EJW
ECON JOURNAL WATCH
Scholarly Comments on Academic Economics

ECON JOURNAL WATCH 12(2)
May 2015: 164–191

Replications in Economics: A Progress Report

Maren Duvendack[1], Richard W. Palmer-Jones[2], and W. Robert Reed[3]

LINK TO ABSTRACT

In producing an econometric study, a researcher makes multifarious decisions in fashioning his bridge from data to estimated results. If a separate researcher were to attempt to replicate the results, she would have great difficulty divining those decisions without accompanying data and code—the computer program that produces the estimates. Publication of data and code that allow other authors to reproduce an original study is necessary if researchers are to be confident they have correctly understood that original study. Thirty years ago, it was very difficult to obtain authors' data and code. Since then, there has been considerable progress, led by the American Economic Association, in making this standard practice, at least at some journals.

By itself, access to data and code might be inadequate to incentivize replication of studies: Researchers also need outlets to publish the results of replication efforts. If all economics journals made their data and code available, but no journals were willing to publish replication studies, then it is unlikely that more than a few such studies would be undertaken. Personal websites, social media, and other

1. School of International Development, University of East Anglia, Norwich NR4 7TJ, UK.
2. School of International Development, University of East Anglia, Norwich NR4 7TJ, UK.
3. Department of Economics and Finance, University of Canterbury, Christchurch 8140, New Zealand. (Corresponding author.) The authors thank Richard Anderson, Bruce McCullough, Badi Baltagi, Tom Stanley, and many others for comments on an earlier draft of this paper. We also 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.

outlets do allow 'unpublished' studies some access to a larger community of scholars, but in the absence of professional review it would be difficult for any but the most prominent replication studies to achieve notice in the profession. Further, the absence of publication credit would provide less incentive to undertake replication.

We provide a "progress report" on the use of replications in economics.[4] At least since the seminal study by William Dewald et al. (1986), there has been a recognition in the economics profession that many empirical results in economics are not reproducible or not generalizable to alternative empirical specifications, econometric procedures, extensions of the data, or other study modifications. A survey of the current literature reveals that ameliorating this state of affairs has not been easy. There have been substantial improvements in the sharing of data and code, but it is still rare for peer-reviewed journals to publish studies that replicate previous research.

The concern that a substantial portion of empirical research is not reproducible or generalizable is not restricted to economics and the social sciences. In December 2011, an issue of *Science* (**link**) was devoted to "Data Replication & Reproducibility" in the so-called 'hard sciences.' The concern with replication in science has become sufficiently widespread that it has crossed over to popular media. *The Economist*, *The New Yorker*, *The Atlantic*, *BBC Radio*, and the *Los Angeles Times* are just a few of the popular media outlets that have recently reported on concerns over reproducibility in scientific research.[5] And, while popular interest tends to focus on academic fraud, others have pointed out that academic practices generate a disproportionate rate of false positives (Maniadis et al. 2014; Ioannidis 2005; Ioannidis and Doucouliagos 2013; Paldam 2013; Camfield et al. 2014). Replication can provide a useful check on the spread of incorrect results. The use of replications should be of interest to many economists, even those not directly involved in the production of empirical research.

Our report provides a brief history of replication and data sharing in economics journals, as well as the results of a survey of replication policies at all 333 economics journals listed in *Web of Science*. Further, we analyse a collection of 162 replication studies published in peer-reviewed economics journals. We then discuss recent replication initiatives and offer suggestions on how replication analysis can be more effectively employed.

4. We offer definitions of 'replication' in this paper as there is currently no consensus among scholars. For an interesting discussion on defining 'replication' see Clemens (2015).

5. Here are links to the articles from *The Economist* (**link**), *The New Yorker* (**link**), *The Atlantic* (**link**), *BBC Radio* (**link**), and the *Los Angeles Times* (**link**).

This paper is part of an ongoing project which includes the website replicationnetwork.com (**link**), which provides additional, regularly updated information on replications in economics. Readers are encouraged to visit the site, both to stay abreast of developments and to contribute information that might be of interest to other researchers interested in replications.

A brief history of replications in economics

Replication and data sharing

From the early days of econometrics it has been acknowledged that sharing of data is desirable. Ragnar Frisch's introductory editorial to the new journal *Econometrica* said:

> 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, 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. But crucial is the availability of data and code. The issues raised by Frisch remain front and center.

In the post-World War II period, several scholars raised concerns about the quality of data and the validity of social and economic statistical analysis (Morgenstern 1950; Tullock 1959). Gordon Tullock was one of the first to draw attention to what is now commonly referred to as "the file drawer problem" (Rosenthal 1979): inconclusive findings are likely to be filed, while results that are statistically significant get published. Tullock 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" (Tullock 1959, 593).

The *Journal of Human Resources* (*JHR*) was an early leader in the publication of replications. Articles in *JHR* included replication as part of their analysis. For

example, an article by Marshall Smith (1968) says 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 … of the report. This is the case" (1968, 386 n.1). Replication, in the sense of repeating a prior analysis, was promoted in *JHR*; 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, 199).[6] Authors reported whether their results replicated or were consistent with the results of others (e.g., Winkler 1975, 202). Others reported replicating, or at least re-estimating, portions of other papers (e.g., Link and Ratledge 1975; see also Akin and Kniesner 1976; Link and Ratledge 1976).

The *Journal of Human Resources* continued to publish papers that contained replications through the 1970s and 1980s, and announced a commitment to doing so in its Winter 1990 issue:

> *JHR* 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 their data available to analysts from a date six months after *JHR* publication data for a period three years thereafter. …
> 3. The *JHR* welcomes replication, fragility, and sensitivity studies of empirical work that has appeared in the *JHR* in the last five years or empirical work judged by the editors to be important to the fields covered by the *JHR*.

The first two planks of this policy were reaffirmed in 2012 (**link**).

In a 1975 *Journal of Political Economy* (*JPE*) article, Edgar Feige asserted that economics journals' 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, 1291). Following its publication of Feige's article, the *JPE* initiated a "Confirmations and Contradictions" section, which existed from 1976 to 1999. The *JPE* editors wrote that confirmations could come from using new data, while contradictions would "be most powerful when based upon the same data" (editorial comment in Feige 1975, 1296). Philip Mirowski and Steven Sklivas (1991, 159), however, reported that only five of 36 notes appearing in this *JPE* section from 1976 to 1987 included replications, of which only one was "successful" in actually replicating the original results. Richard Anderson et al. (2008)

6. It appears that Gallaway and Dykman (1970) regard their paper, in part, as a replication of a report.

counted 13 more notes through 1999, of which only one 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, 108). We are not aware of any formal responses to Mirowski and Sklivas (1991) or Anderson et al. (2008).

In the 1980s few major economics journals had a data sharing or replication policy in place, even though some economists, such as Thomas Mayer (1980), expounded the need for replication. A notable exception at the time was the *Journal of Money, Credit and Banking* (*JMCB*) which requested that authors 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 provide data and code upon request or to deposit their data and code in journal-managed data archives upon submission of their article. Bruce McCullough et al. (2006) and Jan Höffler (2014) argue that the former are ineffective because most authors and editors ignore them (see also McCullough 2007).

Substantial progress has been made in the last two decades with respect to the publishing of replications and to the mandating of the provision of data and code. Several economic journals now have official policies on data sharing/archiving or replication, including the *American Economic Review* (*AER*), *Econometrica*, the *Journal of Applied Econometrics* (*JAE*), and a number of others. Less common is the requirement that authors provide their associated computer code. For example, the *JAE* encourages, but does not require, authors to supply code.

The *AER*'s policy statement, adopted in 2004 following the critical paper of McCullough and H. D. Vinod (2003), has served as a model for other journals. And the *AER* has recently 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 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)" (**link**). But much data manipulation commonly occurs between the 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 2013).[7]

The mandatory submission of raw data sets, along with the programs that produce the final data sets, would enable researchers to understand how the data were 'cleansed' and identify coding errors embodied in the final data set. These

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

issues are little discussed in the replication literature (see Glandon 2011 for an exception). Accordingly, Anderson et al. (2008, 99) 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." Even if more stringent requirements will remain impractical, discussions of the matter keep us mindful of a vast realm of abuse that policies like the *AER*'s do *not* ensure against.

The replication policy adopted by *Econometrica* (**link**) 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 archive (**link**), and a replication section (**link**). *JAE* clearly specifies the format in which data sets and computer code should be made available (again, making data available is mandatory for all papers published in *JAE* while the provision of code is voluntary).

The requirement of making available the author's data and code is a necessary, but by no means sufficient, condition that enables replicators to confirm the original study's results. It may be the case that policies are not strictly enforced; even the *AER* has been faulted for issues of non-compliance (Glandon 2011).[8] Or there could be 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 or data-sharing policies over recent years, but replication activities have only marginally increased. As the saying goes, incentives matter. Dewald et al. (1986) remarked that the incentives are low to undertake replication "however valuable in the search for knowledge." Following Thomas Kuhn (1970), they attribute the weakness of incentives 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 might be thought to be "lacking imagination" or "unable to allocate…time wisely." Replications may be seen as "reflecting a lack of trust in another scientist's integrity and ability" or "a personal dispute between researchers" (Dewald et al. 1986, 587).

8. Our survey about replication policies, discussed below, was administered to a large number of economics journals. Several 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.

Most discussions of incentives for replication include three actors—replicators, journal editors and original authors.[9] More recently one might take account of social commentators, journalists, bloggers, and the like. Would-be replicators reckon 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 getting a reputation of having an unfavourable personality, or advancing 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 arisen from errors in the original research or in the replication. Commentators have discussed such matters repeatedly. A recent example is the heated debate among social psychologists over the replication by David Johnson et al. (2014) of Simone Schnall et al. (2008).[10]

Furthermore, from their proprietary or confidential nature, many data sets are not made available for replication (see the *JAE* "Instructions for Authors" on proprietary data (**link**)), and many researchers are reluctant to share data sets when they have mind yet to appropriate their value in future research (Dewald et al. 1986).

There is a trend among such organizations as the Economic and Social Research Council (UK) and the National Science Foundation (U.S.) to insist that researchers they fund make their data sets publicly available.[11] In the non-profit aid research funding sector, the Gates Foundation has policies on data access (**link**). The International Initiative for Impact Evaluation (3ie) has adopted a policy that all data produced by funded activities be archived (**link**), although at the time of writing no precise protocols for this could be found.

Journal editors are concerned with per-page citations, and replications are thought to perform less well than original studies. Also, the editorial costs of allowing replication may be heavy when controversy between authors ensues. Editors may also be concerned about alienation of established authors. More importantly, they may wish to avoid the reputational consequences of exposing

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

10. For more details on this exchange, see Mukunth (2014) for an overview, Schnall (2014), and Schnall's posts on the blog of the Cambridge Embodied Cognition and Emotion Laboratory (**link**).

11. "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" (**link**). "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" (**link**). Gary King offers a list of further funding agencies with data sharing and archiving policies in place (**link**).

errors in papers published in their own journal. All of these factors attenuate editors' incentives to facilitate and encourage replication of the papers they publish.

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 is small or even negative. If replicating authors can easily confirm the original results, there is no real gain, while the damage to their reputation may be large if the original results cannot be confirmed. Original authors, then, may see their providing data and code as having a potential downside and very little upside.

Reputational issues magnify the difficulties associated with both original authors and replicators getting to 'the truth.' Recent cases involving Carmen Reinhart and Kenneth Rogoff (O'Brien 2013) and Thomas Piketty (Giles 2014) illustrate the amount of attention that can be attracted to controversies surrounding replication. Academics are sometimes 'named and shamed' through academic blogs and other informal media. Frequently the issues center around data errors and nuanced issues of sample selection and empirical procedures, not outright academic fraud.

In our Internet age, criticism can spread quickly and extensively, while rebuttals or more considered views do not necessarily attract much attention. The skills necessary to navigate successfully in social media may be orthogonal to scientific merit. While there are many impediments to replicators, the other side of the coin is that it can be difficult for the original authors to defend themselves when faced with unfair criticisms. Many journals provide the original authors with an opportunity to respond to replicators' findings, but more frequently this is not the case, as we show below.

With reputational issues motivating many disputes between replicating and original authors, there would seem to be a place for the establishment of protocols between replicators and replicatees to mitigate the possibilities of errors or misunderstandings in replications. In this context, Daniel Kahneman (2014) has recently called for the establishment of "A New Etiquette for Replication."[12] However, since reputation often plays out in the wide world of public opinion, and since that world operates under different rules than scientific responsibility, it is unlikely that such protocols would ever be able to fully safeguard against the harms done by malicious replicators or belligerent replicatees (see Hoxby 2007 or Acemoglu et al. 2012 for examples of sharp responses to replicators).

Several authors have suggested that the "push to replicate findings" in science could entail perverse effects (Bissell 2013; Gelman 2013). There is a perceived danger that authors could become more cautious and direct their efforts away from controversial or difficult topics (Schnall 2014). Difficult though these

12. See also the 3ie's "Replication Contracts Notification and Communication Policy" (**link**).

issues may be, the potential gains to the economics profession, and the public, of furthering the practice of replication are, in our view, substantial.

Current replication policies at *Web of Science* economics journals

Our investigation of current replication policies at economics journals began with the list of journals categorized as "Economics journals" by Thomson Reuters's *Journal Citation Reports*, which as of September 2013 was a total of 333 journals.[13] We researched each journal with respect to two questions: (i) Does the journal regularly publish data and code for its empirical research articles?, and (ii) Does the journal's website explicitly mention that it publishes replication studies?

We investigated the website of each journal. To determine whether a journal "regularly" publishes data and code for empirical research articles, we accessed recent online issues of the journal and counted the number of empirical research articles that were published: if at least 50 percent of these articles had attached data and code, the journal was classified as regularly publishing data and code. With respect to determining whether a journal's website explicitly mentions that it publishes replications, we read through website sections such as "About," "Aims and Scope," etc., for some statement that the journal invites submissions of replication studies or publishes 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. After a first draft of the paper was produced, we re-contacted the journal editors and asked them to again verify that our information was up-to-date and accurate. The response rates to these surveys were approximately 20 percent and 30 percent, respectively.[14]

Table 1 reports the results concerning availability of data and code for empirical articles. Twenty-seven of 333 journals met our standard for "regularly" publishing data and code. But many journals publish little content of the empirical sort to which replication pertains, so the absence of data and code should not

13. Journals were identified from the online 2012 *JCR* Social Science Edition and included all journals that were categorized as "Economics" in the Subject Category Selection dialog box.

14. Sixty-six journals responded to the first 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. A followup email was sent in January 2015 in which journal editors were asked to respond to the correctness of the information reported in an earlier draft of this study. Approximately 100 journals responded to that email.

be inferred as lack of support for the general policy of making this information available.

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)	*American Economic Review*
7)	*Brookings Papers on Economic Activity*
8)	*Econometrica*
9)	*Economic Journal*
10)	*Econometrics Journal*
11)	*Economics: The Open-Access, Open-Assessment E-Journal*
12)	*European Economic Review*
13)	*Explorations in Economic History*
14)	*International Journal of Forecasting* [(a)]
15)	*Jahrbücher für Nationalökonomie und Statistik/Journal of Economics and Statistics*
16)	*Journal of Applied Econometrics*
17)	*Journal of Labor Economics*
18)	*Journal of Money, Credit, and Banking* [(b)]
19)	*Journal of Political Economy*
20)	*Journal of the European Economic Association*
21)	*Quarterly Journal of Economics*
22)	*Review of Economic Dynamics*
23)	*Review of Economic Studies*
24)	*Review of Economics and Statistics* [(c)]
25)	*Review of International Organizations*
26)	*Studies in Nonlinear Dynamics and Econometrics* [(d)]
27)	*World Bank Economic Review*
Other: 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 (**link**), 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" (**link**). (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."	

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 as to whether these journals monitor whether published authors follow through on the responsibility, nor how journals might enforce it.

Even if such policies were duly enforced, there would be an advantage in locating the archive at the journal website. The journal can standardize the formatting of data and code. We did not inquire whether journals had policies about formatting, but our unscientific sampling of files suggests that authors are largely uninstructed in this area. We also did not inquire whether journals had internal processes for ensuring that the results of a published study are easily replicated with the files provided. In several ways, 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. Many researchers may find the benefits of providing transparent data and code files do not outweigh the costs. Journals can solve this incentive problem by making provision of data and code a condition for publication.

Table 2 lists the journals whose websites explicitly mention that they invite submission of replications or publish replications.[15] Some journals publish replications without explicitly stating that they do so. If journals are willing to publish replications, it is important that they say so in a public place, so potential authors can easily learn the fact. 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 she can publish her work.

Of the 333 journals examined, only 10 explicitly state that they publish replication studies (these journals are listed in Table 2). Further, some of these are specialty journals that only publish studies in a particular area, such as the journals *Experimental Economics* and *Economics of Education Review*. Others, such as the *Journal of Applied Econometrics*, only publish replications where the original article was published in one of a few elite journals. Thus, as a practical matter, there may only be one or two journals that appear willing to publish a replicating author's research. The lack of publishing outlets is perhaps the most serious obstacle to researchers interested in undertaking replication research.

15. 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.

TABLE 2. Journals whose websites explicitly mention that they publish replications

1)	*Econ Journal Watch*
2)	*Economic Development and Cultural Change*
3)	*Economics of Education Review*
4)	*Empirical Economics*
5)	*Experimental Economics*
6)	*Explorations in Economic History*
7)	*International Journal of Forecasting*
8)	*Jahrbücher für Nationalökonomie und Statistik/Journal of Economics and Statistics*
9)	*Journal of Applied Econometrics*
10)	*Review of International Organizations*

An analysis of published replications

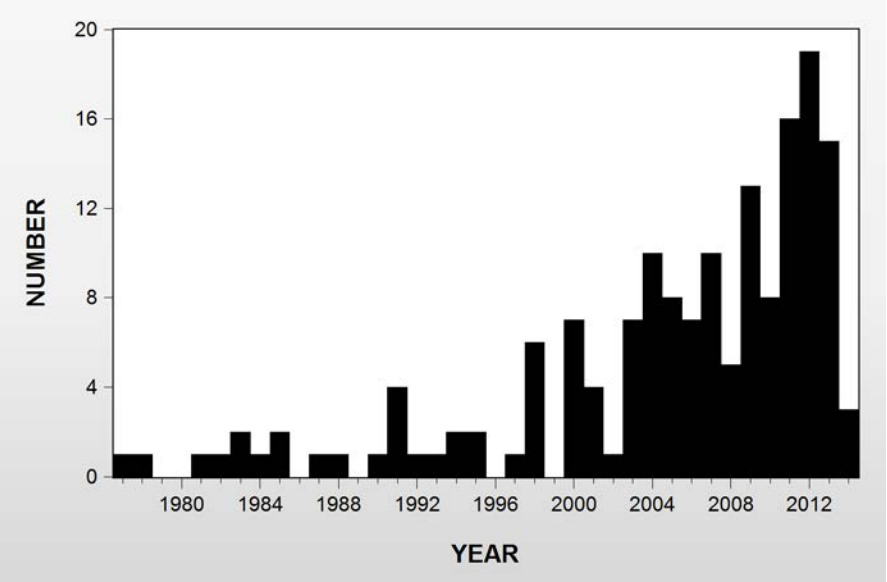
Here we analyze a number of published replication studies found in refereed economics journals. To be considered a replication study, an article had to (i) have as its main purpose to verify the reliability of a previously published study, and (ii) have been published in a peer-reviewed journal.[16] 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" wiki (**link**); (iii) suggestions from journal editors; and (iv) the present authors' own collections. Subsequent to that, we also did a more systematic search that targeted the top 50 economics journals (based on impact factors).[17]

16. 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, being colored by the incentive to defend the author's earlier research.

17. The impact factors were taken from RePEc (**link**). We actually referenced 51 journals, since two journals had identical impact factors. The journals are: *American Economic J.: Microeconomics, American Economic J.: Macroeconomics, American Economic Rev., Econometric Reviews, Econometrica, Econometrics J., Economic J., European Economic Rev., Experimental Economics, Games and Economic Behavior, International Economic Rev., International J. of Central Banking, J. of Accounting and Economics, J. of Applied Econometrics, J. of Development Economics, J. of Econometrics, J. of Economic Dynamics and Control, J. of Economic Growth, J. of Economic Perspectives, J. of Economic Surveys, J. of Economic Theory, J. of Empirical Finance, J. of Environmental Economics and Management, J. of Finance, J. of Financial and Quantitative Analysis, J. of Financial Economics, J. of Financial Intermediation, J. of Financial Markets, J. of Health Economics, J. of Human Resources, J. of International Business Studies, J. of International Economics, J. of International Money and Finance, J. of Labor Economics, J. of Law and Economics, J. of Monetary Economics, J. of Political Economy, J. of Population Economics, J. of Public Economics, J. of Risk and Uncertainty, J. of the European Economic Association, J. of Urban Economics, Labour Economics, Mathematical Finance, Oxford Bulletin of Economics and Statistics, Quarterly J. of Economics, RAND J. of Economics, Rev. of Economic Dynamics, Rev. of Economic Studies, Rev. of Financial Studies,* and *World Bank Economic Rev.*

For the systematic search of top-50 journals, each journal was searched using the term "replicat*." This generated 13,261 potentially relevant articles. Not having the means to screen all of these, we randomly sampled approximately 12% of them (1,601 articles), reviewing the full text to determine if the article satisfied our criteria to be classified as a "replication study." Of these 1,601 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.

Figure 1. Histogram of replication studies by year for our sample of 162 articles



In the end, our searching found 162 replication studies. Figure 1 presents a plot of replication studies (within the sample just described) 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 and was published in *Economic Inquiry* (Siskind 1977). Over the next 14 years (through 1991), fifteen more replication studies were published, eleven of which were published in one journal, the *Journal of Human Resources*. Early replication studies also appeared in *Applied Economics* (1983, 1985), the *Quarterly Journal of Economics* (1984), and the *Journal of Applied Econometrics* (1990).

From 1992, a number of other journals published replication studies: *Review of Economics and Statistics* (1992), *Journal of Development Studies* (1993, 1994, 2001),

Marketing Letters (1994), *Labour Economics* (1995), *Empirical Economics* (1997), *Journal of Environmental Economics and Management* (1998), *Public 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 William G. Dewald, Jerry G. Thursby, and Richard G. 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, edited by Badi Baltagi (Pesaran 2003). From that time on, the *JAE* has become the most prolific publisher of replication studies amongst economics journals. Another notable journal event was the start in 2004 of *Econ Journal Watch*; from the first issue, the journal's line of economic criticism has included replications (see Maberly and Pierce 2004, which itself comes under criticism in Witte 2010). As Figure 1 makes clear, journals have published replication studies with increasing frequency since the early 2000s.

Table 3 provides a listing of the journals that have published replication studies. The *JAE* 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*, the *Journal of Development Studies*, and *Experimental Economics*. These six journals account for almost 60 percent of all replication studies. Only ten economics journals have ever published more than three 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?*** Did the replication article merely summarize the findings of the replication? Or did it report individual estimates that allowed comparison with the original article?
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 these characteristics are described in more detail in Table 4. Table 5 reports the results. The numbers in the table are means of the corresponding dummy variables. As these numbers report population rather than sample values, hypothesis testing is not applicable.

The first characteristic, *Summary?*, is considered largely because of a practice of the *JAE* to sometimes publish paragraph-length summaries of replication studies. An example is David Drukker and Weihua Guan (2003), which reads in its entirety:

> 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.

The subsequent analysis separates out *JAE* replication studies from other journals' replication studies, as roughly a fifth of all *JAE* replications consist of short summaries. We also separate out experimental replication studies, because these 'replications' involve new data collection using different subjects, and often subjects from different countries. This raises issues of how 'reproducibility' should be interpreted. And so, we report the characteristics of replication studies for four categories of journals: (i) Studies from all journals (n=162), (ii) *JAE* studies (n=31), (iii) experimental studies (n=12), and (iv) non-*JAE*/non-experimental studies (n=119).

With respect to the *Exact?* characteristic, 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 frequent reason for not attempting to exactly reproduce an original study's findings is that a replicator attempts to confirm an original study's findings by using a different data set. An example is a replication study by Vegard Iversen and Richard Palmer-Jones (2008), which tested a result of Kaushik Basu et al. (2002) by using more recent data and data from a different country.

The next characteristic, *Extension?*, asks whether a replication study merely reproduces an original study's findings or also has some independent novelty or innovation (e.g., different data or additional hypotheses). On this dimension, there is wide variation across journal categories. Studies published in the *JAE* often

consist exclusively of attempts to confirm the original study's findings. Less than a third of *JAE* replication studies perform extensions of the original study. In contrast, most 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: we cannot tell if the difference between the *JAE* studies and the experimental studies, say, is driven by the preferences of journal editors or by the preferences of replicating authors.

The next characteristic, *Original Results?*, tells whether the replication study re-reports the original results in a way that facilitates comparison with the original study. A large portion of replication studies do not offer easy comparisons, perhaps because of limited journal space. Sometimes the lack of direct comparison is more than a minor inconvenience, as when a replication study refers to results from an original study without identifying the table or regression number from which the results come.

The next three characteristics involve the outcome of replication studies in confirming findings from the original study. Across all categories of journals and studies, 127 of 162 (78%) replication studies disconfirm a major finding from the original study. Interpretation of this number is difficult. One cannot assume that the studies treated to replication are a random sample. Also, researchers who confirm the results of original studies may face difficulty in getting their results published since they have nothing 'new' to report. On the other hand, journal editors are loath to offend influential researchers or editors at other journals. The *Journal of Economic & Social Measurement* and *Econ Journal Watch* have sometimes allowed replicating authors to report on their (prior) difficulties in getting disconfirming results published. Such firsthand accounts detail the reticence of some journal editors to publish disconfirming replication studies (see, e.g., Davis 2007; Jong-A-Pin and de Haan 2008, 57).

The last characteristic, *Reply?*, indicates how frequently journals publish a response by the original authors to the replication study in the same journal issue. Such replies are generally infrequent.[18] 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 33 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 (the exception being the exchange between Muñoz 2012 and Findlay and Santos 2012).

18. We did not search for replies that were published in later issues of the journal because of the right-hand censoring problem that arises from replies that have not yet been published. Our unscientific analysis is that most replies are published in the same issue as the comment/replication.

TABLE 3. Distribution of replications across journals

Journal	Frequency Pct. (Number)	Cumulative Pct.
Journal of Applied Econometrics	19.1 (31)	19.1
Journal of Human Resources	11.7 (19)	30.9
American Economic Review	9.3 (15)	40.1
Econ Journal Watch	6.8 (11)	46.9
Journal of Development Studies	6.2 (10)	53.1
Experimental Economics	5.6 (9)	58.6
Applied Economics	4.3 (7)	63.0
Empirical Economics	4.3 (7)	67.3
Journal of Economic and Social Measurement	3.7 (6)	71.0
Public Choice	3.7 (6)	74.7
Journal of Political Economy	1.9 (3)	76.5
Labour Economics	1.9 (3)	78.4
Economic Inquiry	1.2 (2)	79.6
Journal of Environmental Economics and Management	1.2 (2)	80.9
Quarterly Journal of Economics	1.2 (2)	82.1
Review of International Organizations	1.2 (2)	83.3
American Economic Journal: Applied Economics	0.6 (1)	84.0
American Law and Economics Review	0.6 (1)	84.6
Applied Financial Economics	0.6 (1)	85.2
Conflict Management and Peace Science	0.6 (1)	85.8
Econometrica	0.6 (1)	86.4
Economic Journal	0.6 (1)	87.0
European Economic Review	0.6 (1)	87.7
Health Economics	0.6 (1)	88.3
International Economics and Economic Policy	0.6 (1)	88.9
International Review of Applied Economics	0.6 (1)	89.5
Journal of Development Economics	0.6 (1)	90.1
Journal of Development Effectiveness	0.6 (1)	90.7
Journal of International Development	0.6 (1)	91.4
Journal of International Trade & Economic Development	0.6 (1)	92.0
Journal of Law and Economics	0.6 (1)	92.6
Journal of Money, Credit, and Banking	0.6 (1)	93.2
Journal of the European Economic Association	0.6 (1)	93.8
Journal of Urban Economics	0.6 (1)	94.4
Marketing Letters	0.6 (1)	95.1
Proceedings of the National Academy of Sciences	0.6 (1)	95.7
Public Finance Review	0.6 (1)	96.3
Quarterly Journal of Business and Economics	0.6 (1)	96.9
Review of Austrian Economics	0.6 (1)	97.5
Review of Economics and Statistics	0.6 (1)	98.1
Review of Financial Studies	0.6 (1)	98.8
Social Science & Medicine	0.6 (1)	99.4
World Development	0.6 (1)	100.0

TABLE 4. Description of characteristics

Characteristic	Description
Summary?	This is coded 1 if the replication article merely summarized the findings of the replication.
Exact?	**For non-experimental studies**: 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? (*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 data-generating process, or 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. *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.) **For experimental studies**: 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.
Extension?	**For non-experimental studies**: 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. **For experimental studies**: 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).
Original Results?	This is coded 1 if the replication explicitly reports an important estimate(s) from the original study such that it is easy to make a direct comparison of results without having to go back to the original study.
Negative? *Mixed?* *Positive?*	*Negative?* is coded 1 whenever a significant difference with the original study is found and much attention is given to this. *Mixed?* is coded 1 whenever there are significant confirmations of the original study, but significant differences are also found. *Positive?* is coded 1 whenever the replication study generally affirms all the major findings of the original study.
Reply?	This is coded 1 whenever a reply/response from the original study accompanies the replication study. (*Note*. This was determined by viewing the replication study on the website of the online version of the journal, and seeing if a reply/response from the original authors was located contiguously.)

TABLE 5. Characteristics of replication studies by journal type

Journals	Summary?	Exact?	Extension?	Original Results?	Negative?	Mixed?	Positive?	Reply?
All (162)	0.049	0.642	0.519	0.586	0.660	0.123	0.216	0.204
JAE (31)	0.194	0.742	0.290	0.323	0.452	0.194	0.355	0.032
Experimental (12)	0.000	0.750	0.833	0.583	0.500	0.167	0.333	0.083
Non-*JAE*/Non-Experimental (119)	0.017	0.605	0.546	0.655	0.731	0.101	0.168	0.261

Note: Numbers in the table are averages of the respective 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.

What can we learn from our analysis of replication studies? Most importantly, and perhaps not too surprisingly, the main takeaway is that, conditional on

the replication having been published, there is a high rate of disconfirmation. Over the full set of replication studies, approximately two out of every three studies were unable to confirm the original findings. Another 12 percent disconfirmed at least one major finding of the original study, while confirming others (*Mixed?*). In other words, nearly 80 percent of replication studies have found major flaws in the original research.

Could this be an overestimate of the true rate of Type I 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 are able to reproduce the results," as in Drukker and Guan 2003). 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. One may be tempted to take the *JAE*'s record as representative, and we see that the *JAE*'s rate of replications that disconfirm at least one major finding (that is, *Negative?* + *Mixed?*) is 65 percent (0.452+0.194). By any account, this is still a large number. It raises serious concerns about the reliability of published empirical research in economics.

The future of replications

Outside economics, there have been calls in many fields for an increase in replication activities. A well funded consortium including PLOS and Science Exchange operates a Reproducibility Initiative, which aims to independently verify the results of major scientific experiments (**link**). There have also been renewed calls for replication in the political sciences; Gary King's website (**link**) is a good resource, and another is Nicole Janz's Political Science Replication Blog (**link**). The Berkeley Initiative for Transparency in the Social Sciences (**link**) was started with the objective to make empirical social science research more transparent, promoting replications.

Economics has seen relatively few replication initiatives. One is the Replication in Economics project at Göttingen University, which is funded by the Institute for New Economic Thinking; it offers a wiki containing an extensive number of replication studies published in economic journals (**link**; see also Höffler 2014). Another replication initiative, in the field of development economics, has been launched by 3ie (**link**).

Will the current foothold in the journals expand? On the supply side are producers of replication. The increasing availability of 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 (see Figure 1). Further availability of data and code should result in more resources being devoted to replication research.

But production also depends on professional rewards, which in turn are related to the probability of publication in a respected journal. Stan Liebowitz (2014, 1272–1275) reports that quality of journal in which an author's work appears is the most important publication-related criterion for promotion. 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.'[19] Given that very few journals currently publish replication research, however, even a small increase in their absolute 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 replication market, an important determinant 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). 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. But journals of lesser status 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.

Demand is also affected by technological innovation. The *JAE*'s practice of publishing summaries of replications allows it to allocate less journal space for a replication study relative to an original 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 links to the full-length manuscript as an online supplement. Such innovations could increase the ratio of citations per journal page and hence could 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 increase the demand and supply of replication studies.

19. 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.

Replication as mitigation of publication bias

Annie 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 David Card and Alan Krueger (1995), Orley Ashenfelter et al. (1999), Chris Doucouliagos (2005), and Martin Paldam (2013), among others. Closely related is "HARKing", or "Hypothesizing After the Results are Known" (Kerr 1998). This is effectively data mining, where researchers stumble upon statistically significant results in their regression runs and then work backwards deductively to identify hypotheses consistent with those results.

Replication holds promise to mitigate such biases. 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. 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 in tandem with meta-analysis

Meta-analysis or 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 (Stanley and Doucouliagos 2012; Ringquist 2013; Paldam 2013). To date, replication and meta-analysis have largely lived parallel lives, but we hope to see more use of them 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. T. D.

Stanley and Stephen Jarrell (1989) identified the potential for meta-analysis and replication to work together more than 25 years ago.[20]

In our opinion, replication is an underappreciated and underutilized tool for assessing the reliability and validity of empirical results. It is our hope that this progress report and the accompanying website (**replicationnetwork.com**) will further this development.

Appendix

A file containing our reference list and analysis of replication studies can be downloaded **here**.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2012. The Colonial Origins of Comparative Development: An Empirical Investigation: Reply. *American Economic Review* 102(6): 3077–3110.

Akin, John S., and Thomas J. Kniesner. 1976. Proxies for Observations on Individuals Sampled from a Population. *Journal of Human Resources* 11(3): 411–413.

Anderson, Richard G., William H. Greene, B. D. McCullough, and H. D. Vinod. 2008. The Role of Data/Code Archives in the Future of Economic Research. *Journal of Economic Methodology* 15(1): 99–119.

Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek. 1999. A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias. *Labour Economics* 6(4): 453–470.

Baltagi, Badi H., and Sophon Khanti-Akom. 1990. On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators. *Journal of Applied Econometrics* 5(4): 401–406.

Basu, Kaushik, Ambar Narayan, and Martin Ravallion. 2002. Is Literacy Shared Within Households? Theory and Evidence from Bangladesh. *Labour Economics* 8(6): 649–665.

20. Stanley and Jarrell (1989, 167f.) write: "Meta-regression analysis provides us with a framework for quantitative surveys of the empirical literature, replication, and self-analysis. … MRA also presents interesting possibilities for organizing and encouraging replication of empirical economic research. … It is at once a framework in which to organize and interpret exact and inexact replications, to review more objectively the literature and explain its disparities, and to engage in the self-analysis of investigating the socioeconomic phenomenon of social scientific research itself."

Bissell, Mina. 2013. Reproducibility: The Risks of the Replication Drive. *Nature* 503(7476): 333–334. **Link**

Burman, Leonard E., W. Robert Reed, and James Alm. 2010. A Call for Replication Studies. *Public Finance Review* 38(6): 787–793.

Camfield, Laura, Maren Duvendack, and Richard W. Palmer-Jones. 2014. Things You Wanted to Know About Bias in Evaluations but Never Dared to Think. *IDS Bulletin* 45(6): 49–64.

Camfield, Laura, and Richard W. Palmer-Jones. 2013. Three 'Rs' of Econometrics: Repetition, Reproduction and Replication. *Journal of Development Studies* 49(12): 1607–1614.

Card, David, and Alan B. Krueger. 1995. Time-Series Minimum-Wage Studies: A Meta-Analysis. *American Economic Review* 85(2): 238–243.

Clemens, Michael. 2015. The Meaning of Failed Replications: A Review and Proposal. *Working Paper* 399, Center for Global Development (Washington, D.C.). **Link**

Davis, Graham A. 2007. Reflections of a First-Time Replicator. *Journal of Economic and Social Measurement* 32(4): 263–266.

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson. 1986. Replication in Empirical Economics: the *Journal of Money, Credit and Banking* Project. *American Economic Review* 76: 587–603.

Doucouliagos, Chris. 2005. Publication Bias in the Economic Freedom and Economic Growth Literature. *Journal of Economic Surveys* 19(3): 367–387.

Drukker, David M., and Weihua Guan. 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.

Feige, Edgar L. 1975. The Consequences of Journal Editorial Policies and a Suggestion for Revision. *Journal of Political Economy* 83(6): 1291–1296.

Feigenbaum, Susan, and David M. Levy. 1993. The Market for (Ir)Reproducible Econometrics. *Social Epistemology* 7(3): 215–232.

Findlay, David W., and John M. Santos. 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. **Link**

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. Publication Bias in the Social Sciences: Unlocking the File Drawer. *Science* 345(6203): 1502–1505.

Frisch, Ragnar. 1933. Editor's Note. *Econometrica* 1(1): 1–4.

Gallaway, Lowell E., and Zachary Dyckman. 1970. The Full Employment–Unemployment Rate: 1953–1980. *Journal of Human Resources* 5(4): 487–510.

Gelman, Andrew. 2013. Replication Backlash. *Statistical Modeling, Causal Inference, and Social Science*, December 17. **Link**

Gibson, John, David L. Anderson, and John Tressler. 2014. Which Journal Rankings Best Explain Academic Salaries? Evidence from the University of California. *Economic Inquiry* 52(4): 1322–1340.

Giles, Chris. 2014. Piketty Findings Undercut by Errors. *Financial Times*, May 23. **Link**

Glandon, Philip J. 2011. Appendix to the Report of the Editor: Report on the *American Economic Review* Data Availability Compliance Project. *American Economic Review* 101(3): 695–699.

Höffler, Jan H. 2014. Teaching Replication in Quantitative Empirical Economics. Presented at European Economic Association & Econometric Society 2014 Parallel Meetings (Toulouse), August 28. **Link**

Hoxby, Caroline M. 2007. Does Competition Among Public Schools Benefit Students and Taxpayers? Reply. *American Economic Review* 97(5): 2038–2055.

Ioannidis, John P. A. 2005. Why Most Published Research Findings Are False. *PLOS Medicine* 2(8): 696–701. **Link**

Ioannidis, John P. A., and Chris Doucouliagos. 2013. What's to Know About the Credibility of Empirical Economics? *Journal of Economic Surveys* 27(5): 997–1004.

Iversen, Vegard, and Richard W. Palmer-Jones. 2008. Literacy Sharing, Assortative Mating, or What? Labour Market Advantages and Proximate Illiteracy Revisited. *Journal of Development Studies* 44(6): 797–838.

Iversen, Vegard, and Richard W. Palmer-Jones. 2014. TV, Female Empowerment and Demographic Change in Rural India. *3ie Replication Paper* 2, International Initiative for Impact Evaluation (Washington, D.C.). **Link**

Jasny, Barbara R., Gilbert Chin, Lisa Chong, and Sacha Vignieri. 2011. Again, and Again, and Again. *Science* 334(6060): 1225.

Jensen, Robert, and Emily Oster. 2009. The Power of TV: Cable Television and Women's Status in India. *Quarterly Journal of Economics* 124(3): 1057–1094.

Jensen, Robert, and Emily Oster. 2014. TV, Female Empowerment and Fertility Decline in Rural India: Response to Iversen and Palmer-Jones. International Initiative for Impact Evaluation (Washington, D.C.). **Link**

Johnson, David J., Felix Cheung, and M. Brent Donnellan. 2014. Does Cleanliness Influence Moral Judgments? A Direct Replication of Schnall, Benton, and Harvey (2008). *Social Psychology* 45(3): 209–215.

Jong-A-Pin, Richard, and Jakob de Haan. 2008. Growth Accelerations and Regime Changes: A Correction. *Econ Journal Watch* 5(1): 51–58. **Link**

Kahneman, Daniel. 2014. A New Etiquette for Replication. *Social Psychology* 45(4): 310–311.

Kerr, Norbert L. 1998. HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review* 2(3): 196–217.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: University of Chicago Press.

Liebowitz, Stan J. 2014. Willful Blindness: The Inefficient Reward Structure in Academic Research. *Economic Inquiry* 52(4): 1267–1283.

Link, Charles R., and Edward C. Ratledge. 1975. Social Returns to Quantity and Quality of Education: A Further Statement. *Journal of Human Resources* 10(1): 78–89.

Link, Charles R., and Edward C. Ratledge. 1976. Proxies for Observations on Individuals Sampled from a Population: Reply. *Journal of Human Resources* 11(3): 413–419.

Maberly, Edwin D., and Raylene M. Pierce. 2004. Stock Market Efficiency Withstands Another Challenge: Solving the "Sell in May/Buy After Halloween" Puzzle. *Econ Journal Watch* 1(1): 29–46. **Link**

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2014. One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American Economic Review* 104(1): 277–290.

Mayer, Thomas. 1980. Economics as a Hard Science: Realistic Goal or Wishful Thinking? *Economic Inquiry* 18(2): 165–178.

McCrary, Justin. 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. 2007. Got Replicability? The Journal of Money, Credit and Banking Archive. *Econ Journal Watch* 4(3): 326–337. **Link**

McCullough, B. D., Kerry Anne McGeary, and Teresa D. Harrison. 2006. Lessons from the *JMCB* Archive. *Journal of Money, Credit, and Banking* 38(4): 1093–1107.

McCullough, B. D., and H. D. Vinod. 2003. Verifying the Solution from a Nonlinear Solver: A Case Study. *American Economic Review* 93(3): 873–892.

Mirowski, Philip, and Steven Sklivas. 1991. Why Econometricians Don't Replicate (Although They Do Reproduce). *Review of Political Economy* 3(2): 146–163.

Morgenstern, Oskar. 1950. *On the Accuracy of Economic Observations*. Princeton, N.J.: Princeton University Press.

Mukunth, Vasudevan. 2014. Replication Studies, Ceiling Effects, and the Psychology of Science. *Is Nerd*, May 26. **Link**

Muñoz, Robert Jr. 2012. Beyond Race Cards in America's Pastime: An Appreciative Reply to Findlay and Santos. *Econ Journal Watch* 9(2): 141–148. **Link**

O'Brien, Matthew. 2013. Forget Excel: This Was Reinhart and Rogoff's Biggest Mistake. *The Atlantic*, April 18. **Link**

Paldam, Martin. 2013. Regression Costs Fall, Mining Ratios Rise, Publication Bias Looms, and Techniques Get Fancier: Reflections on Some Trends in Empirical Macroeconomics. *Econ Journal Watch* 10(2): 136–156. **Link**

Pesaran, Hashem. 2003. Introducing a Replication Section. *Journal of Applied Econometrics* 18(1): 111.

Ringquist, Evan J. 2013. *Meta-Analysis for Public Management and Policy*, ed. Mary R. Anderson. San Francisco: Jossey-Bass.

Rosenthal, Robert. 1979. The "File Drawer Problem" and Tolerance for Null Results. *Psychological Bulletin* 86(3): 638–641.

Schnall, Simone. 2014. Clean Data: Statistical Artifacts Wash Out Replication Efforts. *Social Psychology* 45(4): 315–317.

Schnall, Simone, Jennifer Benton, and Sophie Harvey. 2008. With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments. *Psychological Science* 19: 1219–1222.

Siskind, Frederic B. 1977. Minimum Wage Legislation in the United States: Comment. *Economic Inquiry* 15(1): 135–138.

Smith, Marshall S. 1968. Equality of Educational Opportunity: Comments on Bowles and Levin. *Journal of Human Resources* 3(3): 384–389.

Stanley, T. D., and Hristos Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business*. New York: Routledge.

Stanley, T. D., and Stephen B. Jarrell. 1989. Meta-Regression Analysis: A Quantitative Method of Literature Surveys. *Journal of Economic Surveys* 3(2): 161–170.

Tullock, Gordon. 1959. Publication Decisions and Tests of Significance—A Comment. *Journal of the American Statistical Association* 54(287): 593.

Wilhite, Allen W., and Eric A. Fong. 2012. Coercive Citation in Academic Publishing. *Science* 335(6068): 542–543.

Winkler, Donald R. 1975. Educational Achievement and School Peer Group Composition. *Journal of Human Resources* 10(2): 189–204.

Witte, H. Douglas. 2010. Outliers and the Halloween Effect: Comment on Maberly and Pierce. *Econ Journal Watch* 7(1): 91–98. **Link**

About the Authors

Maren Duvendack is a Lecturer in Development Economics in the School of International Development, at the University of East Anglia. Her key research areas cover development economics, applied micro-econometrics, impact evaluation, systematic reviews and meta-analysis, microfinance, replication and reproduction of quantitative analyses, and research ethics. After completing her Ph.D. she joined the International Food Policy Research Institute in Washington, D.C., as a Postdoctoral Fellow before moving to the Overseas Development Institute in London as a Research Fellow in evaluation and impact assessment. She is currently involved in evaluating the impact of IFAD and DFAT/AusAid funded projects in a number of developing countries as well as 2 meta-analyses for DFID. Her email address is m.duvendack@uea.ac.uk.

Richard Palmer-Jones is a development economist who lived and worked in Malawi and Nigeria for most of the 1970s and has been working mainly on Bangladesh and India since the early 1980s. His work encompasses both primary and secondary data on topics including irrigation, agricultural growth, poverty measurement, agricultural wages, gender, education, and nutritional status. He has a Ph.D. from the University of Reading and has worked at the Universities of Reading, Ahmandu Bello (Nigeria), London, Oxford, and East Anglia. He is three times winner of the Dudley Seers Prize for the best article in the *Journal of Development Studies*, of which he is now a Managing Editor. His email address is r.palmer-jones@uea.ac.uk.

Bob Reed received his Ph.D. from Northwestern University in 1985. He is currently Professor of Economics at the University of Canterbury in New Zealand. He has taught at Texas A&M University, the University of Texas at Austin, and the University of Oklahoma. Professor Reed has published in the areas of labor economics, public choice, public finance, and applied econometrics. His work has appeared in numerous professional journals, including the *Journal of Political Economy*, *Journal of Public Economics*, *Journal of Labor Economics*, *Journal of Urban Economics*, *Journal of Human Resources*, *Oxford Bulletin of Economics & Statistics*, *Economic Inquiry*, and the *National Tax Journal*. His recent research interests include replication and meta-analysis, and he has a long-standing research interest in the relationship between taxes and economic growth. His email address is bob.reed@canterbury.ac.nz.

Go to archive of Economics in Practice section
Go to May 2015 issue