Review. *American Economic Review*, 101(3), pp.684–699.
- Morgenstern, O. (1951). On the accuracy of economic observations, In: *Activity Analysis of Production and Allocation: Proceedings of a Conference*. Wiley: New York.
- Mukunth, A.V. (2014). Replication studies, ceiling effects, and the psychology of science. *Sciblogger*. Available at: <http://sciblogger.com/2014/05/26/replication-studies-ceiling-effects-and-the-psychology-of-science/> accessed 28/8/2014.
- Muñoz, R. Jr. (2012). Beyond race cards in america's pastime: An appreciative reply to Findlay and Santos (2012). *Econ Journal Watch*, 9(2), 141-148.
- O'Brien, M. (2013). Forget Excel: This Was Reinhart and Rogoff's Biggest Mistake: Correlation is not Causation. *The Atlantic*, 18 April 2013. Available at: <http://www.theatlantic.com/business/archive/2013/04/forget-excel-this-was-reinhart-and-rogooffs-biggest-mistake/275088/>, accessed 29 August 2014.

- Ortmann, A., Fitzgerald, J., & Boeing, C. (2000). Trust, reciprocity, and social history: A re-examination. *Experimental Economics*, 3(1), 81-100.
- Palmer-Jones, R. and Camfield, L. (2013). Three 'Rs' of Econometrics: Repetition, Reproduction and Replication. *Journals of Development Studies*, 49(12), pp.1607-1614.
- Peng, R.D. (2009). Reproducible research and Biostatistics. *Biostatistics*, 10(3), pp.405–408.
- Pesaran, H. (2003). Introducing a Replication Section. *Journal of Applied Econometrics*, 18(1), pp.111-111.
- Pritchett, L. (2002). It pays to be ignorant: A simple political economy of rigorous program evaluation. *The Journal of Policy Reform*, 5(4), pp.251-269.
- Ringquist, E. (2013). *Meta-analysis for public management and policy*. San Francisco, CA: Jossey-Bass.
- Robbins, L. (1932, 1935). *An Essay on the Nature and Significance of Economic Science*, 2nd ed. London: Macmillan.
- Rothstein, J. (2007). Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000). *American Economic Review*, 97, pp.2026–2037.
- Rosenthal, R. (1979). The “file drawer” problem and tolerance for null results. *Psychological Bulletin*, 86, pp.638–641.
- Sandve, G.K., Nekrutenko, A., Taylor, J. and Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. *PLoS Comput Biol* 9, p.e1003285.
- Schnall, S., Benton, J. and Harvey, S. (2008). With a clean conscience cleanliness reduces the severity of moral judgments. *Psychological Science*, 19, pp.1219–1222.
- Schnall, S. (2014). Clean data: Statistical artefacts wash out replication efforts. Commentary and Rejoinder on Johnson, Cheung, and Donnellan (2014). *Social Psychology*, Advance online publication. doi: 10.1027/1864-9335/a000204.
- Siskind, F.B. (1977). Minimum wage legislation in the United States: Comment. *Economic Inquiry* 15(1), 135–138.
- Smith, M.S. (1968). Equality of Educational Opportunity: Comments on Bowles and Levin. *The Journal of Human Resources*, 3, pp.384–389.
- Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. New York, NY: Routledge.
- The Editors, *Journal of Political Economy* (1975). Editorial Comment. *Journal of Political Economy*, 83, pp.1295–1296.

Tullock, G. (1959). Publication Decisions and Tests of Significance - A Comment. *Journal of the American Statistical Association* 54(287), pp.593-593.

Wilhite, A. and Fong, E. (2012). Coercive Citation in Academic Publishing. *Science* 335, pp. 542-543.

Winkler, D.R. (1975). Educational Achievement and School Peer Group Composition. *The Journal of Human Resources*, 10, pp.189–204.

TABLE 1
Journals that Regularly Publish Data and Code for Empirical Research Articles

- 1) Agricultural Economics
- 2) American Economic Journal: Applied Economics
- 3) American Economic Journal: Economic Policy
- 4) American Economic Journal: Macroeconomics
- 5) American Economic Journal: Microeconomics
- 6) Brookings Papers on Economic Activity
- 7) Econometrica
- 8) Economic Journal
- 9) Economics-The Open Access Open-Assessment E-Journal
- 10) European Economic Review
- 11) International Journal of Forecasting^a
- 12) Jahrbucher fur Nationalokonomie und Statistik/Journal of Economics and Statistics
- 13) Journal of Applied Econometrics
- 14) Journal of Labor Economics
- 15) Journal of Money Credit And Banking^b
- 16) Journal of Political Economy
- 17) Journal of The European Economic Association
- 18) Quarterly Journal of Economics
- 19) Review of Economic Studies
- 20) Review of Economics and Statistics^c
- 21) Review of International Organizations
- 22) Studies in Nonlinear Dynamics and Econometrics^d
- 23) The American Economic Review
- 24) World Bank Economic Review

OTHER^c:

The journal *Experimental Economics* commented: “We don't require individuals to post their data. We have never felt the need since there is a strong norm within the experimental community of sharing the data upon request (as well as instructions & z-tree code).”

The journal *Econ Journal Watch* does not regularly publish code, but they do regularly link their empirical articles to data, and have done so since the first issue of the journal in 2004.

NOTES: “Regularly” is defined as at least 50% of the empirical articles supply their data and code.

^a Some issues publish data and code for at least 50% of the empirical articles. The journal notes that it is currently in the process of moving all supplements to the ScienceDirect website which will make it easier for researchers to access them.

^b Data and code are published on the journal’s website: <http://www.jmcb.osu.edu/journal-index-and-archive> , but not on the Wiley online journal website.

^c The journal commented, “The Review of Economics and Statistics has an online data archive to which we require all of our published authors to post their Data and Code which is available to the public (<http://thedata.harvard.edu/dvn/dv/restat>).”

^d SNDE responded to our survey by noting that the journal “has required the inclusion of data and code for 17 years, before virtually any other journal.

^e Additional notes and comments from journals may be found on the accompanying Excel spreadsheet.

TABLE 2
Journals that Explicitly Mention that They Publish Replications

- 1) Econ Journal Watch
- 2) Economic Development and Cultural Change
- 3) Empirical Economics
- 4) Experimental Economics
- 5) International Journal of Forecasting
- 6) Jahrbucher Fur Nationalokonomie Und Statistik /Journal of Economics And Statistics
- 7) Journal of Applied Econometrics

OTHER^a:

The journal *Studies in Nonlinear Dynamics and Econometrics* noted, “The journal also has a replication section which publishes, albeit infrequently, replication studies.”

^a Additional notes and comments from journals may be found on the accompanying Excel spreadsheet.

TABLE 3
Distribution of Replications Across Journals

<i>Journal</i>	<i>Frequency Pct (Number)</i>	<i>Cumulative Pct</i>
Journal of Applied Econometrics	19.4 (30)	19.4
Journal of Human Resources	12.3 (19)	31.6
American Economic Review	9.7 (15)	41.3
Econ Journal Watch	6.5 (10)	47.7
Experimental Economics	5.8 (9)	53.5
Journal of Development Studies	5.8 (9)	59.4
Applied Economics	4.5(7)	63.9
Empirical Economics	4.5(7)	68.4
Journal of Economic & Social Measurement	3.9 (6)	72.3
Public Choice	3.9 (6)	76.1
Journal of Political Economy	1.9 (3)	78.1
Labour Economics	1.9 (3)	80.0
Economic Inquiry	1.3 (2)	81.3
Quarterly Journal of Economics	1.3 (2)	82.6
American Economic Journal: Applied Economics	0.6 (1)	83.2
American Law and Economics Review	0.6 (1)	83.9
Applied Financial Economics	0.6 (1)	84.5
Conflict Management and Peace Science	0.6 (1)	85.2
Econometrica	0.6 (1)	85.8
Economic Journal	0.6 (1)	86.5
European Economic Review	0.6 (1)	87.1
Health Economics	0.6 (1)	87.7
International Economics and Economic Policy	0.6 (1)	88.4
International Review of Applied Economics	0.6 (1)	89.0
Journal of Development Economics	0.6 (1)	89.7
Journal of Development Effectiveness	0.6 (1)	90.3
Journal of Environmental Economics and Management	0.6 (1)	91.0
Journal of International Development	0.6 (1)	91.6
Journal of International Trade & Economic Development	0.6 (1)	92.3
Journal of Law and Economics	0.6 (1)	92.9
Journal of Money, Credit, and Banking	0.6 (1)	93.5

<i>Journal</i>	<i>Frequency Pct (Number)</i>	<i>Cumulative Pct</i>
Journal of the European Economic Association	0.6 (1)	94.2
Marketing Letters	0.6 (1)	94.8
Proceedings of the National Academy of Sciences	0.6 (1)	95.5
Public Finance Review	0.6 (1)	96.1
Quarterly Journal of Business and Economics	0.6 (1)	96.8
Review of Economics and Statistics	0.6 (1)	97.4
Review of Financial Studies	0.6 (1)	98.1
Review of International Organizations	0.6 (1)	98.7
Social Science & Medicine	0.6 (1)	99.4
World Development	0.6 (1)	100.0

TABLE 4
Description of Characteristics

<i>Characteristic</i>	<i>Description</i>
<i>Summary?</i>	This is coded 1 if summarizes the results of a replication without reporting individual estimates.
<i>Exact?</i>	<p><u>FOR NON-EXPERIMENTAL STUDIES:</u> This is coded 1 if the replication uses the exact same data, specification, and estimation procedures as the original study (as much as possible). In other words, did the replication attempt to exactly reproduce the original results?</p> <p><i>NOTE#1: There are grey areas here. If a replication uses data or techniques that are similar to the original study (for example, simulation studies with the same DGP; maximum likelihood estimation of nonlinear models using different software), it is coded 1 even if the replication is not “exactly” the same. Another example: If a replication is working from a common data source, say Census data, and extracts data using the same criteria as the original study, it is coded 1 if the number of observations are the same or very similar.</i></p> <p><i>NOTE#2: Some replications mention in passing that they were able to reproduce the original results. If this is explicitly stated, it is coded 1.</i></p> <p><u>FOR EXPERIMENTAL STUDIES:</u> If the study attempted to create the same experimental environment – e.g., same payoffs, same instructions, same number of options, etc. – it is coded 1.</p>
<i>Extension?</i>	<p><u>FOR NON-EXPERIMENTAL STUDIES:</u> This is coded 1 if the replication attempts to extend the original findings (e.g., to see if the results are valid for a different country, or a different time period). It is coded 0 if it limits itself to determining whether the original results are valid (e.g., uses the same data, same country, same time period or slightly modified time period, but modifies the specification and/or estimation procedure).</p> <p><u>FOR EXPERIMENTAL STUDIES:</u> Experimental replications are coded 1 if they attempt to extend the original findings (e.g., by adding an hypothesis not considered by the original study).</p>
<i>Original Results?</i>	This is coded 1 if the replication explicitly reports an important estimate(s) from the original study so that it easy to make a direct comparison of results without having to go back to the original study.

<p><i>Negative?</i> <i>Mixed?</i> <i>Positive?</i></p>	<p>Negative is coded 1 whenever a significant difference with the original study is found and much attention is given to this.</p> <p>Mixed is coded 1 whenever there are significant confirmations of the original study, but significant differences are also found.</p> <p>Positive is coded 1 whenever the replication study generally affirms all the major findings of the original study.</p>
<p><i>Reply?</i></p>	<p>This is coded 1 whenever a reply/response from the original study accompanies the replication study.</p> <p><i>NOTE: This was usually determined by sighting the replication study on the website of the online version of the journal, and seeing if a reply/response was located contiguously.</i></p>

TABLE 5
Characteristics of Replication Studies by Journal Type

<i>Journals</i>	<i>Summary?</i>	<i>Exact?</i>	<i>Extension?</i>	<i>Original Results?</i>	<i>Negative?</i>	<i>Mixed?</i>	<i>Positive?</i>	<i>Reply?</i>
<i>All (155)</i>	0.052	0.639	0.503	0.594	0.665	0.123	0.213	0.200
<i>JAE (30)</i>	0.200	0.733	0.267	0.333	0.467	0.200	0.333	0.033
<i>Experimental (11)</i>	0.000	0.636	0.818	0.545	0.545	0.182	0.273	0.091
<i>Non-JAE/Non-Experimental (114)</i>	0.018	0.614	0.535	0.667	0.728	0.096	0.175	0.254

NOTE: Numbers in the table are averages of the respective 0-1 dummy variables (see TABLE 4 for explanation of categories and coding). The numbers in parentheses in the Journals column indicates the number of replication studies in each journal category.

FIGURE 1
Histogram of Replication Studies by Year

