Research Policy 48 (2019) 62–83

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Replication studies in economics—How many and which papers are chosen for replication, and why?[☆]

Frank Mueller-Langer[a,b,*], Benedikt Fecher[c,d], Dietmar Harhoff[b], Gert G. Wagner[c,d,e]

[a] European Commission, Joint Research Centre, Digital Economy Unit, Spain
[b] Max Planck Institute for Innovation and Competition, Germany
[c] Alexander von Humboldt Institute for Internet and Society, Germany
[d] German Institute for Economic Research (DIW Berlin), Germany
[e] Max Planck Institute for Human Development, Germany

ARTICLE INFO

JEL classification:
A1
B4
C12
C13

Keywords:
Replication
Economics of science
Science policy
Economic methodology

ABSTRACT

We investigate how often replication studies are published in empirical economics and what types of journal articles are replicated. We find that between 1974 and 2014 0.1% of publications in the top 50 economics journals were replication studies. We consider the results of published formal replication studies (whether they are negating or reinforcing) and their extent: Narrow replication studies are typically devoted to mere replication of prior work, while scientific replication studies provide a broader analysis. We find evidence that higher-impact articles and articles by authors from leading institutions are more likely to be replicated, whereas the replication probability is lower for articles that appeared in top 5 economics journals. Our analysis also suggests that mandatory data disclosure policies may have a positive effect on the incidence of replication.

1. Introduction

Scientific research plays an important role in the advancement of technologies and the fostering of economic growth (Dasgupta and David, 1994; Dosi, 1988; Murray et al., 2016; Romer, 1986; Sorenson and Fleming, 2004; Stephan, 1996). Hence, the production of thorough and reliable scientific results is crucial from a social welfare and (science) policy perspective (Furman et al., 2012; Kiri et al., 2018; Lacetera and Zirulia, 2011; Stephan, 2012).

However, in times of increasing retractions and frequent instances of scientific fraud and misconduct, scientific quality assurance mechanisms are subject to a high level of scrutiny (McNutt, 2014; Steen et al., 2013; Lacetera and Zirulia, 2011; Cokol et al., 2008). Therefore, formal published replication studies and informal replication can be seen as important post-publication quality checks in addition to the established pre-publication peer review process (Coffman et al., 2017; Coffman and Niederle, 2015). Replicability has been described as an ideal standard of good scientific practice (Popper, 1959; Jasny et al., 2011; Campbell and Stanley, 1963). Maintaining this ideal standard becomes even more important in light of the increase in publications and growing demand from publishers, funding bodies, and policy makers to make research more transparent (European Commission, 2012, 2016). Increasingly powerful research infrastructures to support these demands, e.g., services that host data and code, reinforce this dynamic (Bohannon, 2015; Hoeffler, 2017).

While issues regarding the replicability of scientific research have been reported in multiple scientific fields including economics (Anderson et al., 2008; Dewald et al., 1986; Camerer et al., 2016), systematic empirical evidence of the incidence of replication studies in economics and their determinants is scarce.[1] Considering the relevance and influence of economic research beyond academia—in a field like evidence-based policy making, for instance—as well as the increasing impact (measured by extramural citations) of empirical economic research on other scientific disciplines (Angrist et al., 2017), there is a particular need to explore and understand

[☆] The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

* Corresponding author at: European Commission, Joint Research Centre (JRC), Directorate Growth & Innovation, Digital Economy Unit, Edificio Expo. C/ Inca Garcilaso 3, 41092 Seville, Spain.

E-mail address: frank.mueller-langer@ip.mpg.de (F. Mueller-Langer).

[1] Notable exceptions are Berry et al. (2017), Duvendack et al. (2015), and Hamermesh (2017).

https://doi.org/10.1016/j.respol.2018.07.019
Received 22 February 2017; Received in revised form 24 July 2018; Accepted 26 July 2018
Available online 02 October 2018

the drivers of replication studies in economics. Such evidence is a necessary precursor for designing favorable boundary conditions for professional self-control and for promoting trust in the scientific enterprise (Anderson and Kichkha, 2017; McCullough et al., 2006). A better understanding of the incidence and drivers of published formal replication studies is also important in light of recent predictions made by economic theory (Lacetera and Zirulia, 2011; Kiri et al., 2018). For instance, in their game-theoretical analysis of high-quality scientific production, Kiri et al. (2018, p. 835) underline "that not only are scientific findings never complete or definitive and are always prone to improvement; but, also, that observing only apparently definitive or undisputed findings may be a sign of weakness of a scientific field rather than a proof of its solidity."

Here, we investigate formal, i.e., published, replication studies in economics by examining which and how many published papers are selected for replication and what factors drive replication in these instances. While there are also informal replication studies that are not published in scientific journals (especially replications conducted in teaching) and an increasing number of other forms of post-publication review (e.g., meta-analyses or discussions on websites such as PubPeer), these are not covered with our approach. The replication studies in our sample may differ in terms of (a) their results, i.e., they may negate or (partially) reinforce the results of the replicated article, and (b) the extent of replication (narrow or scientific). Narrow replication studies are typically entirely devoted to the replication of a particular result. By contrast, the replication work in scientific replication studies is often embedded in a broader analysis. Note that all replication studies in our sample, irrespective of the replication result and extent, have in common that their *main* purpose is to evaluate the accuracy of prior research, i.e., to replicate another article. We use metadata about all articles published in the top 50 economics journals between 1974 and 2014.

We find that replication studies are a matter of impact: high-impact articles and articles by authors from leading institutions are more likely to be replicated. Furthermore, our results suggest that the probability of a replication study being conducted is lower for articles published in the top five economics journals. We also find that mandatory data disclosure policies may have a positive effect on the incidence of replication. Finally, we discuss the implications of these results. Our results suggest that replication efforts could be incentivized by reducing the cost of replication.

2. Background: replication in economics

In 1982, the *Journal of Money, Credit and Banking* (JMCB) initiated the JMCB Data Storage and Evaluation Project (Dewald et al., 1986). Within this project, the JMCB required authors to make the data and code used in their articles available to other researchers upon request. In a second part of this project, Dewald et al. (1986) conducted replications of nine articles for which the data was made fully available.[2] They were able to replicate the results of two articles in their entirety. Later, McCullough et al. (2006) tried to replicate 62 articles submitted to the same journal and could fully replicate the results of 14 articles.[3] Also, McCullough et al. (2008) tried to replicate 133 research articles published by the *Federal Reserve Bank of St. Louis Review* and could replicate 29. These results raised concerns with respect to the technical and factual reproducibility of empirical work in economics.

The debate surrounding Reinhart and Rogoff (2010) and the non-supportive replications by Herndon et al. (2014) and Bell et al. (2015) further advanced the debate about replication in the field of economics.[4] The study by Camerer et al. (2016), in which the authors attempted to replicate 18 studies published in two top economic journals—the *American Economic Review* and the *Quarterly Journal of Economics*—between 2011 and 2014 drew renewed attention to the issue of replicability in economics. The authors were able to find a significant effect in the same direction as proposed by the original research in only 11 out of 18 replications (61%). It should be noted, however, that the inability of researchers to replicate each other's results is well known from other disciplines (in biology, for example; Begley and Ellis, 2012). In these disciplines, a single paper may be linked to a patent or a product in certain cases (Fehder et al., 2014; Gans et al., 2017; Murray and Stern, 2007).[5]

While replications for entire studies are rarely attempted, there is evidence for indirect forms of validation and informal replication. For example, Berry et al. (2017) assess the rate of replication for empirical papers in the 2010 *American Economic Review* and find that across 70 empirical papers, 29% have 1 or more citations that partially replicate the original result. While only a minority of the papers under scrutiny have a published replication, a majority (60%) have been subject to a replication, a robustness test, or an extension. In addition, Hamermesh (2017) examines the most heavily cited publications in labor economics from the early 1990s. He shows that many highly cited papers are replicated using data from other time periods and/or economies. Therefore, he suggests that the validity of the central ideas put forward by high-impact articles in labor economics have usually been tested. Finally, Fecher et al. (2016) examine replication efforts by the large community of users of the German Socio-Economic Panel Study (SOEP) and find that most replications are conducted in the context of teaching. These informal replications are usually not part of the published academic knowledge base and are therefore absent from the documentation of replication efforts.

3. Conceptual framework

In order to better understand the mechanics behind replication, it is helpful to review the different strands of literature on replicability in economics. Here we focus on research, impact, and competition, as well as costs of replication. We also distinguish between narrow and scientific replications in our examination of the incidence of replication studies (Baltagi, 2003; Dewald et al., 1986; Duvendack et al., 2015; Hamermesh, 2007; Hunter, 2001).

[2] From the first 54 data sets submitted to JMCB, 14 were judged incomplete. For instance, the identification of the source of data was the most frequent problem in this respect.

[3] From 186 empirical articles, 69 had entries in the data archive. Replication of seven articles was not possible because of confidentiality of data or lack of software.

[4] Notably, Carmen M. Reinhart and Kenneth S. Rogoff facilitated the detection of their coding error by making the data available upon request to Thomas Herndon, Michael Ash, and Robert Pollin. In addition, Reinhart et al. (2012) provide an erratum to Reinhart and Rogoff (2010) where they correct their coding error. Reinhart et al. (2012) also address some of the methodology and exclusion issues raised by Herndon et al. (2014). Finally, Bell et al. (2015) used Reinhart and Rogoff's data, made available by Herndon et al. (2014), to re-examine the relationship between growth and debt in developed countries. Using a multilevel distributed lag model, Bell et al. (2015, p. 470) provide evidence for a reverse causal link "predominantly in the opposite direction to that mooted by Reinhart and Rogoff."

[5] In addition, the scientific community has repeatedly experienced instances of misconduct and erroneous analyses (Lacetera and Zirulia, 2011; Azoulay et al., 2015). The data fraud scandal concerning social psychologist Diedrik Stapel (Levelt et al., 2012), Hwang's fraudulently reported breakthroughs in stem-cell research (Cyranoski, 2006), or Schoen's entirely fabricated results on organic transistors in over 40 publications (Grant, 2002) are only the most prominent examples.

3.1. Impact, competition, and costs

Several authors relate replication to impact and competition. For example, Hamermesh (2007, 2017) proposes that the probability that an individual will attempt a replication increases with the visibility of the published results that then become subject to replication. Furman et al. (2012) suggest that results from frequently cited articles generate more interest and scrutiny and—possibly due to a higher probability of replication—have a higher likelihood of retraction. There is also empirical evidence that replication is a matter of impact. Sukhtankar (2017) examined replication studies of empirical papers in development economics published in the top five and subsequent five general interest journals between 2000 and 2015. Of the 1056 articles, 57 (5.4%) were replicated in another published article or working paper. The strongest predictor of whether a study would be replicated or not is the article's Google Scholar citation count, followed by the year of publication (older articles are less likely to be replicated). Card and Della Vigna (2013) provide evidence that competition for publication in the top economics journals has significantly intensified over the last 50 years, i.e., the acceptance rate in the top economics journals has fallen from 15 to 6% between 1970 and 2012. The authors propose that due to increased competition, researchers have improved the quality of their work. Furthermore, journal editors can be more selective in choosing from a large number of submissions (Feigenbaum and Levy, 1993). It is for this reason that the editors of JPE (1975, p. 1295) suggest that "(t) he true remedy (against careless empirical research) is to resort to the powerful force of competition." In a similar fashion, Lacetera and Zirulia (2011) argue that while referees examine the accuracy of submitted articles, it is ultimately the competitive environment that will lead to critical review and replication.

In a game-theoretical analysis of high-quality scientific production, Kiri et al. (2018) examine the role of top journals as quality certifiers. In particular, they explore the question of whether the reliability of published results is higher for more highly rated journals. The results of their analysis suggest that high-status journals will not publish low-quality papers because reputational mechanisms deter them from publishing articles of lower quality. In this scenario, the optimal strategy for a researcher is to first submit high-quality work to a top journal and only submit to second-tier journals after rejection from top journals. Even if this submission and publication strategy really does ensure the highest quality of articles, it also has important implications for potential replicators as they will never try to replicate articles published in a top journal. Kiri et al. (2018, p. 835) also show that "if journals can perfectly sort out the quality of papers, then the equilibrium in which a scientist exerts high effort (…) and no verification activities occur is sustainable." Notably, this result depends on the assumption of a perfect selection mechanism. While we do not believe that a perfect selection mechanism is possible, we argue that top five economics journals will be more likely to be able to assess the quality of papers. Based on this theoretical prediction, we therefore propose that articles published in top five economics journals are less likely to be eventually replicated.

Dewald et al. (1986) and Hamermesh (2007) suggest that replication studies are more likely to be published when a central result of the original study is contradicted, e.g., they detect error or fraud, and that a replicator's main objective is to publish a correction or comment. On the basis of this strand of literature, we propose that competition in general and the news value of replication results in particular can serve as an explanation for the instances of replication studies in academic journals.

Another strand of literature focuses on the role of access to data and supplementary material for replication via mandatory data disclosure policies. Several authors suggest that data disclosure may enhance the quality of articles as it reduces the cost of checking empirical results and encourages more careful research (Frisch, 1933; Lacetera and Zirulia, 2011). Dewald et al. (1986) suggest that data disclosure decreases the frequency and magnitude of errors in published articles. Analogously, the (effort) costs of replication can be considered an important factor for conducting replication studies. In their game-theoretical analyses of scientific fraud and high-quality scientific production, Lacetera and Zirulia (2011) and Kiri et al. (2018) suggest that the cost of replication is likely to be lower if data is made available under a data disclosure policy. In addition, the cost of replication is likely to be particularly high when confidential or self-created data used in an article is not made available under a data disclosure policy. Previous research in this area addressed the reluctance among scientists to make research data available (Andreoli-Versbach and Mueller-Langer, 2014; Haeussler, 2011; Savage and Vickers, 2009; Tenopir et al., 2011; Fecher et al., 2016), poor data documentation in published research (Ioannidis, 2005), and missing data availability policies in scientific journals (McCullough et al., 2008; Vlaeminck, 2013).[6]

In addition, we consider the incentives for researchers to engage in replication studies. Under the current scientific reward system, which depends heavily on article publication, a researcher would typically choose between allocating time to writing an original research article and conducting a replication study. According to Feigenbaum and Levy (1993, p. 217), the optimization problem of the researcher "is dictated by the relative returns in citations yielded by original versus replication work, as compared to the relative time cost of the two alternative endeavors." Feigenbaum and Levy (1993) also suggest that the expected citation returns for replication studies are lower than those of original research articles. This provides an explanation why replication studies are rarely conducted. Or, as Dewald et al. (1986, p. 587) put it, "replication (…) does not fit within (…) the reward structure in scientific research."[7] However, under the aforementioned conditions, it will only be beneficial for a researcher to engage in replication studies if it is substantially less costly to produce them. Feigenbaum and Levy (1993) argue that the availability of data and code will decrease these costs. Recent evidence supports this theoretical prediction. Following their attempt to replicate 67 macroeconomic papers, Chang and Li (2017) suggest that it was substantially easier to replicate previously published findings when they were published in journals that have mandatory data disclosure policies. Furthermore, Chang and Li (2017, p. 63) conclude that "the most common cause of our inability to replicate findings was that authors did not provide files to the journal replication archives."

Finally, from the perspective of journals, Feigenbaum and Levy (1993) suggest that an editor's rationale for publishing replications of their own journal's articles differs from the decision to publish replications of another journal's articles. In particular, they suggest that a rational, citation-maximizing editor of a high-ranking journal may decline the publication of a replication study on an article published in her own journal anticipating that editors of lower-ranking journals, in expectation of positive citation effects, may still find it beneficial to publish this replication study. It is in this respect that the editor of a high-ranking journal "captures the citation externality without rewarding the replicator" (Feigenbaum and Levy, 1993, p. 223).

3.2. Types of replication

Following Hamermesh (2007), we distinguish between two types of replication studies: pure and scientific replication. In a pure replication (often also referred to as narrow replication), authors of a replication study use the same data and the same methods as the authors of the replicated article.[8] By contrast, authors of scientific replications use

[6] See also Mueller-Langer and Andreoli-Versbach (2018) for a theoretical analysis of the welfare effects of universal mandatory data disclosure.

[7] See also Kuhn (1970) and Wible (1991).

[8] For example, Zhang and Ortmann (2014, p. 415) provide a "replication of Engel's (2011) study using his data and statistical methods."

different data and, possibly, different methods than the authors of the replicated articles. Following the taxonomy provided by Baltagi (2003) and the ReplicationWiki,[9] we consider three subtypes of scientific replications: (1) wide replications using different data but the same methods and models as the replicated article,[10] (2) replications using the same data but new methods and models,[11] and (3) replications using new data and new methods and models.[12] Hence, the extent of replication studies examined in the present study ranges from the extremes of pure replication on the one hand and scientific replication using new data and new methods on the other. Arguably, the different types of replication studies (scientific or pure) and the different purposes of replication, i.e., creation of new knowledge or mere checking of results, are relevant with respect to the question of where replication studies should be published or posted. For instance, one may argue that scientific replications should undergo the same quality assessment as regular journal articles. Therefore, scientific replications should be published in the regular issues or replication sections of peer-reviewed journals (Baltagi, 2003). By contrast, sites such as RePEc, SSRN, arXiv.org, or ReplicationWiki appear to be suitable outlets for pure replications in which the mere checking of results is at the core of the replication effort. Replication journals such as the recently established *International Journal for Re-Views in Empirical Economics* appear to be suitable outlets for all types of replication studies.[13] Finally, we consider the replication result, i.e., we examine whether replication studies support or contradict prior published findings.[14]

3.3. Optimal amount of published formal replication studies

In this subsection, we discuss the optimal amount of published formal replication studies in order to guide our empirical analysis. First, we put the subject under study, i.e., published formal replication studies, in the context of the economics of science. Second, we distinguish between informal and formal replication. Third, we discuss other mechanisms that help ensure the veracity of published scientific results.

3.3.1. Published formal replication studies in the context of the economics of science

Review and publication are the most important filter mechanisms in science. Our approach allows us to explore replication studies that are subject to these filter mechanisms, i.e., formal published replication studies.[15] The empirical analysis presented in this paper is closely related to a recent strand of theoretical economics literature exploring the incentives for researchers to conduct high-quality empirical work and the incentives for their peers to verify the reliability of previously published findings (Lacetera and Zirulia, 2011; Kiri et al., 2018).[16]

Notably, the observed number of published formal replication studies is the result of an equilibrium outcome of a series of incentives for researchers, editors, and potential replicators, i.e., readers.[17] To give an example, the incentives for a researcher to conduct a replication study will depend on the probability of detecting an error or fraud in previously published findings (Dewald et al., 1986; Hamermesh, 2007), the impact of the article under scrutiny (Sukhtankar, 2017), the cost of replication (Lacetera and Ziruli, 2011; Kiri et al., 2018), and the willingness of editors to publish replication studies (especially those confirming a published result). The replicator's probability of finding something interesting will depend on the effort that the creator of original research has invested in careful empirical work. This effort will also depend on the probability that a published article will eventually be replicated. Finally, the cost of replication will depend on the editors' decisions whether or not to adopt data disclosure policies for published papers.

Prior theoretical literature aims at disentangling these potentially countervailing effects. For instance, in a game-theoretical model of the research and publication process, Lacetera and Zirulia (2011) explore the incentives for a researcher (henceforth, researcher) to commit scientific fraud and of another researcher (henceforth, replicator) to conduct a replication. In their model, the researcher decides to either engage in "radical" research that may result in a major scientific contribution or in "incremental" research that may result in a minor contribution to existing knowledge. Lacetera and Zirulia (2011) assume that the effort costs of the researcher are higher for major scientific contributions than for minor contributions to existing knowledge. They also assume that the replication costs of the replicator are higher for major contributions than for minor contributions. Their game-theoretical approach allows Lacetera and Zirulia (2011) to analyze the effect of policies aiming to facilitate replication, i.e., the extent to which such policies reduce the cost of replication through data disclosure policies, as well as their impact on incentives to commit scientific fraud and the probability of undetected fraud.[18]

On the basis of the theoretical predictions of their model, Lacetera and Zirulia (2011) offer two important insights for empirical research regarding the optimal amount of published formal replication studies: first, the effect of the cost of replication on the carefulness of researchers, and second, criteria for suitable data samples.

With respect to replication costs, our data allows us to disentangle two potential drivers of published formal replication studies. Regarding the first driver, Lacetera and Zirulia (2011) propose that replication costs are higher for major novel results than for minor contributions to the literature. Arguably, the better the rank of the journal in which an article is published, the higher the probability that this article contains major novel results. This argument may be particularly valid for articles published in top five economics journals for the following reasons. First, publication of major novel results is one of the core objectives of top five journals because major results are—ceteris paribus—more likely to attract citations. Second, due to the fierce competition for publication in top five journals, editors can be more selective in choosing articles for publication from a large pool of good submissions. Third, top five journals are more likely to attract higher quality referees than other journals because refereeing for top five journals generates larger visibility and reputation effects for referees. By attracting higher quality referees whose main task is to assess the quality and novelty of

[9] http://replication.uni-goettingen.de/wiki/index.php/Replication_in_economics (last accessed 21 August 2017)

[10] For example, see Taylor et al.'s (2010) wide replication of Hastings (2004).

[11] For example, Rock et al. (2000) use the same data as the respective replicated article (Bhushan, 1989), but use count-data econometrics instead of OLS.

[12] See Bali et al.'s (2005) replication of Goyal and Santa-Clara (2003) where they use an extended sample period and provide alternative measures of idiosyncratic risk.

[13] See https://www.zbw.eu/en/about-us/key-activities/research-data-management/iree/ (last accessed 14 September 2017).

[14] See also Section 4.1, where we provide detailed examples of different replication results.

[15] Our approach does not allow us to make statements about informal replications that have not been published in journals. We discuss this aspect in more detail in Section 6.

[16] See also Mirowski and Sklivas (1991), Wible (1991), and Feigenbaum and Levy (1993) for early theoretical analyses of replication, data disclosure, and the cost of replication.

[17] We thank an anonymous referee for this comment.

[18] Lacetera and Zirulia (2011) suggest that the "intuitive" result, i.e., a reduction of replication cost reduces the probability of undetected fraud, is only obtained if the researcher does not change the type of research. Notably, a reduction in replication cost does not necessarily result in a lower probability of undetected fraud. The underlying driver of this counterintuitive result is that a reduction in the replication cost may change the overall nature of the research and publication game as follows: A reduction in the replication costs induces the researcher to shift from incremental to radical research, which is more costly for the replicator to replicate.

submissions, editors of top five journals achieve a higher probability of publication of major novel results than editors of other journals. In our data, we take these aspects into account by including a binary variable for publication in a top five economics journal.[19]

Regarding the second potential driver of published formal replication studies, Lacetera and Zirulia (2011) argue that empirical studies based on detected scientific misconduct only, i.e., studies based on cases of negative replication or retractions only, might suffer from bias in the sense that high-impact articles that generate more interest are also more likely to be subject to scrutiny from their peers and therefore might have a higher probability of retraction or negative replication. In particular, Lacetera and Zirulia (2011) suggest that a suitable data sample should also potentially include correct works, i.e., non-fraudulent works, in addition to flawed works in order to mitigate concerns related to the abovementioned bias. Our data sample satisfies this criterion because we consider both positively as well as negatively replicated articles in addition to nonreplicated articles while controlling for article impact (see Section 4).

Notably, prior game-theoretical works exploring high-quality scientific production and replication also offer important insights with respect to the optimal amount of replication (Lacetera and Zirulia, 2011; Kiri et al., 2018). In particular, the optimal amount of published formal replication studies will strike a balance between the benefits and costs of replication while taking into account the incentives associated with science.

For instance, in terms of benefits of replication, Lacetera and Zirulia (2011) address the question of whether an increase in the recognition of replication studies may reduce the incentives for scientific misconduct. Lacetera and Zirulia (2011, p. 589f) argue that the "limited recognition for replication works that characterizes the scientific community (…) can also be seen as a limit to the pressure to publish and compete with other scientists in a given field." In addition, they argue that a researcher making a radical contribution in any given field obtains a quasi-monopoly in that field as long as no attempts at replication are undertaken. Consequently, incentives for other scientists to engage in research in the same field may be reduced. Based on these arguments, the authors suggest that establishing higher incentives for replication may help to deter scientific fraud. However, an increase in the incentives for replication in order to make published works more reliable, e.g., the adoption of data disclosure policies, may come at the cost of slowing down the creation of novel scientific results (Lacetera and Zirulia, 2011; Mueller-Langer and Andreoli-Versbach, 2018).[20] Kiri et al. (2018) come to the same conclusion in their theoretical analysis of high-quality scientific production in which the correctness of a new published result is endogenous to effort for careful empirical work. In addition, Kiri et al. (2018) suggest that the researcher's and replicator's private incentives and their social incentives are not necessarily well aligned. In particular, policies aimed at reducing the cost of replication, such as data disclosure policies, may have a negative impact on a researcher's incentives to start a new, potentially socially valuable project. Based on this theoretical prediction, Kiri et al. (2018) argue that the potentially countervailing effects of a reduction of the costs of replication may be taken into consideration when designing optimal replication policies.

3.3.2. Informal vs. formal replications

In terms of the benefits of replication, one may distinguish between (a) false results that are removed ex post from the cumulative body of knowledge and (b) higher incentives for careful empirical work ex ante (Kiri et al., 2018; Lacetera and Zirulia, 2011). In this respect, it is important to distinguish between formal replication, i.e., formal published replication studies, and informal replication, i.e., replication in working papers or replication exercises in econometrics classes.

In terms of quality and reliability, note that the result of a replication is not necessarily correct and may, in turn, be subject to debate.[21] Arguably, replicators might be biased toward finding results that contradict previously published findings (Dewald et al., 1986; Hamermesh, 2007). Therefore, one may argue that the most important filter mechanisms in science, i.e., review and publication, should not only apply to original research but also to replication studies. By contrast, one may also argue that the time lag between the publication of an article in a journal and a potentially negating replication study should not be too long so that false knowledge can be removed from the cumulative body of knowledge in due time. In order to take this aspect into account, we control for the time lag between the publication of an article in a journal and publication of the respective replication study in our empirical analysis.

In terms of visibility of replication, note that publication of a replication study in a journal is likely to increase the visibility of the replication. For instance, Coffman et al. (2017, p. 44) claim that "replication work (…) needs the visibility afforded by top journals." This aspect is particularly important in light of recent empirical findings that authors are typically very unsure about whether or not a paper has been subject to a replication study (Berry et al., 2017). Arguably, an increase in the visibility associated with publication in a top journal will increase the probability that the abovementioned benefits of replication are achieved. For instance, readers will not be aware of a negative replication if it is only part of a classroom exercise. In addition, replication work presented in a working paper will be less visible for the readers, including the replicated authors, than replication work presented in a formal published replication study. Notably, recent initiatives to collect and spread information on replication studies, such as the ReplicationWiki or the Replication Network, may help to mitigate some of these concerns in terms of visibility.

However, regarding the ex post effects of formal published replication studies, note that actual replication has to take place on a case-by-case basis, i.e., a single negating replication study typically negates the results of a single article. By contrast, regarding the ex ante effect of replication, note that a credible threat of replication could have an impact on a larger scale. The ex ante effect of replication is particularly important because, as pointed out by Lacetera and Zirulia (2011), neither the editors nor the referees of journals are expected to check for misconduct.

A sufficient level of formal published replication studies is required for replications to constitute a credible threat. For instance, consider a researcher who channels her efforts into a new empirical research project. Thereby, she faces the trade-off between the private benefits and costs of careful empirical work (Lacetera and Zirulia, 2011; Kiri et al., 2018). For instance, assume that the private benefits of carefulness are given by citations of the published article. Following Feigenbaum and Levy (1993, p. 219), in terms of citation streams, "an original article can die a violent death upon the publication of a negative replication." It is in this respect that the researcher's cost resulting from the publication of a negative replication study can be interpreted in terms of foregone citations.[22] Arguably, under these conditions there is a minimum level of expected foregone citation streams due to the publication of a negative replication that is needed for negative replication to constitute a credible threat.[23]

[19] We also include the impact factor of journals in which replicated and nonreplicated articles are published.

[20] Lacetera and Zirulia (2011) also suggest that replication is sensitive to a free rider problem because replication costs are borne by the individual, whereas the benefits of replication are social.

[21] See, for instance, Gilbert et al. (2016)'s critical assessment of a replication study in psychology (Open Science Collaboration, 2015).

[22] This provides an explanation for the observation that the authors of original work frequently perceive replication attempts as hostile (Mirowski and Sklivas, 1991; Maniadis et al., 2015).

[23] To illustrate this point, consider the following two extreme cases. First, assume that expected foregone citations equal zero because the incidence of negative replication equals zero. In this case, a formal published replication study does not constitute a credible threat. Second, assume that the probability

3.3.3. Other mechanisms to ensure the veracity of published scientific results

Finally, it is important to note that there are other mechanisms than formal published replication studies that may help to ensure the veracity of published scientific results. First, in pre-analysis plans, researchers specify and register a fixed plan of how to collect and analyze the data before a project begins (Coffman and Niederle, 2015). While a pre-analysis plan aims at minimizing the researcher's flexibility over the econometric approaches and specifications, it may fail to solve the problem of false positives as indicated by recent evidence on discontinuities around *p*-values of 5%, for instance (Brodeur et al., 2016; Coffman et al., 2017). Due to the large number of publications and researchers' bias toward finishing and publishing positive findings, the problem of false positives is likely to persist (Coffman and Niederle, 2015; Franco et al., 2014; Ioannidis, 2005). In light of these findings, Coffman et al. (2017, p. 41) suggest that "a solution is required to try and separate what is true from what is not after a paper is disseminated: replication."

Second, hypothesis registries are typically publicly available databases in which all attempted projects, i.e., published and unpublished works for a given class of hypotheses, are listed. These registries aim at mitigating the problem that statistically insignificant results remain unpublished. However, Coffman and Niederle (2015, p. 90) suggest that hypothesis registries "would not necessarily increase the probability that a published result is true." Notably, Coffman and Niederle (2015) stress the important function replications serve as an effective mechanism to identify false positives in their comparative analysis of pre-analysis plans, hypothesis registries, and replications.

Third, meta-analysis is another mechanism alongside replication that serves the goal of research synthesis (Anderson and Kichkha, 2017; Cooper and Hedges, 2009; Cooper, 2017). Articles that adopt meta-analysis techniques typically draw conclusions by combining the results from a set of empirical works. Comparing replications and meta-analysis, Anderson and Kichka (2017, p. 58) challenge the statistical consistency of meta-estimates on the grounds that (1) "meta-analysis delivers correlation," (2) "significance of variables in meta-analysis does not signal important directions for further research," and thus (3) "meta-analytic explanatory variables will become endogenous" when researchers derive directions for future research from meta-analytical studies.

4. Data and variables

The period under study is 1974 to 2014. We obtain our sample from the population of all articles published in the top 50 economics journals (in total, 126,505 articles). From these 126,505 articles we identified 130 published replication studies, i.e., 0.10% were replication studies. Following Duvendack et al. (2015), we categorized a published journal article as a replication study if its main purpose is to test the reliability of a previously published study.[24] Consequently, this definition does not include original research articles in which indirect replication, e.g., an initial "sanity check" to test the validity of a data set, either does not occur at all or *does* occur but is not the main purpose of an article. In particular, while the narrow replications under study are typically entirely devoted to replication, the replication works in the scientific replications under study are embedded in a broader analysis. However, the testing of a previously published result always plays the main role in all replication studies under study.

Having identified 130 published replication studies, which is consistent with prior research on the frequency of published replication studies in economics,[25] we then identified the respective journal articles (henceforth, replicated articles) that were eventually replicated. In order to study differences between replicated and nonreplicated articles, we selected all articles published in top 50 economics journals that contained at least one replicated article. Our sample thus consists of 1243 articles, 130 of which were replicated. Including nonreplicated articles from the same journal issues allows us to control for latent effects at the level of journal issues while keeping the data collection effort manageable. Note that our approach only covers instances of formal replications, i.e., replication studies published in journals. Informal replication studies that are not published (e.g., replication efforts that are conducted in teaching or are part of an initial step in a research project) are not covered (see also Section 6.1).

4.1. Data and sample creation

We followed two distinct strategies in order to identify replication studies. First, we considered Web of Science (WoS) metadata for all articles published in the top 50 economics journals (126,505 articles). We retrieved information on the titles and abstracts of the 126,505 articles under study from WoS. We used this information to count, for each article, how often indicative key (stem) words such as "repli*," "reexamin*," "comment," "revisit," "retesting," or "reappraisal" (among others), as well as references to other articles, appear in its title and abstract.[26] Both the frequency and the location of these key words allowed us to determine the likelihood of the article being a replication study for all articles under study. To illustrate this point, an article that contains "repli*," "retesting," or "note" in the title and contains a reference to another article in the abstract—indicated by "(19*)" or "(20*)"—has a high likelihood of being a replication study.[27] We ranked all articles in terms of their likelihood of being a replication study.[28] We

(*footnote continued*)
of publication of a negative replication equals one. This leads to the maximum level of expected foregone citations. In fact, if the negative replication is immediately published, i.e. it is published before the original article receives its first citation, a citation-maximizing researcher may not have an incentive to conduct empirical research in the first place. Formal negative replication constitutes a credible threat in this case. Arguably, the minimum level of expected foregone citation streams due to the publication of a negative replication study needed to achieve a credible threat will lie between these two extremes.

[24] See also Section 3.2, where we define replication type (pure or scientific), and Section 4.1, where we define replication result (negating or reinforcing).

[25] Notably, Duvendack et al. (2015) also identified published replication studies in economics. They used (a) keyword searches in Google Scholar and in all 333 economics journals listed in WoS, (b) entries in the ReplicationWiki, (c) suggestions from editors and (d) their own collections of replication studies. They then performed a more systematic search within the top 50 economics journals. Overall, Duvendack et al. (2015) found 162 published replication studies in the economics literature. The number of replication studies in our sample (130) is lower than the one in Duvendack et al. (2015) for the following reason: In our sample, the replication studies and respective replicated articles are both published in the top 50 economics journals according to WoS impact factors. By contrast, Duvendack et al. (2015) also consider replication studies published in the 283 WoS-listed economics journals that have a lower impact factor.

[26] Initially, we used a vector of 100 key words, which was finally reduced to a vector of 70 stem words to avoid double entries. For instance, we used the stem word "repli*" for both keywords "replication" and "replicable" or "reexamin*" for "reexamination" and "reexamine," respectively.

[27] However, our filtering algorithm also identifies comments and other publications that only use adjectives (and their stem words) such as "original" or verbs like "re-examine" in the abstract.

[28] We used the information obtained from the previous counting exercise to determine this ranking by journal as follows: First, we counted how often the 70 key stem words appear in the title and abstract of any given article in any given journal. Second, for any given journal, we ordered the articles according to the accumulated appearance of these key stem words. To illustrate this point, we consider the results of the ranking for the *American Economic Review*. Iversen and Söderström (2014) is ranked 2nd in terms of probability of being a replication study. The word "comment" appears in the title of the paper. In addition, stem words such as "comment," "correct*," or "(20*," appear in its abstract. By contrast, in Sheshinksi (1971), which is ranked 100th in terms of being a replication study, only the key word "note" appears in the title.

studied the 100 highest ranking articles in each journal in detail in order to identify replication studies. In addition, we included all eligible replication studies published on the website of ReplicationWiki[29] in our data set. We then identified the respective journal article that was eventually replicated.[30]

For the analysis, we only considered empirical research articles and removed purely theoretical articles. We defined a purely theoretical article as an article that does not use any data. We searched for summary statistics and statistical tables in the PDFs in order to distinguish between empirical and purely theoretical articles.

We studied all replication studies to determine whether they either negate or (partially) reinforce the replicated article. Prominent examples of the former—in two separate authorship teams—are Foote and Goetz's (2008) and Joyce's (2009) non-supportive replications of Donohue and Levitt's (2001) results on the association between the abortion rate and age-specific crime rates. Foote and Goetz (2008) identify a coding error in the concluding regressions that has a substantial impact on the main results obtained by Donohue and Levitt (2001). Foote and Goetz (2008, p. 407) acknowledge that the "paper could not have been written without the assistance of John Donohue and Steven Levitt, who made their original data and programs available on the Internet and who supplied us with their new data and programs as soon as they became available." This highlights the potential value of the disclosure of data and code to replication, particularly for high-profile results. In addition, Joyce (2009, p. 112) provides empirical evidence "that there is no association between abortion and age-specific homicide rates or age-specific arrest rates for murder."[31] In another example, Dyl and Maberly (1986, p. 1149) negate Cornell's (1985) results on the weekly pattern of stock returns indicating the following: "(…) In an attempt to resolve this puzzling inconsistency, we replicated Cornell's study for the period from May 3, 1982, through July 24, 1984. We were unable to duplicate his results." In a more recent example, Abrevaya and Puzzello (2012) negate Adda and Cornaglia's (2006) empirical findings on the compensatory behaviour of smokers when cigarette taxes increase. Abrevaya and Puzzello (2012, p. 1760) conclude that "(i)n this comment, we have re-examined this claim by AC and find little systematic evidence to support it." To give an example for a reinforcing replication study, consider Gerdtham et al. (1999, p. 117), who aim "to validate Wagstaffs and van Doorslaer's approach of constructing a continuous health measure" and whose "results (…) support the validity of the WvD method."[32] In addition, Hung and Plott (2008, p. 1518) reinforce Anderson and Holt's (1997) results on information cascades in the laboratory and conclude the following: "The results of Anderson and Holt replicate (Result 1). In our experiments we observe the phenomena they report."

Among the top 50 economics journals (according to the WoS impact factor), 23 had published at least one replicated article in the past.[33] We retrieved article metadata from WoS, i.e., publication date, number of references, pages and authors, and journal information. We gathered information on the rank position of the institutions that the authors are affiliated with from the Ranking Web of Universities 2014 and obtained author citation metrics from Scopus. Following Andreoli-Versbach and Mueller-Langer (2014), we identified which journals have data disclosure policies as well as the first volume in which the policy was adopted in order to identify articles that are subject to a data disclosure policy.

Funding bodies may require authors to make data or program code available. Therefore, we also analyzed the funding guidelines of 36 research funding bodies across the globe regarding their data management policies. To this end, we randomly selected 27 public funding bodies from the 15 countries with the highest public expenditure for research according to the OECD (2016). We added nine funding bodies that are not necessarily public but that support noncommercial research (e.g., Bill and Melinda Gates Foundation), that are international (e.g., the European Commission's Horizon 2020), or that are from countries that are too small to appear in the list of top funding countries (e.g., the Swiss National Science Foundation). In addition to screening the websites manually for guidelines, we contacted every funding body in our sample via email in two waves and asked for relevant documents. Twenty funding bodies replied to our information request. We were able to gather the guidelines from all 36 funding bodies. We treated the guidelines as textual data and code if they mention data management, replication studies, and if they specify where and how data should be stored, its terms of access (e.g., on request or public), and documentation standards. From the 36 guidelines under study, 22 (61%) mention data sharing; 19 (53%) specify how or where to publish data;[34] 20 (56%) mention data documentation; 16 (44%) require data management plans; and 20 (56%) mention an embargo period for data in which the principal investigator has exclusive rights to access the data and use it in empirical articles. Notably, none of the funding bodies under study has an explicit replication policy.

4.2. Variables

Table 1 provides an overview of the dependent and independent variables used in our study.[35]

4.2.1. Dependent variable

The dependent variable, *ReplicatedArticle*, is a binary variable indicating whether the article under study was eventually replicated. This variable measures the joint likelihood of a replication being undertaken and then being published. Merely analyzing the incidence of replication studies being undertaken (measurable by replications described in discussion papers or other prepublication media) would not capture the most important filter mechanism in academic communication: review and ultimate publication. We obtained *ReplicatedArticle* by identifying replication studies published in the top 50 economics journals and the respective replicated articles published in these journals.

Fig. 1 shows that the total number of journal articles increased at a slightly higher rate than the total number of published replication studies. It is also noteworthy that the share of published replication studies from the total number of journal articles per year never exceeds 0.26% in the period between 1974 and 2014. From our sample, we also estimate that the share of empirical articles increased from about 73% in 1975 to about 80% in 2010 (Fig. 2).

Replication studies may negate or (partially) reinforce the results of the replicated article. 61 of the 130 replication studies under study are negating (partially reinforcing: 69). In addition, 102 of the 130 replication studies under study, i.e., 78.5%, are published in the same journal as the respective replicated articles. 52.9% of the 102 replicated articles with replications in the same journal are positively replicated (negatively replicated: 47.1%). The impact factors of journals publishing positive and negative same-journal replications are, on average, very similar (positive: 3.2; negative: 3.3). In terms of positive versus negative replication, a similar picture emerges for the 28 replicated

[29] http://replication.uni-goettingen.de/wiki/index.php/Main_Page (last accessed 30 January 2017)

[30] We provide an overview of the replication studies and respective replicated articles in the supplementary materials.

[31] See also Joyce (2004), Donohue and Levitt (2004) and Fryer et al. (2013).

[32] See Wagstaff and van Doorslaer (1994).

[33] Appendix 1 provides on overview of the journals under study, year of enactment of mandatory data disclosure policies, and number of publications.

[34] For instance, 12 guidelines (33%) indicate that data from funded projects should be stored in a public repository, while 2 (6%) mention that data needs to be made available upon request.

[35] Appendix 2 provides a correlation matrix for the dependent variable and main variables of interest.

Table 1
Descriptive statistics.

	mean	sd	min	max	N
Dependent variable					
Replicated article	0.105		0	1	1243
Negated replicated article	0.049		0	1	1243
(Partially) reinforced replicated article	0.056		0	1	1243
Main variables of interest					
Total citations before publication of replication	20.89	64.18	0	1508	1243
Lag between publication of replicated article and replication	4.851	3.601	0	23	1243
Journal impact factor	3.516	1.153	2.137	6.033	1243
Top five economics journal	0.512		0	1	1243
Top 50 university	0.606		0	1	1243
Mandatory data disclosure policy	0.230		0	1	1243
Data or program code available	0.169		0	1	1243
Control variables					
Self-created data	0.124		0	1	1243
Confidential or proprietary data	0.012		0	1	1243
Article published in conference proceedings	0.118		0	1	1243
Number of references	29.28	17.83	0	130	1243
Number of pages	19.51	10.94	1	65	1243
Number of authors	2.057	1.097	1	16	1243
h-index of best author	17.42	12.90	0	106	1225
Third-party funding	0.185		0	1	1243
Funder's support for data availability	0.598	1.366	0	5	1243

Notes: In the regressions, we also include dummy variables for publication year intervals, i.e., 1970–1974, 1975–1979, …, 2010–2014, and journal dummy variables (variables not reported in Table 1). Articles under study fall into one of the following Business & Economics research areas according to WoS: General (76.4%), Mathematical Methods (10.2%), Environmental Sciences (3.6%), Health Care Sciences (0.9%) or Transportation (0.01%). WoS assigns the same research area to all articles in any given journal. By including binary journal variables in the regressions, we therefore control for WoS research areas.

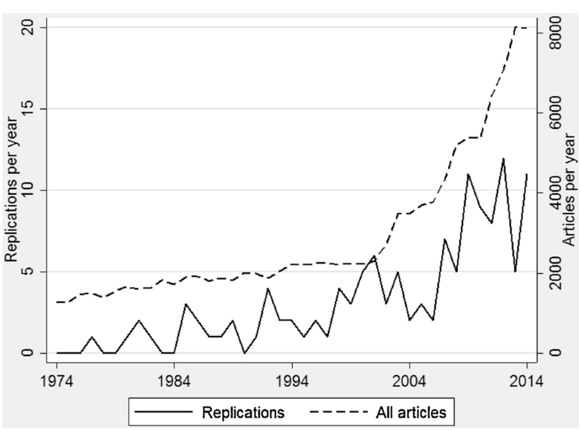


Fig. 1. Total number of published replication studies by publication year (as indicated by the vertical axis on the left-hand side) and of all journal articles by publication year (vertical axis on the right-hand side). Fig. 1 suggests that the total number of journal articles increased at a slightly higher rate than the total number of published replication studies. Sample: All articles published in the top 50 economics journals (according to WoS) between 1974 and 2014 (including 130 published replication studies).

articles of which a replication study is published in a different journal (positively replicated: 53.6%; negatively replicated: 46.4%). We then considered what fraction of these 28 different-journal replication studies is published in higher ranking (and lower ranking) journals than the respective replicated article. Of the different-journal replications, 32.1% are published in higher ranking journals and 67.9% in lower ranking journals, respectively.

Finally, we distinguished between two main types of replication studies and replicated articles. First, narrowly (or purely) replicated articles are eventually replicated using the same data and code. 20 articles under study are narrowly replicated articles. Second, 107 articles under study are scientifically replicated articles. This type of replicated article comprises widely replicated articles (new data but same method: 29 articles), articles that are replicated using the same data but new methods (48 articles), and articles that are replicated using new methods and new data (30 articles).[36]

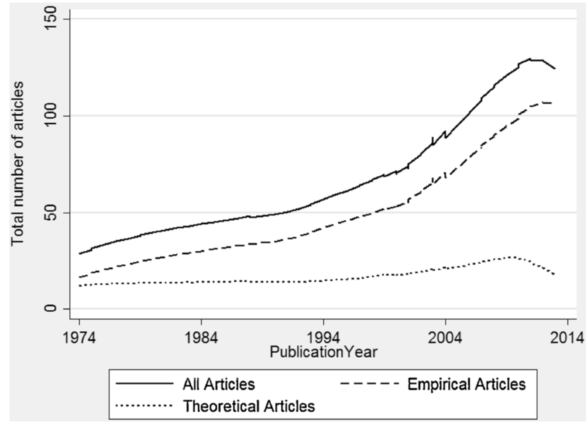


Fig. 2. Total number of empirical and purely theoretical articles by publication year. Sample: All empirical and purely theoretical articles in issues of top 50 economics journals that contained at least one replicated article. Recall that purely theoretical articles are excluded from our main sample of 1243 empirical articles that we use in the regression analysis. Here, we refrain from excluding 445 purely theoretical articles. We use the lowess command in STATA, which performs locally weighted scatterplot smoothing (Cleveland, 1979). We refrain from displaying the three scatterplots to improve the readability of the figure.

4.2.2. Independent variables

We distinguished between the main variables of interest and control variables by indicating article, author, journal, and institutional characteristics (see Table 1).

Regarding the main variables of interest, *CitesPreReplication* indicates the total number of citations of replicated and nonreplicated articles in the same issue one year prior to the publication of a replication study. To illustrate this point, consider the replication study by Couch and Placzek (2010) of Jacobson et al. (1993). Total citations one year before the publication of this replication study are given by the cumulative citations generated from 1993 to 2009 (241 in this case). Arguably, replicated articles may attract more citations prior to the publication of a replication study if there is a longer time lag between the publication of the article and the publication of the respective replication study. *LagReplication* measures this time lag in years.

Table 2 provides extended descriptive statistics on *CitesPreReplication* and *LagReplication* by subgroup while distinguishing between replicated and nonreplicated articles. The descriptive statistics indicate that prior to replication, replicated articles attracted more citations than nonreplicated articles for virtually all subgroups.[37] Replicated and nonreplicated articles published in the top five journals attracted more total citations (by a factor of four) than articles published in other top 50 journals. In addition, the time lag between the publication of an article and the respective replication study is significantly longer for top

[36] We refrained from coding three particular replication studies (Salas and Raftery, 2001; Lee, 2008; Fraas and Lutter, 2012) as either pure or scientific because they examine the econometric approaches adopted in the respective replicated articles theoretically without using either the same or new data as the replicated articles.

[37] The two only subgroups in which nonreplicated articles attracted slightly more citations than replicated articles are *SelfCreatedData* = 1 and *Pyear80* = 1.

Table 2
Extended descriptive statistics.

	Replicated articles		Nonreplicated articles	
	CitesPreReplication mean	*LagReplication* mean	*CitesPreReplication* Mean	*LagReplication* mean
Main variables of interest				
Top5Journal = 1	100.57	6.71	24.51	6.10
Top5Journal = 0	25.23	3.64	7.40	3.45
Top50University = 1	73.07	5.04	19.67	5.07
Top50University = 0	26.69	5.17	11.32	4.47
MandatoryDisclosure = 1	35.22	3.89	12.06	4.04
MandatoryDisclosure = 0	64.66	5.26	17.58	5.07
DataOrCode = 1	29.00	4.26	20.15	3.87
DataOrCode = 0	65.99	5.22	15.45	5.02
Control variables				
SelfCreatedData = 1	13.00	3.93	13.65	4.42
SelfCreatedData = 0	66.79	5.23	16.63	4.88
ProceedingsArticle = 1	27.40	2.60	6.92	4.33
ProceedingsArticle = 0	61.91	5.18	17.62	4.90
Funded = 1	118.23	5.67	23.66	5.06
Funded = 0	43.29	4.90	14.63	4.77
Year variables				
Pyear1970 = 1	78.00	8.00	12.43	8.00
Pyear1975 = 1	3.25	3.50	3.02	3.98
Pyear1980 = 1	10.29	6.57	11.25	5.04
Pyear1985 = 1	71.62	5.85	10.73	4.48
Pyear1990 = 1	49.40	7.13	22.25	8.32
Pyear1995 = 1	106.57	5.71	21.11	5.75
Pyear2000 = 1	99.48	5.28	30.04	5.57
Pyear2005 = 1	24.54	3.96	14.39	3.90
Pyear2010 = 1	19.25	2.08	5.08	3.17

five journals (6.71 years) than for other top 50 journals (3.64 years). Replicated and nonreplicated articles by authors affiliated with a top 50 university attracted more citations on average than articles by authors affiliated with lower ranked universities. While we obtain similar results for articles that received external research funding, the opposite is true for articles published in conference proceedings.

The quality of a journal is given by the impact factor recorded by WoS in 2014, *ImpactFactor*. We include a dummy variable, *Top5Journal*, identifying articles published in one of the top five economics journals according to Card and DellaVigna (2013), i.e., *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. For single-authored papers, *Top50University* indicates that the author is affiliated with a top 50 university according to the Ranking Web of Universities 2014. For papers with multiple authors, it indicates that the author from the highest ranked university of all authors is affiliated with a top 50 university. *MandatoryDisclosure* is a binary variable indicating whether an article is subject to a mandatory data disclosure policy. It is equal to zero for articles published in journals without mandatory data disclosure. It is also equal to zero for articles published in journals with a data disclosure policy if the article was published in a volume before the policy was enacted. It equals one for articles published in the volume of enactment and all subsequent volumes. *DataOrCode* is a dummy variable indicating whether the data or program code used in an article are available on the journal website. For instance, journals with mandatory data disclosure policies typically provide a link to the data and program code together with the PDF of the article.[38]

As for the control variables, *SelfCreatedData* is a dummy variable indicating whether the data used in an article is self-created (in contrast to archived data, which is re-analyzed), e.g., via laboratory or field experiments, surveys, or interviews. In order to make this classification, we read all articles in detail. Publications that exclusively use publicly available data, e.g., census data, did not qualify as articles using self-created data. We read all explanatory notes in order to determine whether the data used in an article was confidential or proprietary to generate the dummy variable *ConfidentialData*. *ProceedingsArticle* is a binary variable indicating whether articles were published in conference proceedings. *References* and *Pages* are defined as the total number of references and pages. *Authors* indicates the number of authors. To control for author quality, we created the variable *BestH*, which indicates the h-index of authors of single-authored articles or the highest h-index of all co-authors in the case of multi-authored articles, respectively. *Funded* is a dummy variable indicating whether an article received third-party funding.[39] *FunderDataSupport* is a variable ranging from 0 to 5 that indicates the number of data policies and data management tools that external research funders provide to the authors they support. This variable can be thought of as the extent to which external funders have policies that facilitate data availability.[40] We also created binary variables indicating the time of publication in five-year intervals (henceforth, year dummies). Finally, according to WoS, articles under study fall into one of the following Business & Economics research areas (not reported in Table 1): General (76.4%), Mathematical Methods (10.2%), Environmental Sciences (3.6%), Health Care Sciences (0.9%) or Transportation (0.01%). Notably, WoS assigns the same research area to all articles in any given journal. By including

[38] Journals with such a policy typically follow the mandatory data disclosure policy of the *American Economic Review*, which states: "Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. These will be posted on the AER Web site."

[39] We retrieved this information from the acknowledgment sections of the articles under study and from WoS article metadata. In our sample, the five most frequent funding bodies in terms of articles funded are the National Science Foundation (12.6% of all articles), Economic and Social Research Council (1.9%), Social Sciences and Humanities Research Council of Canada (1.5%), Medical Research Council (0.5%), and Leverhulme Trust (0.5%).

[40] Appendix 3 provides an overview of the data guidelines of the funding agencies under study.

binary journal variables in the regressions, we therefore control for WoS research areas.

5. Empirical results

We begin by reporting descriptive results on the prevalence and enforcement of mandatory data disclosure policies of journals. We then perform probit regressions to examine the article-level correlates of the probability that a journal article is eventually replicated.

5.1. Enforcement of mandatory data disclosure policies

In our sample, nine of the 23 journals under study have a mandatory data disclosure policy (see Appendix 1). Notably, two explicit replication policies (*JPE* and *Labour Economics*) were suspended, according to the editors, due to a lack of interest in replication studies (Hamermesh, 2007). Moreover, 286 of the empirical articles under study are subject to a mandatory data disclosure policy, i.e., they were published after the data policy was introduced. Notably, the data sets or program codes used in articles are available on the respective journal website for 183 of these articles (63.9%). Out of the 286 articles that are subject to a mandatory data disclosure policy, we identified one article for which the data or program code was not available because it was proprietary (which leads to an exemption from disclosure). This suggests that for 35.7% (102) of the 286 empirical articles subject to mandatory data disclosure policies, the data or program code was not available even though the data was not proprietary. This result raises concerns regarding the enforcement of mandatory data disclosure policies. Fig. 3 illustrates the total number of articles published under a mandatory data disclosure policy and the total number of articles published under a mandatory data disclosure policy that is not strictly enforced. This suggests that both numbers increase over time and that for a large share of journals, mandatory data disclosure policies are announced but not always enforced or monitored.

However, these findings should be interpreted with caution given the relatively small number of observations.

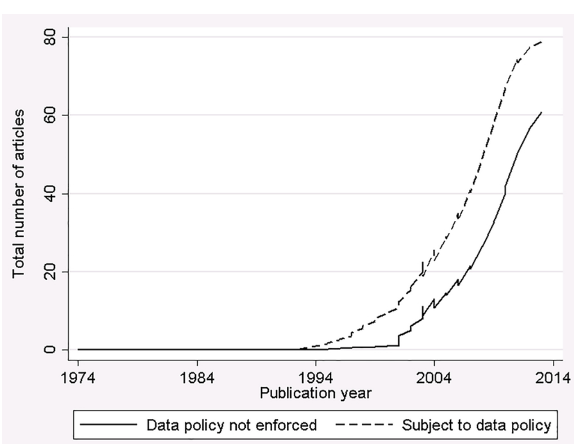


Fig. 3. Prevalence and enforcement of mandatory data disclosure by publication year. The vertical axis shows the total number of articles subject to a mandatory data disclosure policy (indicated by the dotted line) and the total number of articles subject to a mandatory data disclosure policy that is not strictly enforced (indicated by the solid line). Fig. 3 suggests that both numbers increase over time and that for a large share of journals, mandatory data disclosure policies are announced but not always enforced or monitored. We use the *lowess* command in STATA, which performs locally weighted scatterplot smoothing (Cleveland, 1979). We refrain from displaying the two scatterplots to improve the readability of the figure.

5.2. Empirical analysis of replication studies

To examine the article-level correlates of the probability that a journal article is eventually replicated, we run probit (and OLS) regressions as given by: prob(ReplicatedArticle) $= \beta_1 \cdot$ ProceedingsArticle $+ \beta_2 \cdot$ LogReferences $+ \beta_3 \cdot$ LogPages $+ \beta_4 \cdot$ Authors $+ \beta_5 \cdot$ LogBestH $+ \beta_6 \cdot$ SelfCreatedData $+ \beta_7 \cdot$ ConfidentialData $+ \beta_8 \cdot$ Funded $+ \beta_9 \cdot$ FunderDataSupport $+ \beta_{10} \cdot$ LogCitesPreReplication $+ \beta_{11} \cdot$ LogLagReplication $+ \beta_{12} \cdot$ Top5Journal $+ \beta_{13} \cdot$ LogImpactFactor $+ \beta_{14} \cdot$ Top50University $+ \beta_{15} \cdot$ MandatoryDisclosure $+ \beta_{16} \cdot$ DataOrCode.

We also include journal and year dummy variables. Note that author citation metrics are not available for 18 out of 1243 observations. We use log transformations in the regressions. As some researchers have an h-index of zero, we define *LogBestH* = log(*LogBestH* + 1). We follow the same procedure for *LogReferences*. We compute robust standard errors clustered at the journal level. All specifications reported in this paper are straightforward modifications of this baseline specification.

5.2.1. Full sample analysis

We run our regressions with nine different specifications as reported in Table 3. In specification [1], we consider control variables (article, author, and institutional characteristics), journal dummies, and year dummies.[41] Table 3 does not report marginal effects of the control variables (Appendix 4 provides the full version of Table 3). For all specifications, we provide the Wald test statistics from the probit regressions for the control variables, journal dummies, and year dummies at the bottom of the table. We include the log of total citations before publication of a replication study and the log lag between the publication of replicated articles and of the respective replication studies in specification [2] to examine the effect of article impact on the replication probability. In specification [3], we include the top-five-economics-journal dummy variable and the log of the journal impact factor to account for the effect of journal quality on the probability of replication. In specification [4], we add the affiliation with top 50 universities. We include the binary variable for mandatory data disclosure in specification [5] to examine the impact of the existence of mandatory data disclosure policies on the replication probability. In specification [6], we include the binary variable for actual data or code availability. We separately include *MandatoryDisclosure* and *DataOrCode* in columns [5] and [6] because of their high correlation (0.69; see Appendix 2).[42] In column [7], we examine articles published in issues that contain at least one scientifically replicated article.[43] Finally, we examine the subsamples of journal articles published in issues with at least one negated replicated article (reported in column [8]) and at least one (partially) reinforced replicated article (reported in column [9]).

The results reported in Table 3 provide empirical evidence that the impact of journal articles, which is measured in citations, positively affects the probability that they are eventually replicated. If not stated differently, we henceforth report marginal effects at the means. For the binary

[41] In the regressions, we specify all dummy variables as factor variables because the *margins* command in Stata would otherwise treat these binary variables as continuous (Williams, 2012). In addition, we omit the factor variables for the top five journals to avoid collinearity with *Top5Journal*.

[42] The high correlation reflects that authors of articles published in journals with mandatory data disclosure policies are more likely to make their data and code available. Results are qualitatively unchanged when we include both variables in column [5]. In particular, effects of both variables remain insignificant. Our results are also virtually the same when we run the regressions separately for the subsample of articles that fall under the WoS research area "Business & Economics: General".

[43] We refrain from running the regressions separately for the subsample of articles published in issues that contain at least one narrowly replicated article because of the low number of narrowly replicated articles in our sample (20).

Table 3
Marginal effects at the mean after probit regression.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Issues/Sample:	All	All	All	All	All	All	With sci. repl. art.	With neg. repl. art.	With reinf. repl. art.
Dependent variable:	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article	Scient. repl. art.	Negated repl. art.	Reinf. repl art.
Log total citations before publication of replication		0.055***	0.056***	0.053***	0.053***	0.053***	0.053***	0.067***	0.040***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.010)	(0.007)
Log lag between publication of replicated article and replication		−0.106***	−0.106***	−0.100***	−0.101***	−0.101***	−0.105***	−0.142***	−0.073***
		(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.013)	(0.022)	(0.016)
Top 5 economics journal = 1			−0.096***	−0.092***	−0.081**	−0.090***	−0.072***	−0.037	−0.062**
			(0.024)	(0.025)	(0.030)	(0.025)	(0.018)	(0.043)	(0.021)
Log impact factor			−0.040	−0.036	−0.042	−0.038	−0.072**	−0.008	−0.071
			(0.022)	(0.023)	(0.027)	(0.026)	(0.025)	(0.040)	(0.037)
Top 50 university = 1				0.038	0.038	0.038	0.044*	0.019	0.045*
				(0.020)	(0.020)	(0.020)	(0.019)	(0.027)	(0.021)
Mandatory data-disclosure policy = 1					−0.011				
					(0.016)				
Data or program code available = 1						−0.002	−0.016	−0.014	0.003
						(0.012)	(0.017)	(0.025)	(0.008)
Observations	1225	1225	1225	1225	1225	1225	973	563	714
Pseudo R-squared	0.0723	0.111	0.113	0.120	0.120	0.120	0.139	0.129	0.145
Log Pseudo Likelihood	−384.5	−368.4	−367.4	−364.9	−364.8	−364.9	−290.1	−168.2	−193.8
Journal Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Wald Test Stat., Control Vars.[a]									
Chi-squared	177.9	29.94	29.85	30.04	28.11	29.71	28.18	5.483	66.38
Degrees of freedom	9	9	9	9	9	9	8	8	8
p-value	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.705	0.000
Wald Test Statistics, Journal Dummies									
Chi-squared	215074	3.000e+08	4.140e+07	620923	497803	503847	8598	1.189e+06	2505
Degrees of freedom	16	18	18	18	18	18	15	17	9
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Wald Test Statistics, Year Dummies									
Chi-squared	380.8	549.4	669.8	510.7	394.5	367	322.6	36.53	58.29
Degrees of freedom	8	8	8	8	8	8	8	8	7
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Robust standard errors clustered at the journal level in parentheses. We used the *margins* command in Stata to obtain the marginal effects at the mean reported in this table. Dummy variables specified as factor variables.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.
[a] Control variables (not reported): *ProceedingsArticle*, *LogReferences*, *LogPages*, *Authors*, *LogBestH*, *SelfCreatedData*, *ConfidentialData*, *Funded*, *FunderDataSupport*.

independent variables under study, the marginal effect at the means shows how an article's probability of replication changes as the binary variable changes from 0 to 1 while holding all other independent variables at their means (Williams, 2012). Regarding the interpretation of the marginal effects of log-transformed variables, note that an increase of 1 in a log-transformed (continuous) variable corresponds to multiplying the variable by $e = 2.718$. This corresponds to a 171.83% change in the variable.

For the full sample analysis reported in columns [1] to [6], the marginal effects at the mean of *LogCitesReplication* range from 0.053 to 0.056 and are statistically significant at the 0.1% level across all specifications.[44] Hence, an increase in citations before publication of a replication study by about 170% is associated with an increase of the replication probability by about 5.3 to 5.6 percentage points. To illustrate this point, if citations before the publication of a replication study increase by 171.83% from 20.89 (i.e., the mean for the full sample, see Table 1) to 56.79, the replication probability increases by 5.3 to 5.6 percentage points. Recall that the mean replication probability is 10.5% for our sample (see Table 1). Therefore, one may argue that an increase in the replication probability by more than five percentage points reflects an important magnitude.

In addition, the log lag between the publication of the replicated articles and of the respective replication studies negatively affects the probability of replication. This effect is statistically significant at the 0.1% level across specifications. For the full sample analysis reported in columns [1] to [6], an increase in *logLagReplication* by one unit is associated with a decrease in the replication probability by 10.0 to 10.6 percentage points.[45] To illustrate, if the lag between publication of the replicated article and the respective replication study increases by 171.83% from 4.85 years (i.e., the mean for the full sample, see Table 1) to 13.18 years, the replication probability decreases by 10.0 to 10.6 percentage points. Interestingly, the effect of *LogLagReplication* is larger in magnitude than that of *LogCitesPreReplication*. This indicates that articles initially published longer ago are less likely to be the subject of a published replication study, irrespective of their citation performance.

Our analysis further suggests that articles published in top 5 journals are less likely to be eventually replicated. The marginal effects of

[44] We obtain similar results for the subsample analyses reported in columns [7] to [9]. The marginal effects at the mean of *LogCitesReplication* range from 0.040 to 0.067 and are statistically significant at the 0.1% level across all specifications.

[45] For the subsample analyses reported in columns [7] to [9] we obtain similar results. An increase in *logLagReplication* by one unit is associated with a decrease in the replication probability by 7.3 to 14.2 percentage points.

Top5Journal are negative and statistically significant at least at the 1% level across specifications. The only exception is the negative, statistically insignificant marginal effect of *Top5Journal* reported in column [8]. Our results for the full sample analysis reported in columns [1] to [6] suggest that publication in a top 5 journal is associated with a decrease in the replication probability by 8.1–9.6 percentage points. Taking into account the mean replication probability for our sample, i.e., 10.5%, we argue that a decrease in replication probability by more than 8 percentage points reflects an important magnitude.

The marginal effects of *LogImpactFactor* are also negative but statistically insignificant across specifications. The only exception is the negative, statistically significant marginal effect of *LogImpactFactor* reported in column [7].[46]

For the full sample analysis reported in columns [2] to [6] and for column [8], the marginal effects of *Top50University* are positive but statistically insignificant. For the subsample analysis reported in columns [7] and [9], however, we find that the marginal effects of *Top50University* are positive and statistically significant at the 5% level. The latter results suggest that journal articles by authors from higher ranked institutions have a higher incidence of replication in both scientific and reinforcing replication studies.

Finally, in the full sample analysis reported in columns [5] and [6], we find no significant effect of *MandatoryDisclosure* or *DataOrCode* on the probability of replication, respectively. We further explore this nexus in the following subsample analysis.

5.2.2. Subsample analysis: Articles published in or after 2004

It is important to note that, while the period under study is 1974 to 2014, disclosure of data and code is a relatively new phenomenon. For instance, *Econometrica* and the *AER* are the first journals in our sample to adopt a mandatory data disclosure policy in 2004 and 2005, respectively (see Appendix 1). Therefore, we run the regressions separately for the subsample of articles published in or after 2004. Results are reported in Table 4, where we include *MandatoryDisclosure* in columns [1] to [3] and *DataOrCode* in columns [4] to [6]. In columns [2] and [5], we refrain from including *Top5Journal*. Instead, we include a dummy variable for *AER* publication, *AER_Publication*. In columns [3] and [6], we include *AER_Publication* and its interaction with *MandatoryDisclosure* and *DataOrCode*, respectively. We include the *AER* interaction to address concerns that our results are driven by a single journal. For instance, 35 of the 130 replicated articles under study are published in the *AER*. In addition, 254 of the 286 articles published under a mandatory data disclosure policy are *AER* articles.

Comparing Table 4 with columns [1] to [6] from Table 3, we can see that the results on *LogCitesReplication, LogLagReplication,* and *Top5-Journal* remain qualitatively unchanged. In addition, the results on *LogImpactFactor* remain qualitatively unchanged except for columns [1] and [4].

In contrast to the results reported in Table 3, we find evidence for a positive effect of mandatory data disclosure polices on the probability of replication in the subsample of articles published in or after 2004. The marginal effect of *MandatoryDisclosure* is statistically significant at the 0.1% level in columns [1] and [2] ranging from 0.043 to 0.064. Our results suggest that publication in a journal is associated with an increase in the replication probability by 4.3 to 6.4 percentage points if the journal has a mandatory data disclosure policy. The marginal effect of *MandatoryDisclosure* remains statistically significant at the 0.1% level and is similar in magnitude (i.e., 4.5 percentage points) when we include the interaction of *MandatoryDisclosure* with the *AER* dummy in column [3]. Arguably, an increase in the replication probability by more than 4 percentage points reflects an important magnitude. Overall, our results suggest that mandatory data disclosure policies may decrease the cost of replication, thereby increasing the probability of replication.

We do not find evidence for a positive effect of data or code availability on the replication probability in columns [4] and [5]. However, once we include the interaction of *DataOrCode* with the *AER* dummy in column [6], the marginal effect of data or code availability on the replication probability (i.e., 9.3 percentage points) is statistically significant at the 1% level, while the effect of the interaction term is negative and statistically significant at the 1% level.

Finally, we find evidence for a positive effect of *Top50University* on the replication probability. The marginal effect of *Top50University* is statistically significant at the 1% level across columns ranging from 0.030 to 0.033. This result suggests that the replication probability is 3.0 to 3.3 percentage points higher for articles by authors affiliated with a top 50 university.

5.3. Robustness

We also perform linear probability regressions for robustness. Table 5 reports OLS regression coefficients for the full sample. These essentially yield the same results.

In addition, we explore whether the marginal effects of the main variables of interest are different in terms of sign and magnitude for different levels of article impact. Fig. 4 shows the marginal effects at the mean after probit regression of (A) *logCitesPreReplication*, (B) *Top5-Journal*, (C) *Top50University*, and (D) *DataOrCode* for specification (6) of Table 3 for different levels of article impact. We show 95% confidence intervals. The horizontal axis indicates article impact, i.e., *logCitesPreReplication*, while the vertical axis indicates the marginal effect on the replication probability.

Fig. 4A suggests that the marginal effect of *logCitesPreReplication* on the replication probability is positive for all levels of article impact and increasing in article impact. The marginal effect of *Top5Journal* on the replication probability is negative for all levels of article impact and decreasing in article impact (Fig. 4B). The marginal effect of *Top50University* on the replication probability is positive for all levels of article impact and increasing in article impact (Fig. 4C). Fig. 4D suggests that the marginal effect of *DataOrCode* on the replication probability remains at almost zero percentage points for different levels of article impact.

Finally, in a similar fashion as in Fig. 4, we explore whether the marginal effects of the main variables of interest are different in terms of sign and magnitude for different levels of author quality indicated by the H-index of the best author. Appendix 5 suggests that the marginal effects of *logCitesPreReplication*, *Top5Journal*, *Top50University*, and *DataOrCode* are not sensitive in terms of sign or magnitude for different levels of author quality.

5.4. Characteristics of replicated and replicating authors

We examine the characteristics of authors whose journal articles have been replicated (replicated authors) and compare them with the characteristics of the authors of the respective replicating articles (replicating authors) and with the characteristics of authors in the full sample of 1243 articles, i.e., authors of replicated and nonreplicated articles.[47] In Fig. 5, we consider the characteristics of authors from all articles in the sample (indicated by the four respective bars on the left-hand side), replicated authors (respective bars in the middle), and replicating authors (respective bars on the right-hand side). We obtained

[46] To illustrate this point, in column [7], a 172% increase in the impact factor of a journal in which an article is published is associated with a decrease of the replication probability by 7.2 percentage points.

[47] It is beyond the scope of the present article to examine the channels through which prospective replicating authors become aware of articles that are eventually replicated. In this respect, an analysis based on citation networks appears to be a particularly interesting avenue for further research.

Table 4
Marginal effects at the mean after probit regression: Articles published in or after 2004.

	[1]	[2]	[3]	[4]	[5]	[6]
Sample:	≥2004	≥2004	≥2004	≥2004	≥2004	≥2004
Dependent variable:	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article
Log total citations before publication of replication study	0.048***	0.050***	0.050***	0.047***	0.049***	0.050***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Log lag between publication of replicated article and replication	−0.110***	−0.118***	−0.119***	−0.108***	−0.114***	−0.118***
	(0.013)	(0.013)	(0.014)	(0.013)	(0.012)	(0.013)
Top 5 economics journal = 1	−0.190***			−0.100***		
	(0.051)			(0.028)		
AER publication = 1		−0.067***	−0.065**		−0.068***	−0.039***
		(0.016)	(0.023)		(0.012)	(0.011)
Log impact factor	0.128***	−0.010	−0.009	0.074**	−0.017	−0.012
	(0.036)	(0.020)	(0.021)	(0.027)	(0.019)	(0.022)
Top 50 university = 1	0.033**	0.032**	0.032**	0.032**	0.031**	0.030**
	(0.011)	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)
Mandatory data disclosure policy = 1	0.064***	0.043***	0.045**			
	(0.016)	(0.009)	(0.014)			
Mandatory data disclosure interacted with *AER* publication = 1			−0.002			
			(0.016)			
Data or program code available = 1				0.032	0.040	0.093**
				(0.020)	(0.023)	(0.033)
Data or program code available interacted with *AER* publication = 1						−0.045**
						(0.015)
Observations	560	560	560	560	560	560
Pseudo R-squared	0.184	0.187	0.187	0.181	0.187	0.190
Log Pseudo Likelihood	−131.7	−131.2	−131.2	−132.2	−131.2	−130.7
Journal Dummies	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES
Wald Test Stat., Control Vars. [a]						
Chi-squared	232.9	473.8	524.9	160.6	52.65	133.3
Degrees of freedom	9	9	9	9	9	9
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Wald Test Statistics, Journal Dummies						
Chi-squared	14387	1859	1125	20533	48157	571.1
Degrees of freedom	11	11	11	11	11	11
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Wald Test Statistics, Year Dummies						
Chi-squared	76.68	15.23	13.42	2.339	4.638	4.760
Degrees of freedom	2	2	2	2	2	2
p-value	0.000	0.000	0.001	0.311	0.099	0.093

Notes: Robust standard errors clustered at the journal level in parentheses. We used the *margins* command in Stata to obtain the marginal effects at the mean reported in this table. Dummy variables specified as factor variables.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.
[a] Control variables (not reported): *ProceedingsArticle*, *LogReferences*, *LogPages*, *Authors*, *LogBestH*, *SelfCreatedData*, *ConfidentialData*, *Funded*, *FunderDataSupport*.

the results reported in Fig. 5 from (A) the co-author with the highest H-index of all co-authors, (B) the co-author with the highest number of citations of all co-authors, and (C & D) the co-author that is affiliated with the highest ranked university of all co-authors.[48]

On average, the H-index of replicated authors (21.54) is almost twice as high as the H-index of replicating authors (10.98). It is also higher than the H-index of the authors of all articles in the sample (17.42). Replicated authors' total citations are, on average, four times higher than total citations of replicating authors. On average, the citations are also higher than the total citations of the authors of all articles in the sample (4119 vs. 2462). In addition, replicated authors are more likely to be affiliated with a Top 50 university (0.73) than replicating authors (0.33) and all authors (0.61). Finally, the universities of replicated authors are, on average, higher ranked than the universities of replicating authors and all authors.

6. Discussion and further research

In order to grapple with the phenomenon of replication in broader ways and to interpret our results against this background, it is instrumental to discuss and contextualize our results. This relates in particular to the difference between informal and formal replication (see 6.1), the implications of our results for other scientific disciplines (see 6.2), and the difference between self-created and widely available data (6.3).

6.1. Informal vs. formal replications

As discussed in subsection 3.3, our approach does not allow us to make statements about informal replications, e.g., replications published in working papers, initial sanity checks, or replications done in course work. For example, Sukhtankar (2017) analyzed empirical papers in development economics published in the top five and next five general interest journals between 2000 and 2015. Of the 1056 papers, 57 (5.4%) were replicated in another published paper or working paper. The number drops to 29 (2.7%) when considering only papers published in journals. Many replications might not be published in the traditional sense at all. For example, a university teacher might want her students to "learn from the best" and choose a study for replication in class that was particularly well executed. This kind of replication effort is unlikely to be published as a peer-reviewed paper (or a discussion paper) (Fecher et al., 2016). Furthermore, journals—in anticipation of reader expectations and impact scores—publish according to a certain news value (Feigenbaum and Levy, 1993). One could argue that replications generally lack these qualities and are therefore rarely published in peer-reviewed journals.

However, informal replications, as well as similarly indirect forms of

[48] Note that higher ranked universities have a lower position in the university ranking.

Table 5
Ordinary least squares (OLS) regression coefficients.

Issues:	[1] All	[2] All	[3] All	[4] All	[5] All	[6] All	[7] With sci. repl. art.	[8] With neg. repl. art.	[9] With reinf. repl. art.
Dependent variable:	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article	Scient. repl. art.	Negated repl. art.	Reinf. repl art.
Article published in conference proceedings = 1	−0.011	0.004	0.011	0.007	0.008	0.001	0.006	0.000	0.006
	(0.026)	(0.029)	(0.029)	(0.029)	(0.030)	(0.028)	(0.029)	(0.100)	(0.032)
Log number of references	−0.004	−0.019	−0.022	−0.020	−0.020	−0.020	−0.020	0.041	−0.064**
	(0.026)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)	(0.025)	(0.031)	(0.020)
Log number of pages	0.067*	0.041	0.046	0.042	0.042	0.044	0.064*	−0.018	0.084*
	(0.026)	(0.024)	(0.025)	(0.026)	(0.026)	(0.027)	(0.025)	(0.037)	(0.029)
Number of authors	0.017	0.015	0.015	0.014	0.014	0.014	0.018	0.013	0.018
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.021)	(0.012)
Log H-index of the best author	0.028	0.008	0.009	0.003	0.003	0.003	0.007	0.033	−0.018
	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.020)	(0.022)	(0.016)
Self-created data = 1	−0.017	−0.003	−0.001	0.002	0.002	0.004	0.023	0.002	−0.001
	(0.015)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.016)	(0.036)	(0.026)
Data proprietary according to notes on data & code = 1	−0.017	−0.020	−0.019	−0.018	−0.017	−0.014	[omitted]	[omitted]	[omitted]
	(0.091)	(0.077)	(0.077)	(0.081)	(0.084)	(0.084)			
Third-party funding = 1	−0.021	−0.028	−0.027	−0.026	−0.026	−0.025	−0.064	−0.020	−0.046
	(0.061)	(0.056)	(0.056)	(0.053)	(0.053)	(0.053)	(0.070)	(0.083)	(0.057)
Funder's support for data availability	0.012	0.011	0.011	0.010	0.010	0.010	0.021	0.002	0.020
	(0.016)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.019)	(0.021)	(0.016)
Log total citations before publication of replication		0.062***	0.063***	0.061***	0.061***	0.061***	0.062***	0.074***	0.051***
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.012)
Log lag between publication of replicated article and of respective replication		−0.104***	−0.105***	−0.102***	−0.103***	−0.104***	−0.109***	−0.144***	−0.091**
		(0.016)	(0.017)	(0.016)	(0.016)	(0.016)	(0.014)	(0.021)	(0.022)
Top 5 economics journal = 1			−0.088**	−0.090**	−0.086*	−0.077**	−0.057*	−0.028	−0.064
			(0.024)	(0.024)	(0.033)	(0.026)	(0.022)	(0.059)	(0.032)
Log impact factor			−0.037	−0.031	−0.033	−0.043	−0.080	−0.033	−0.109
			(0.034)	(0.033)	(0.038)	(0.037)	(0.039)	(0.048)	(0.058)
Top 50 university = 1				0.048	0.048	0.048	0.057*	0.026	0.053
				(0.026)	(0.026)	(0.026)	(0.026)	(0.029)	(0.031)
Mandatory data disclosure policy = 1					−0.003				
					(0.020)				
Data or program code available = 1						−0.017	−0.037	−0.027	−0.015
						(0.014)	(0.020)	(0.027)	(0.013)
Constant	−0.207**	0.002	0.126	0.105	0.107	0.116	0.071	0.017	0.245
	(0.065)	(0.066)	(0.069)	(0.067)	(0.067)	(0.070)	(0.095)	(0.109)	(0.119)
Observations	1225	1225	1225	1225	1225	1225	973	563	714
R-squared	0.045	0.069	0.070	0.075	0.075	0.076	0.088	0.087	0.082
Journal Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Wald Test Stat., Control Vars. [a]									
F statistic	9.337	1.535	1.687	1.189	1.179	1.279	3.938	0.712	4.823
Degrees of freedom	9	9	9	9	9	9	8	8	8
p-value	0.000	0.197	0.152	0.349	0.355	0.302	0.007	0.679	0.006
Wald Test Statistics, Journal Dummies									
F statistic	6429	63147	30200	54805	44618	80214	5712	1202	25.96
Degrees of freedom	16	18	18	18	18	18	15	17	9
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Wald Test Statistics, Year Dummies									
F statistic	41.84	23.77	21.58	15.12	15.68	18.28	24.91	14.57	21.51
Degrees of freedom	8	8	8	8	8	8	8	8	7
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000

Notes: Robust standard errors clustered at the journal level in parentheses. Dummy variables specified as factor variables. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.
[a] Control variables: *ProceedingsArticle*, *LogReferences*, *LogPages*, *Authors*, *LogBestH*, *SelfCreatedData*, *ConfidentialData*, *Funded*, *FunderDataSupport*.

falsification and verification, may also help the field to strengthen its knowledge base. We suggest that a better documentation of the instances of replication would be beneficial for the scientific community as it enables the identification of high-quality research and allows the rewarding of frequent replicators. It would furthermore facilitate an answer to the question of how many (more) replications are actually needed.

This also raises the question of whether traditional journals should be the sole place to document validation and falsification efforts. In line with Coffman et al. (2017), we argue that journals should indeed publish more replication studies, for instance in the form of short replication reports. In addition, attention should be given to novel infrastructures for scientific self-control and post-publication review, for example through watch blogs like RetractionWatch,[49] databases like the ReplicationWiki, replication journals (e.g., Hoeffler, 2017), or platforms like PubPeer[50]. Considering the issues raised in the introduction of this paper (increasing retraction rates, issues with replicability in many empirical disciplines), it seems reasonable to document and reward work that contributes to the quality of

[49] https://retractionwatch.com/ (last accessed 28 June 2018)
[50] https://pubpeer.com/static/about (last accessed 28 June 2018)

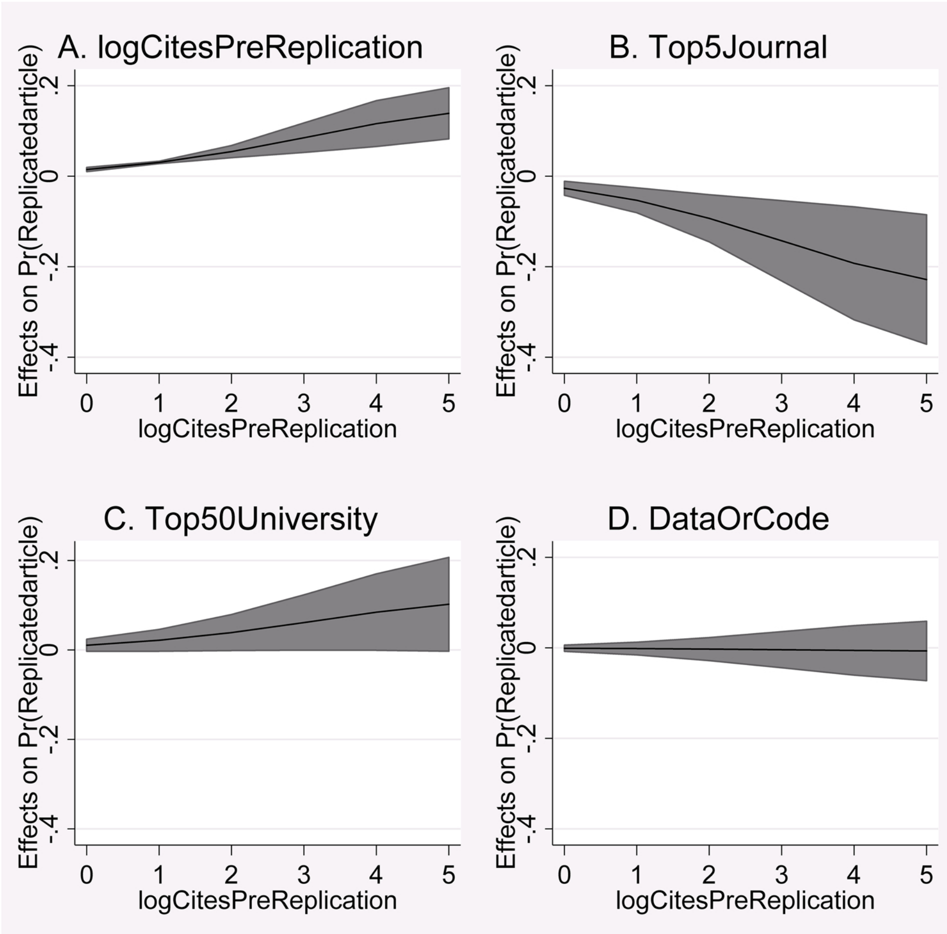


Fig. 4. Marginal effects at the means after probit regression (by article impact). We use the *marginsplot* command in Stata to create this figure. Fig. 4 shows the marginal effects at the mean after probit regression of (A) *logCitesPreReplication*, (B) *Top5Journal*, (C) *Top50University*, and (D) *DataOrCode* for specification (6) of Table 3 for different levels of *logCitesPreReplication*. 95% confidence intervals are shown. *logCitesPreReplication* is indicated by the horizontal axis. The marginal effect on the replication probability is indicated by the vertical axis. Figure A suggests that the marginal effect of *logCitesPreReplication* on the replication probability is positive. It increases in *logCitesPreReplication*. The marginal effect of *Top5Journal* on the replication probability is negative and decreases in *logCitesPreReplication* (Figure B). The marginal effect of *Top50University* on the replication probability is positive. It increases in *logCitesPreReplication* (Figure C). Finally, Figure D suggests that the marginal effect of *DataOrCode* on the replication probability remains at about 0 percentage points for different levels of *logCitesPreReplication*.

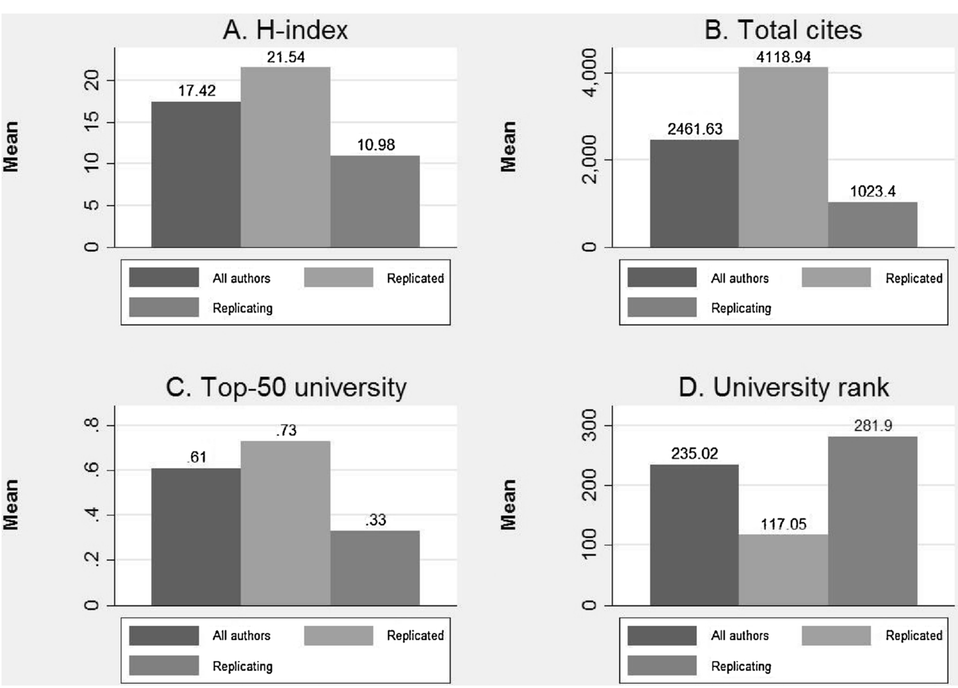


Fig. 5. Characteristics of authors from all articles in the sample (indicated by the four respective bars on the left-hand side), replicated authors (middle), and replicating authors (right-hand side). For the 1243 articles in the full sample, 130 replicated articles and their respective replicating articles, the results reported in this figure are obtained from (A) the co-author that has the highest H-index of all co-authors, (B) the co-author that has the highest number of citations of all co-authors, and (C & D) the co-author that is affiliated with the highest ranked university of all co-authors. Higher ranked universities have a lower position in the university ranking.

the academic knowledge base. This seems particularly relevant for the field of economics because, on the one hand, it has a high relevance for political and economic decision-making and because it has a strong tradition of impactful non-peer-reviewed publication in form of discussion papers on the other.

6.2. Implications for other scientific disciplines

Recent empirical evidence suggests that false positives are a major source for concern in many social and biomedical sciences (Bettis, 2012; Begley and Ellis, 2012; Brodeur et al., 2016; Freedman et al., 2015; Gall et al., 2017; Ioannidis, 2012; Open Science Collaboration, 2015; Prinz et al., 2011). In addition, recent theoretical work on high-quality scientific production and replication suggests that the underlying incentives for researchers and replicators associated with scientific research are likely to be similar in scientific fields where academic competition, i.e., publish or perish, is the leading paradigm (Gall et al., 2017; Lacetera and Zirulia, 2011; Kiri et al., 2018).[51] Hence, there is no a priori reason to believe that the incentives of an academic economist to conduct novel research and/or replication studies should be structurally different from the incentives of a political scientist, biomedical scientist, or social psychologist. It is in this respect that studying replication in empirical economics may shed light on replication in other fields. Or, as Coffman et al. (2017, p. 44) put it: “Clearly the issue of replicability is a source for substantial concern in economics as well. While we believe the economics profession is not doing too badly on the dimension of replicability, this is an area that economists can be leaders in designing better mechanisms for promulgating academic research.”

6.3. Self-created vs. widely available data

Although we differentiate between self-created and widely available data in our analysis, it cannot unequivocally be concluded that noncompliance with data disclosure policies led to fewer replication studies, nor whether making program code available is more important than a general increase in data availability. For instance, it is possible that authors in our sample did not have to make their data available (and thereby comply with a journal's data policy) because their data was already widely available (e.g., census data).

7. Conclusions

Our results confirm previous assumptions that relate replication to impact (Dewald et al., 1986; Furman et al., 2012; Hamermesh, 2007, 2017). Researchers tend to replicate high-impact research from renowned researchers and institutions. In this regard, private incentives are well aligned with societal interests, since high-impact publications are also the studies that are most likely to influence decision-makers in private and public organizations. It seems expedient with regards to the evolution of knowledge to critically examine new insights that could change the discourse in a field (Loasby, 2002; Popper, 1959; Freedman et al., 2015). In a game-theoretical analysis of high-quality scientific production, Kiri et al. (2018) propose that replicators will not attempt to replicate articles published in high-quality journals that specialize in separating out the high-quality articles from the low-quality articles. Our results, which show that the replication probability is lower for articles published in the top five economics journals, provide empirical support for Kiri et al. (2018)'s proposition. In addition, while our sample does not allow us to make statements regarding the optimal amount of replication, the observation that the amount of published articles in economics increased at a higher rate than the amount of published replication studies suggests that more attention should be given to published formal replication studies.

[51] See also Bobtcheff et al. (2017) and Stephan (2012).

Our results concerning data disclosure policies appear ambiguous at first sight. While we cannot detect a statistically strong impact of data disclosure policies on the replication probability for the full sample of articles published between 1974 and 2014, the picture changes when we restrict the analysis to the time period when mandatory data disclosure was introduced around 2004. Exploring the subsample of articles published in or after 2004, we find evidence for a statistically significant positive effect of mandatory data disclosure policies on the replication probability. Arguably, mandatory data disclosure policies may reduce the replicator's cost of replication, thereby increasing the replication probability. The lack of significance in the overall sample may simply be due to the fact that a genuine regime change occurred around 2004 that is not captured in the overall regression. We also show that for 37% of the empirical articles studied that are subject to mandatory data disclosure, the data or program code was not available even though the data was not proprietary. This raises concerns regarding the enforcement of mandatory data disclosure policies (Vlaeminck and Hermann, 2018).

Hence, our results suggest that replication efforts could be incentivized by reducing the cost of replication, for example by promoting data disclosure (see also Hoeffler, 2017).[52] Our results further suggest that the decision to conduct a replication study is, at least partly, driven by the replicator’s reputation considerations. Other possible explanations are the importance of the topics explored in the replicated articles and their impact on public policy as well as editors' publication strategies vis-à-vis replication studies. We argue that the low number of replication studies being conducted could potentially increase if replication studies received more formal recognition (for instance, through publication in [high-impact] journals), specific funding, (for instance, for the replication of articles with a high impact on public policy), or awards.[53]

Our empirical study does not cover informal replication studies that are not published in peer-reviewed journals. These, however, seem to gain in importance with a recognizable move toward open science and post-publication review. Informal replication practices include, for instance, replications conducted in teaching, discussion in a public forum (e.g., PubPeer), and replication studies published as working papers on preprint servers and archives. It appears highly relevant to examine these practices in a consecutive study in order to get an indication of the sufficient amount of replication as well as the potential of novel review mechanisms. Since replication is, at least partly, driven by reputational reward, it may be a viable strategy to document and reward formal as well as informal replication practices. A potentially viable strategy could be to promote replication studies in teaching and as an (optional) chapter of dissertations (Fecher et al., 2016). This would be easy to implement because cumulative dissertations are standard practice for dissertations in economics.

Acknowledgements

The proprietary data used in this paper were collected and first analyses done when Frank Mueller-Langer was Senior Research Fellow at Max Planck Institute for Innovation and Competition (MPI-IC). This work was supported by the German Research Foundation (DFG) under the project European Data Watch Extended (DFG-Grant Number WA 547/6-2). The funders had no role in study design, in the writing of the manuscript, and in the decision to submit the article for publication. We thank Georgios Alaveras, Néstor Duch-Brown, Fabian Gässler, Estrella Gómez Herrera,

[52] Considering the growth in publication of articles (Bornmann and Mutz, 2015), it seems highly relevant to examine further whether the control mechanisms can cope with a higher publication output. If more replications in economics are needed, it would make sense to invest in lowering the costs of replication, for instance by promoting the availability of data and code, and to increase the reputational gain from replication efforts (e.g., through awards).

[53] We find that funding bodies typically do not have replication policies (Appendix 3). We propose that funding bodies introduce replication policies if one of their objectives is to improve the boundary conditions for replication practice.

Bronwyn Hall, Karin Hoisl, Bertin Martens, Olaf Siegert, Sven Vlaeminck, Joachim Wagner, Stefan Wagner, and Peter Winker, as well as three anonymous referees and seminar and conference participants at MPI-IC, LMU Munich, TUM School of Management, and JRC Seville for their valuable comments. We are very grateful to the journal editors who provided us with information about the introduction (or absence) of data disclosure policies. Moritz Mosenhauer, Wolfgang Rösch, and Christoph Winter provided excellent research assistance.

Appendix 1. Overview of the journals, data policies and articles under study

Journal	Impact Factor Rank	Mandatory Data Disclosure Policy	Year / Volume of Enactment	# Articles	# Repl. Articles
JOURNAL OF ECONOMIC LITERATURE	1				
JOURNAL OF FINANCE	2			118	10
QUARTERLY JOURNAL OF ECONOMICS	3			72	10
JOURNAL OF ECONOMIC PERSPECTIVES	4				
TRANSPORTATION RESEARCH PART B-METHODOLOGICAL	5				
JOURNAL OF FINANCIAL ECONOMICS	6			59	9
JOURNAL OF POLITICAL ECONOMY	7	YES	2005/113	72	13
REVIEW OF FINANCIAL STUDIES	8			4	1
ECONOMETRICA	9	YES	2004/72	10	3
JOURNAL OF THE EUROPEAN ECONOMIC ASSOC.	10	YES	2011/9	28	1
REVIEW OF ENVIRONMENTAL ECONOMICS AND POLICY	11				
PHARMACOECONOMICS	12				
AMERICAN ECONOMIC REVIEW	13	YES	2005/95	478	35
ECONOMIC GEOGRAPHY	14				
REVIEW OF ECONOMIC STUDIES	15	YES	2006/73		
JOURNAL OF ECONOMIC GROWTH	16				
BROOKINGS PAPERS ON ECONOMIC ACTIVITY	17	YES	2008/39	6	1
AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS	18	YES	2009/1		
JOURNAL OF FINANCIAL STABILITY	19				
VALUE IN HEALTH	20				
AMERICAN ECONOMIC JOURNAL-MACROECONOMICS	21	YES	2009/1		
ECONOMIC POLICY	22			5	1
JOURNAL OF ACCOUNTING & ECONOMICS	23			58	9
JOURNAL OF ECONOMIC GEOGRAPHY	24				
TECHNOL. AND ECONOMIC DEVELOPMENT OF ECONOMY	25				
REVIEW OF ECONOMICS AND STATISTICS	26	YES	2010/92	73	4
ECONOMIC JOURNAL	27	YES	2012/122	10	1
ENERGY ECONOMICS	28				
AMERICAN ECONOMIC JOURNAL-ECONOMIC POLICY	29	YES	2009/1		
EXPERIMENTAL ECONOMICS	30			9	1
TRANSPORT. RES. PART A-POLICY AND PRACTICE	31			4	1
JOURNAL OF ENVIR. ECONOMICS AND MANAGEMENT	32			12	2
ECOLOGICAL ECONOMICS	33			33	4
ANNUAL REVIEW OF ECONOMICS	34				
ECONOMICS & HUMAN BIOLOGY	35				
JOURNAL OF INTERNATIONAL ECONOMICS	36			10	1
JOURNAL OF DEVELOPMENT ECONOMICS	37	YES	2014/111	12	1
ECONOMIC SYSTEMS RESEARCH	38				
FOOD POLICY	39				
JOURNAL OF BUSINESS & ECONOMIC STATISTICS	40	YES	2011/29	34	3
JOURNAL OF AGRARIAN CHANGE	41				
JOURNAL OF POLICY ANALYSIS AND MANAGEMENT	42				
JOURNAL OF HEALTH ECONOMICS	43			72	12
JOURNAL OF TRANSPORT GEOGRAPHY	44				
TRANSPORTATION RESEARCH PART E-LOGISTICS AND TRANSPORTATION REVIEW	45			5	1
HEALTH ECONOMICS	46			41	6
WORLD BANK RESEARCH OBSERVER	47				
JOURNAL OF MONETARY ECONOMICS	48				
JOURNAL OF REGIONAL SCIENCE	49				
JOURNAL OF LABOR ECONOMICS	50	YES	2010/28		
			TOTAL:	1225	130

Notes: Table based on the 1,225 observations used in the regressions. 23 journals under study that published at least one replicated article in bold.

Appendix 2. Correlation matrix for the dependent variable and main variables of interest

	Replicated Article	*CitesPre Replication*	*Lag Replication*	*Top5 Journal*	*Impact Factor*	*Top50 University*	*Mandatory Disclosure*	*DataOr Code*
ReplicatedArticle	1.00							
CitesPreReplication	0.21	1.00						
LagReplication	0.02	0.36	1.00					
Top5Journal	−0.03	0.17	0.37	1.00				
ImpactFactor	−0.01	0.11	0.08	0.11	1.00			
Top50University	0.09	0.11	0.08	0.20	0.15	1.00		
MandatoryDisclosure	−0.07	−0.06	−0.12	0.45	−0.12	0.14	1.00	
DataOrCode	−0.02	0.00	−0.12	0.35	−0.10	0.10	0.69	1.00

Appendix 3. Overview of data guidelines of funding bodies

Funding Body	[1] Policy/ guideline mentions data sharing	[2] Policy/ guideline specifies how or where to publish data	[3] Policy/ guideline mentions data documentation/ metadata	[4] Policy/ guideline requires data management plan	[5] Policy/ guideline mentions embargo period
National Science Foundation (NSF)	1	0	0	1	1
US National Institutes of Health (NIH)	1	1	1	1	1
National Natural Science Foundation of China (NSFC)	0	0	0	0	0
Deutsche Forschungsgemeinschaft (DFG)	1	1	1	0	0
Bundesministerium für Bildung und Forschung (BMBF)	0	0	0	0	0
Japanese society for the promotion of science (JSPS)	0	0	0	0	0
Japan Science and Technology Agency (JST)	0	0	0	0	0
Russian Foundation for Basic Research (RFBR)	0	0	0	0	0
French National Research Agency (ANR)	1	0	0	0	0
National Research Foundation of Korea (NRF)	0	0	0	0	0
Ministry of Knowledge Economy (MKE)	0	0	0	0	0
Science & Technology Facilities Council (STFC)	1	1	1	1	1
Medical Research Council (MRC)	1	1	1	1	1
Engineering and Physical Science Research Council (EPSRC)	1	1	1	0	1
Biotechnology and Biological Science Research Council	1	1	1	1	1
Natural Environment Research Council (NERC)	1	1	1	1	1
Economic and Social Research Council (ESRC)	1	1	1	1	1
Arts and Humanities Research Council (AHRC)	1	1	1	0	1
National Research Council (CNR)	0	0	0	0	0
Ministry of Education, Universities and Research (MIUR)	0	0	0	0	0
Spanish Ministry of Economy and Competitiveness	0	0	0	0	0
Social Sciences and Humanities Research Council of Canada	1	1	1	1	1
Natural Science and Engineering Research Council of Canada	1	1	1	1	1
Netherlands Organisation for Scientific Research (NWO)	1	1	1	1	1
Swedish Research Council (VR)	1	1	1	0	1
Swedish Foundation for Strategic Research (SSF)	0	0	0	0	0
Swedish Agency for Innovation Systems (VINNOVA)	0	0	0	0	0
Australian Research Council (ARC)	1	1	1	0	1
Norwegian Research Council	1	1	1	1	1
Portuguese Found. Sci. & Techn. (FCT)	1	1	1	1	1
Bill and Melinda Gates Foundation	1	1	1	1	1
Wellcome Trust	1	1	1	1	1
NASA	1	0	1	1	1
Leverhulme Trust	0	0	0	0	0

European Commission (Horizon 2020)	1	1	1	1	1
Swiss National Science Foundation	0	0	0	0	0

Notes: FunderDataSupport is a variable ranging from 0 to 5 that indicates the number of data policies and data management tools that external research funders provide to the authors they support. This variable can be thought of as the extent to which external funders have policies in place that facilitate data availability. To obtain this variable, we analyzed the funding guidelines of 36 research funding bodies worldwide regarding their data management policies. We treat the guidelines as textual data and code if they mention data management, if they specify where and how data should be stored, its terms of access (e.g., on request or public), documentation standards, and if they mention replication studies.

Appendix 4. Full version of Table 3 (Marginal effects at the mean after probit)

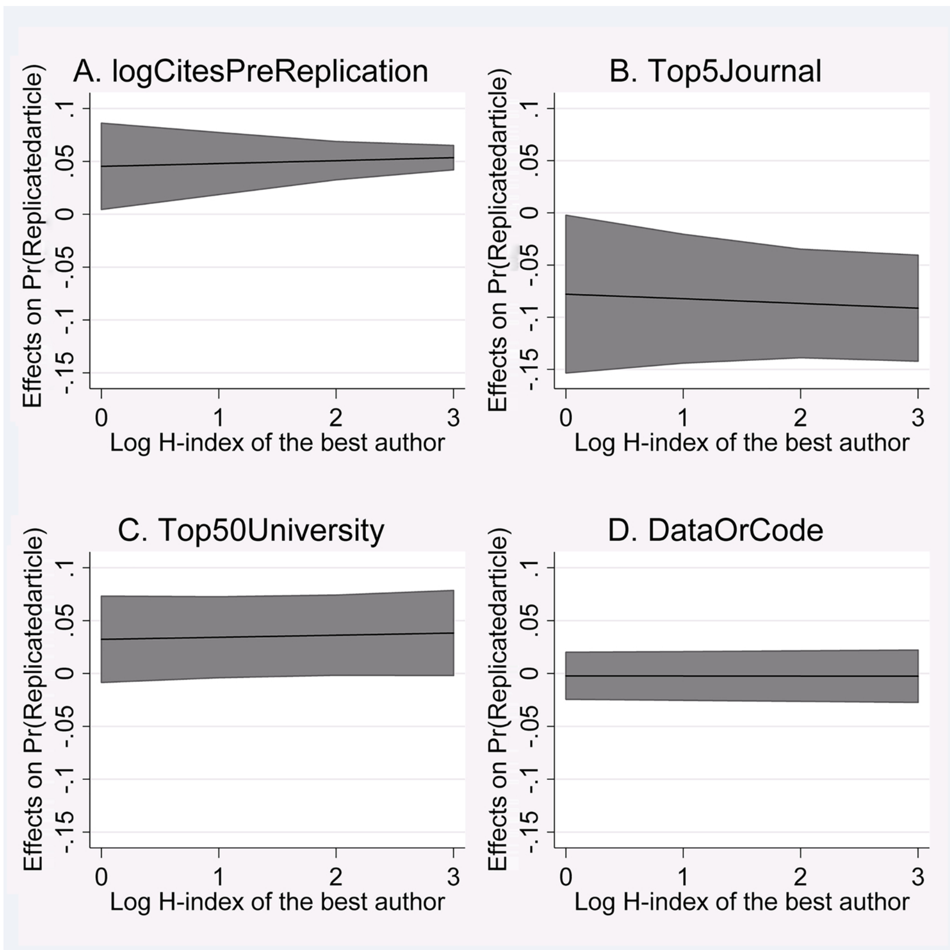
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Issues:	All	All	All	All	All	All	With sci. repl. art.	With neg. repl. art.	With reinf. repl. art.
Dependent variable:	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article	Repl. article	Scient. repl. art.	Neg. repl. art.	Reinf. repl art.
Article published in conference proceedings = 1	−0.047*	−0.039	−0.035	−0.036	−0.035	−0.037	−0.030	0.014	−0.043**
	(0.019)	(0.023)	(0.023)	(0.023)	(0.024)	(0.022)	(0.023)	(0.073)	(0.015)
Log number of references	−0.005	−0.016	−0.019	−0.016	−0.017	−0.016	−0.014	0.039	−0.049**
	(0.027)	(0.024)	(0.024)	(0.025)	(0.024)	(0.025)	(0.025)	(0.033)	(0.019)
Log number of pages	0.072**	0.045*	0.053*	0.048	0.049	0.048	0.068**	−0.016	0.071**
	(0.027)	(0.022)	(0.025)	(0.026)	(0.026)	(0.026)	(0.023)	(0.042)	(0.023)
Number of authors	0.012	0.009	0.009	0.008	0.008	0.008	0.010	0.007	0.011
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.013)	(0.008)
Log H-index of the best author	0.029	0.007	0.008	0.005	0.005	0.005	0.008	0.032	−0.012
	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.017)	(0.021)	(0.014)
Self-created data = 1	−0.018	−0.002	−0.001	0.001	0.002	0.001	0.017	0.007	−0.006
	(0.012)	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.015)	(0.034)	(0.018)
Data proprietary according to notes on data & code = 1	−0.005	−0.011	−0.011	−0.013	−0.010	−0.012	[dropped]	[dropped]	[dropped]
	(0.101)	(0.079)	(0.079)	(0.080)	(0.085)	(0.081)			
Third-party funding = 1	−0.019	−0.029	−0.029	−0.029	−0.028	−0.029	−0.048	−0.012	−0.057
	(0.044)	(0.035)	(0.035)	(0.034)	(0.034)	(0.034)	(0.038)	(0.050)	(0.037)
Funder's support for data availability	0.011	0.011	0.011	0.011	0.011	0.011	0.018	−0.002	0.027
	(0.013)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.015)	(0.014)	(0.019)
Log total citations before publication of replication		0.055***	0.056***	0.053***	0.053***	0.053***	0.053***	0.067***	0.040***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.010)	(0.007)
Log lag between publication of replicated article and replication		−0.106***	−0.106***	−0.100***	−0.101***	−0.101***	−0.105***	−0.142***	−0.073***
		(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.013)	(0.022)	(0.016)
Top 5 economics journal = 1			−0.096***	−0.092***	−0.081**	−0.090***	−0.072***	−0.037	−0.069**
			(0.024)	(0.025)	(0.030)	(0.025)	(0.018)	(0.043)	(0.026)
Log impact factor			−0.040	−0.036	−0.042	−0.038	−0.072**	−0.008	−0.071
			(0.022)	(0.023)	(0.027)	(0.026)	(0.025)	(0.040)	(0.037)
Top 50 university = 1				0.038	0.038	0.038	0.044*	0.019	0.042*
				(0.020)	(0.020)	(0.020)	(0.019)	(0.027)	(0.018)
Mandatory data disclosure policy = 1					−0.011				
					(0.016)				
Data or program code available = 1						−0.002	−0.016	−0.014	0.003
						(0.012)	(0.017)	(0.025)	(0.008)
Observations	1225	1225	1225	1225	1225	1225	973	563	714
Pseudo R-squared	0.0723	0.111	0.113	0.120	0.120	0.120	0.139	0.129	0.145
Log Pseudo Likelihood	−384.5	−368.4	−367.4	−364.9	−364.8	−364.9	−290.1	−168.2	−193.8
Journal Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Wald Test Stat., Control Vars. [a]									
Chi-squared	177.9	29.94	29.85	30.04	28.11	29.71	28.18	5.483	66.38
Degrees of freedom	9	9	9	9	9	9	8	8	8
p-value	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.705	0.000
Wald Test Statistics, Journal Dummies									
Chi-squared	215074	3.000e+08	4.140e+07	620923	497803	503847	8598	1.189e+06	2505
Degrees of freedom	16	18	18	18	18	18	15	17	9

p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Wald Test Statistics, Year Dummies									
Chi-squared	380.8	549.4	669.8	510.7	394.5	367	322.6	36.53	58.29
Degrees of freedom	8	8	8	8	8	8	8	8	7
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Robust standard errors clustered at the journal level in parentheses. We used the *margins* command in Stata to obtain the marginal effects at the mean reported in this table. Dummy variables specified as factor variables. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

[a] Control variables: *ProceedingsArticle, LogReferences, LogPages, Authors, LogBestH, SelfCreatedData, ConfidentialData, Funded, FunderDataSupport.*

Appendix 5. Marginal effects at the means after probit regression (by H-Index of best author)



Notes: This figure shows the marginal effects at the mean after probit regression of (A) *logCitesPreReplication*, (B) *Top5Journal*, (C) *Top50University*, and (D) *DataOrCode* for specification (6) of Table 3 for different levels of *logBestH*. We use the *marginsplot* command in Stata to create this figure. 95% confidence intervals are shown. *logBestH* is indicated by the horizontal axis. The marginal effect on the replication probability is indicated by the vertical axis. Figure A suggests that the marginal effect of *logCitesPreReplication* on the replication probability is positive and remains almost unchanged in magnitude if *logBestH* increases. The marginal effect of *Top5Journal* on the replication probability is negative and remains almost unchanged in magnitude if *logBestH* increases (Figure B). The marginal effect of *Top50University* on the replication probability is positive and remains almost unchanged when *logBesH* increases (Figure C). Finally, Figure D suggests that the marginal effect of *DataOrCode* on the replication probability remains at about 0 percentage points if *logBestH* increases. Overall, Appendix 5 suggests that the marginal effects of the main variables of interest are not sensitive to changes in *logBestH*.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.respol.2018.07.019.

References

Abrevaya, J., Puzzello, L., 2012. Taxes, cigarette consumption, and smoking intensity: comment. Am. Econ. Rev. 102 (4), 1751–1763.
Adda, J., Cornaglia, F., 2006. Taxes, cigarette consumption, and smoking intensity. Am. Econ. Rev. 96 (4), 1013–1028.
Anderson, L.R., Holt, C.A., 1997. Information cascades in the laboratory. Am. Econ. Rev. 87 (5), 847–862.
Anderson, R.G., Kichkha, A., 2017. Replication, meta-analysis, and research synthesis in economics. Am. Econ. Rev.: Papers Proc. 107 (5), 56–59.
Anderson, R.G., Greene, W.H., McCullough, B.D., Vinod, H.D., 2008. The role of data/code archives in the future of economic research. J. Econ. Methodol. 15 (1), 99–119.
Andreoli-Versbach, P., Mueller-Langer, F., 2014. Open access to data: an ideal professed but not practised. Res. Policy 43 (9), 1621–1633.
Angrist, J.D., Azoulay, P., Ellison, G., Hill, R., Lu, S.F., 2017. Inside job or deep impact? Using extramural citations to assess economic scholarship. NBER Working Paper No. 23698.
Azoulay, P., Furman, J.L., Krieger, J.L., Murray, F., 2015. Retractions. Rev. Econ. Stat. 97 (5), 1118–1136.
Bali, T.G., Cakici, N., Yan, X., Zhang, Z., 2005. Does idiosyncratic risk really matter? J. Finance 60 (2), 905–929.
Baltagi, B., 2003. Introducing a replication section. J. Appl. Econom. 18 (1) 111-111.
Begley, C.G., Ellis, L.M., 2012. Drug development: raise standards for preclinical cancer research. Nature 483, 531–533.
Bell, A., Johnston, R., Jones, K., 2015. Stylised fact or situated messiness? The diverse effects of increasing debt on national economic growth. J. Econ. Geogr. 15 (2), 449–472.
Berry, J., Coffman, L.C., Hanley, D., Gihleb, R., Wilson, A.J., 2017. Assessing the rate of replication in economics. Am. Econ. Rev.: Papers Proc. 107 (5), 27–31.
Bettis, R.A., 2012. The search for asterisks: compromised statistical tests and flawed theories. Strateg. Manag. J. 33 (1), 108–113.
Bhushan, R., 1989. Firm characteristics and analyst following. J. Account. Econ. 11 (2-3), 255–274.
Bobtcheff, C., Bolte, J., Mariotti, T., 2017. Researcher's dilemma. Rev. Econ. Stud. 84 (3), 969–1014.
Bohannon, J., 2015. Many psychology papers fail replication test. Science 349, 910–911.
Bornmann, L., Mutz, R., 2015. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J. Assoc. Inf. Sci. Technol. 66 (11), 2215–2222.
Brodeur, A., Le, M., Sangnier, M., Zylberberg, Y., 2016. Star wars: the empirics strike back. Am. Econ. J. Appl. Econ. 8 (1), 1–32.
Camerer, C.F., Dreber, A., Forswell, E., et al., 2016. Evaluating replicability of laboratory experiments in economics. Science. https://doi.org/10.1126/science.aaf0918.
Campbell, D.T., Stanley, J.C., 1963. Experimental and Quasi-experimental Designs for Research. Rand McNally, Chicago.
Card, D., DellaVigna, S., 2013. Nine facts about top journals in economics. J. Econ. Lit. 51, 144–161.
Chang, A.C., Li, P., 2017. A preanalysis plan to replicate sixty economics research papers that worked half of the time. Am. Econ. Rev.: Papers Proc. 107 (5), 60–64.
Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74 (368), 829–836.
Coffman, L.C., Niederle, M., 2015. Pre-analysis plans have limited upside, especially where replications are feasible. J. Econ. Perspect. 29 (3), 81–98.
Coffman, L.C., Niederle, M., Wilson, A.J., 2017. A proposal to organize and promote replications. Am. Econ. Rev.: Papers Proc. 107 (5), 41–45.
Cokol, M., Ozbay, F., Rodriguez-Esteban, R., 2008. Retraction rates are on the rise. EMBO Rep. 9 (2).
Cooper, H., 2017. Research Synthesis and Meta-analysis. SAGE Publications, New York.
Cooper, H., Hedges, L.V., 2009. In: Cooper, H., Hedges, L.V., Valentine, J.C. (Eds.), Potentials and Limitations. In: The Handbook of Research and Synthesis and Meta-Analysis. Russell Sage Foundation, New York, pp. 561–572.
Cornell, B., 1985. The weekly pattern in stock returns: cash versus futures: a note. J. Finance 40 (2), 583–588.
Couch, K.A., Placzek, D.W., 2010. Earnings losses of displaced workers revisited. Am. Econ. Rev. 100 (1), 572–589.
Cyranoski, D., 2006. Rise and fall. news@nature. https://doi.org/10.1038/news060109-8.
Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Res. Policy 23 (5), 487–521.
Dewald, W., Thursby, J., Anderson, R.G., 1986. Replication in empirical economics: the Journal of Money, Credit and Banking project. Am. Econ. Rev. 76 (4), 587–603.
Donohue, J., Levitt, S.D., 2001. The impact of legalized abortion on crime. Q. J. Econ. 116 (2), 379–420.
Donohue, J., Levitt, S.D., 2004. Further evidence that legalized abortion lowered crime: a reply to Joyce. J. Hum. Resour. 39 (1), 29–49.
Dosi, G., 1988. Sources, procedures, and microeconomic effects of innovation. J. Econ. Lit. 26 (3), 1120–1171.
Duvendack, M., Palmer-Jones, R.W., Reed, W.R., 2015. Replications in economics: a progress report. Econ J. Watch 12 (2), 164–191.
Dyl, E.A., Maberly, E.D., 1986. The weekly pattern in stock index futures: a further note. J. Finance 41 (5), 1149–1152.
Engel, C., 2011. Dictator games: a meta study. Exp. Econ. 14 (4), 583–610.
European Commission, 2012. Towards Better Access to Scientific Information: Boosting the Benefits of Public Investments in Research. European Commission, Brussels.
European Commission, 2016. H2020 Programme: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. European Commission, Directorate-General for Research & Innovation, Brussels.
Fecher, B., Fraessdorf, M., Wagner, G.G., 2016. Perceptions and practices of replication by social and behavioral scientists: making replications a mandatory element of curricula would be useful. IZA Discussion Paper Series No. 9896.
Fehder, D.C., Murray, F., Stern, S., 2014. Intellectual property rights and the evolution of scientific journals as knowledge platforms. Int. J. Ind. Organ. 36, 83–94.
Feigenbaum, S., Levy, D.M., 1993. The market for (ir)reproducible econometrics. Soc. Epistemol. 7 (3), 215–232.
Foote, C.L., Goetz, C.F., 2008. The impact of legalized abortion on crime: comment. Q. J. Econ. 123 (1), 407–423.
Fraas, A., Lutter, R., 2012. Efficient pollution regulation: getting the prices right: comment. Am. Econ. Rev. 102 (1), 602–607.
Franco, A., Malhotra, N., Simonovits, G., 2014. Publication bias in the social sciences: unlocking the file drawer. Science 345 (6203), 1502–1505.
Freedman, L.P., Cockburn, I.M., Simcoe, T.S., 2015. The economics of reproducibility in preclinical research. PLoS Biol. 13 (6).
Frisch, R., 1933. Editor's note. Econometrica 1 (1), 1–4.
Fryer, R.G., Heaton, P.S., Levitt, S.D., Murphy, K.M., 2013. Measuring crack cocaine and its impact. Econ. Inq. 51 (3), 1651–1681.
Furman, J.F., Jensen, K., Murray, F., 2012. Governing knowledge in the scientific community: exploring the role of retractions in biomedicine. Res. Policy 41 (2), 276–290.
Gall, T., Ioannidis, J.P.A., Maniadis, Z., 2017. The credibility crisis in research: can economics tools help? PLoS Biol. 15 (4), e2001846.
Gans, J.S., Murray, F., Stern, S., 2017. Contracting over the disclosure of scientific knowledge: intellectual property and academic publication. Res. Policy 46 (4), 820–835.
Gerdtham, U.G., Johannesson, M., Lundberg, L., Isacson, D., 1999. A note on validating Wagstaff and van Doorslaer's health measure in the analysis of inequalities in health. J. Health Econ. 18 (1), 117–124.
Gilbert, D.T., King, G., Pettigrew, S., Timothy, D., Wilson, T.D., 2016. Comment on "Estimating the reproducibility of psychological science". Science 351 (6277), 1037.
Goyal, A., Santa-Clara, P., 2003. Idiosyncratic risk matters. J. Finance 58 (3), 975–1008.
Grant, P., 2002. Scientific credit and credibility. Nat. Mater. 1, 139–141.
Haeussler, C., 2011. Information-sharing in academia and the industry: a comparative study. Res. Policy 40 (1), 105–122.
Hamermesh, D., 2007. Viewpoint: replication in economics. Can. J. Econ. 40 (3), 715–733.
Hamermesh, D., 2017. Replication in labor economics: evidence from data and what it suggests. Am. Econ. Rev.: Papers Proc. 107 (5), 37–40.
Hastings, J., 2004. Vertical relationships and competition in retail gasoline markets: empirical evidence from contract changes in Southern California. Am. Econ. Rev. 94 (1), 317–328.
Herndon, T., Ash, M., Pollin, R., 2014. Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. Cambridge J. Econ. 38 (2), 257–279.
Hoeffler, J.H., 2017. Replication and economics journal policies. Am. Econ. Rev.: Papers Proc. 107 (5), 52–55.
Hung, A.A., Plott, C.R., 2018. Information cascades: replication and an extension to majority rule and conformity-rewarding institutions. Am. Econ. Rev. 91 (5), 1508–1520.
Hunter, J., 2001. The desperate need for replications. J. Consum. Res. 28 (1), 149–158.
Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med. 2, e124.
Ioannidis, J.P.A., 2012. Why science is not necessarily self-correcting. Perspect. Psychol. Sci. 7 (6), 645–654.
Iversen, J., Söderström, U., 2014. The dynamic behavior of the real exchange rate in sticky price models: comment. Am. Econ. Rev. 104 (3), 1072–1089.
Jacobson, L.S., LaLonde, R.J., Sullivan, D.G., 1993. Earnings losses of displaced workers. Am. Econ. Rev. 83 (4), 685–709.
Jasny, B.R., Chin, G., Chong, L., Vignieri, S., 2011. Data replication & reproducibility. again, and again, and again …. Introduction. Science 334, 1225.
Journal of Political Economy, 1975. Editorial and Comment. J. Polit. Econ. 83, 1295–1296.
Joyce, T., 2004. Did legalized abortion lower crime? J. Hum. Resour. 39 (1), 1–28.
Joyce, T., 2009. A simple test of abortion and crime. Rev. Econ. Stat. 91 (1), 112–123.
Kiri, B., Lacetera, N., Zirulia, L., 2018. Above a swamp: a theory of high-quality scientific production. Res. Policy 47 (5), 827–839.
Kuhn, T.S., 1970. The Structure of Scientific Revolutions. University of Chicago Press, Chicago.
Lacetera, N., Zirulia, L., 2011. The economics of scientific misconduct. J. Law Econ. Organ. 27 (3), 568–603.
Lee, R.H., 2008. Future costs in cost effectiveness analysis. J. Health Econ. 27 (4), 809–818.
Levelt, W.J.M., Drenth, P., Noort, E., 2012. Flawed science: the fraudulent research practices of social psychologist Diederik Stapel. Commissioned by the Tilburg University. University of Amsterdam and the University of Groningen.
Loasby, B.J., 2002. The evolution of knowledge: beyond the biological model. Res. Policy 31 (8-9), 1227–1239.
Maniadis, Z., Tufano, F., List, J.A., 2015. How to make experimental economics research more reproducible: lessons from other disciplines and a new proposal. In: In: Deck, C., Fatas, E., Rosenblat, T. (Eds.), Replication in Experimental Economics. Research in Experimental Economics, vol. 18. Emerald Group Publishing, Bingley, pp. 215–230.
McCullough, B.D., McGeary, K.A., Harrison, T.D., 2006. Lessons from the JMCB archive. J. Money Credit Bank. 38 (4), 1093–1107.
McCullough, B.D., McGeary, K.A., Harrison, T.D., 2008. Do economic journal archives promote replicable research? Can. J. Econ. 41 (4), 1406–1420.

McNutt, M., 2014. Reproducibility. Science 343 229-229.
Mirowski, P., Sklivas, S., 1991. Why econometricians don't replicate (although they do reproduce). Rev. Polit. Econ. 3 (2), 146–163.
Mueller-Langer, F., Andreoli-Versbach, P., 2018. Open access to research data: strategic delay and the ambiguous welfare effects of mandatory data disclosure. Inf. Econ. Policy 42 (1), 20–34.
Murray, F., Stern, S., 2007. Do formal intellectual property rights hinder the free flow of scientific knowledge?: an empirical test of the anti-commons hypothesis. J. Econ. Behav. Organ. 63 (4), 648–687.
Murray, F.E., Aghion, P., Dewatripont, M., Kolev, J., Stern, S., 2016. Of mice and academics: examining the effect of openness on innovation. Am. Econ. J. Econ. Policy 8 (1), 212–252.
Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349 (6251) aac4716.
Organisation for Economic Co-Operation and Development (OECD), 2016. Gross Domestic Expenditure on R-D by Sector of Performance and Field of Science. At < http://stats.oecd.org/Index.aspx?DataSetCode=GERD_SCIENCE > . .
Popper, K.R., 1959. The Logic of Scientific Discovery. Hutchinson & Co, London.
Prinz, F., Schlange, T., Asadullah, K., 2011. Believe it or not: How much can we rely on published data on potential drug targets? Nat. Rev. Drug Discov. 10 (9) 712-712.
Reinhart, C.M., Rogoff, K.S., 2010. Growth in a time of debt. Am. Econ. Rev. Papers Proc. 100 (2), 573–578.
Reinhart, C.M., Reinhart, V.R., Rogoff, K.S., 2012. Public debt overhangs: advanced-economy episodes since 1800. J. Econ. Perspect. 26 (3), 69–86.
Rock, S., Sedo, S., Willenborg, M., 2000. Analyst following and count-data econometrics. J. Account. Econ. 30 (3), 351–373.
Romer, P.M., 1986. Increasing returns and long-run growth. J. Polit. Econ. 94 (5), 1002–1037.
Salas, D., Raftery, J.P., 2001. Econometric issues in testing the age neutrality of health care expenditure. Health Econ. 10 (7), 669–671.
Savage, C.J., Vickers, A.J., 2009. Empirical study of data sharing by authors publishing in PLoS journals. PLoS One 4, e7078.
Sheshinksi, E., 1971. Welfare aspects of a regulatory constraint: note. Am. Econ. Rev. 61 (1), 175–178.
Sorenson, O., Fleming, L., 2004. Science and the diffusion of knowledge. Res. Policy 33 (10), 1615–1634.
Steen, R.G., Casadevall, A., Fang, F.C., 2013. Why has the number of scientific retractions increased? PLoS One 8 (7).
Stephan, P.E., 1996. The economics of science. J. Econ. Lit. 34 (3), 1199–1235.
Stephan, P.E., 2012. How Economics Shapes Science. Harvard University Press, Cambridge, MA.
Sukhtankar, S., 2017. Replications in development economics. Am. Econ. Rev.: Papers Proc. 107 (5), 32–36.
Taylor, C.T., Kreisle, N.M., Zimmerman, P.R., 2010. Vertical relationships and competition in retail gasoline markets: empirical evidence from contract changes in Southern California: comment. Am. Econ. Rev. 100 (3), 1269–1276.
Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M., 2011. Data sharing by scientists: practices and perceptions. PLoS One 6, e21101.
Vlaeminck, S., 2013. Data management in scholarly journals and possible roles for libraries – some insights from EDaWaX. Lib. Q. 23 (1), 48–79.
Vlaeminck, S., Hermann, L.K., 2018. Data policies and data archives: a new paradigm for academic publishing in economic sciences? Schmidt, B., Dobreva, M. (Eds.), New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale, Openness and Trust. Proceedings of the 19th International Conference on Electronic Publishing 145–155.
Wagstaff, A., Van Doorslaer, E., 1994. Measuring inequalities in health in the presence of multiple-category morbidity indicators. Health Econ. 3 (4), 281–291.
Wible, J.R., 1991. Maximization, replication, and the economic rationality of positive economic science. Rev. Polit. Econ. 3 (2), 164–186.
Williams, R., 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. Stata J. 12 (2), 308–331.
Zhang, L., Ortmann, A., 2014. The effect of the take-option in dictator-game experiments: A comment on Engel's (2011) meta-study. Exp. Econ. 17 (3), 414–420.