The Economics of Research Reproducibility*

Jean-Edouard Colliard[†] Christophe Hurlin[‡] Christophe Pérignon[§]

January 4, 2022

Abstract: We investigate why economics displays a relatively low level of research reproducibility. We first study the benefits and costs of reproducibility for readers (demand side) and authors (supply side), as well as the role of academic journals in matching both sides. Second, we prove that competition between journals to attract authors can lead to a suboptimally low level of reproducibility. Third, we show how to optimize the costs of reproducibility and estimate that reaching the highest level of reproducibility could cost USD 365 per paper. Finally, we discuss how leading journals can move economics out of a low-reproducibility equilibrium.

Keywords: reproducibility, peer-review process, data and code availability policy, confidential data

JEL Codes: C80, C81, C88.

---

*We are grateful to Bruno Biais, Abel Brodeur, Thierry Foucault, Co-Pierre Georg, Peter Hoffmann, Johan Hombert, Joan Llull, Albert Menkveld, Jean-Charles Rochet, seminar participants at HEC Paris, and participants at the 2021 HEC Liège Research Day and 18th Corporate Finance Day at the Rotterdam School of Management for their comments and suggestions. We thank Olivier Akmansoy and Sébastien Saurin for valuable research assistance. Disclosure: Hurlin and Pérignon are the founders of *cascad* (Certification Agency for Scientific Code And Data, www.cascad.tech), a non-profit academic organisation promoting and verifying research reproducibility.

†HEC Paris, 1 rue de la Libération, 78351 Jouy-en-Josas, France. Email: colliard@hec.fr.

‡University of Orléans, Rue de Blois, 45067 Orléans, France. Email: christophe.hurlin@univ-orleans.fr.

§HEC Paris, 1 rue de la Libération, 78351 Jouy-en-Josas, France. Email: perignon@hec.fr.

Introduction

Like many scientific disciplines, economics has experienced a data and computational revolution. Today, most economics papers are empirical and rely on complex scripts analysing rich and often non-shareable datasets (Christensen and Miguel (2018); Vilhuber (2020)). As computational results now account for a large part of the output and contribution of an academic paper, it is more important than ever that empirical results are computationally *reproducible*. This means that another researcher can regenerate the published quantitative results from a set of files and written instructions provided by the original authors (Kitzes *et al.* (2017)). In short, one can check whether "same data + same code = same results" (Barba, 2018).[1] This step is fundamental to guarantee that empirical results can be trusted by both academics and policymakers. However, in contrast to the swift evolution of research practices, economics journals have slowly evolved on reproducibility, at least until recently. As a result, the average reproducibility level remains low in economics.

In this paper, we aim to understand this situation by conducting a comprehensive study of the *economics of research reproducibility*. Specifically, we see the level of reproducibility as an equilibrium between the supply of reproducibility by authors and the demand by readers, mediated by academic journals. Our study first encapsulates the existing literature in this simple supply and demand framework. We then provide new results on three determinants of reproducibility that have been neglected in the literature: the

[1] In contrast to reproducibility, replicability refers to the ability of a researcher to generate similar results by implementing the same methodology in another context or time period ("same code + new data = same results") or a different methodology on the same data ("new code + same data = same results") (Peng *et al.* (2006)). For examples of replication studies in economics, see McCullough and Vinod (2003); Camerer *et al.* (2016); Hou *et al.* (2018); Drazen *et al.* (2021); Mitton (2021); DellaVigna and Pope (2021).

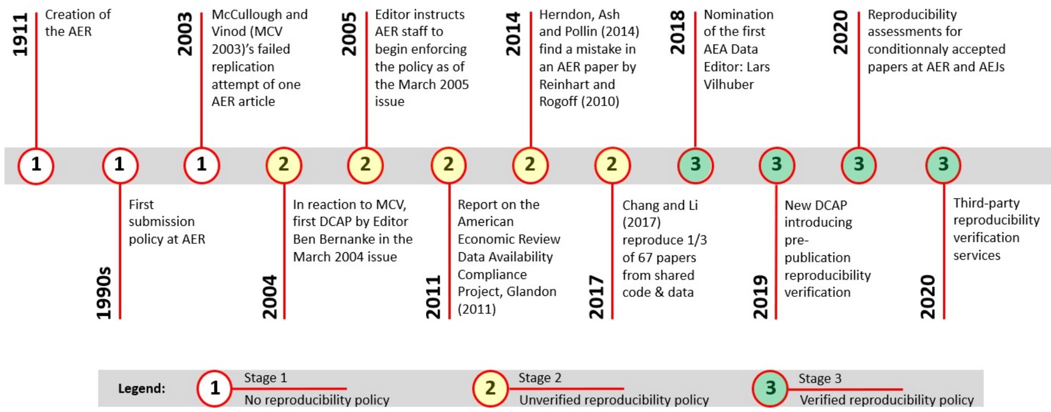

Figure 1: **Making Research Reproducible: The *AER* Timeline**

authors' preferences for reproducibility, the costs journals face when reproducing empirical results, and the journals' incentives to set a high reproducibility level in a theoretical framework. In particular, we show that competition between journals leads to a suboptimally low level of reproducibility, thus calling for remedial actions.

Empirically, we distinguish three levels or stages of reproducibility for a given journal. Stage 1 is the default situation of no reproducibility policy. Stage 2, or unverified reproducibility, is reached when the journal introduces its first Data and Code Availability Policy (DCAP). Stage 3, or verified reproducibility, requires conducting a systematic verification of the reproducibility of all accepted papers prior to publication. Figure 1 illustrates how the *American Economic Review* gradually moved from Stage 1 to Stage 3.

Interestingly, the reproducibility level in economics has not moved monotonically over time, possibly due to shifts in supply or demand. In 1933, Ragnar Frisch, founding editor of Econometrica and first Nobel laureate in economics, wrote that "*in Econometrica the original raw data will, as a rule, be published, unless their volume is excessive. This is important in order to* *stimulate criticism, control, and further studies*" (Frisch, 1933). This rule was later abandoned until pioneer journals such as the *Journal of Money, Credit, and Banking* introduced their DCAP and hence entered Stage 2 as early as in the late 1990's (see Figure 1 in McCullough (2009a)). The *American Economic Review* was the first top-5 economics journal to introduce a DCAP back in 2004 (Bernanke, 2004), whereas the *Quarterly Journal of Economics* was a late adopter in 2016. Despite the generalization of DCAP, the capacity for other researchers to reproduce published results has remained surprisingly low in economics (McCullough *et al.* (2008); Chang and Li (2017); Gertler *et al.* (2018); Herbert *et al.* (2021)). This is because DCAP are often only partially enforced in practice (Duvendack *et al.* (2017)): for many papers numerical resources are unavailable, improperly documented, or of insufficient quality.

One way to improve on this situation is to systematically verify the reproducibility of accepted papers. However, to date only a handful of economics journals have moved to Stage 3 reproducibility. We argue that the sluggish move to Stage 3 reproducibility in economics can be understood as an equilibrium phenomenon. We investigate two driving forces: a "supply effect" based on the high costs of reproducibility for both authors and journals, and a "competition effect" based on journals competing to attract the best papers.

The supply effect has two parts. First, reproducibility imposes high costs on authors. While this point is often mentioned in the literature, the existing evidence is only based on surveys (e.g., Baker (2016); Swanson *et al.* (2020)). To provide more direct evidence, we analyze the propensity of authors of papers accepted in the *Journal of Financial Economics* to publicly share their code and data. We find that only 4.97% of papers have some data or code shared, confirming the view that for most authors the costs of sharing outweigh the benefits. However, we find no evidence that authors are more reluctant to share their code when they are more senior, more cited, or affiliated with more prestigious universities. The evidence thus does not support the received idea that increasing reproducibility standards would discourage the best researchers from submitting their work (Harvey, 2014).

Second, verifying reproducibility imposes high costs on journals, a point which has been neglected in the literature. We use the example of the *cascad* certification agency to provide concrete estimates of the costs of verifying the reproducibility of empirical research in economics.[2] The main components of verification costs are labor costs and the cost of accessing new commercial or administrative datasets. We find that verifying a paper takes on average 10 hours, and estimate that accessing a new dataset costs on average around 5,000 EUR (5,600 USD). As several papers can use the same dataset, economies of scale are large. We estimate that verifying the results of all papers published by ten leading economics journals would lead to an average cost around USD 365 per paper.

We then analyze the competition effect. Following Jeon and Rochet (2010), we model journals as platforms intermediating between authors and readers. Two academic journals choose their submission fees, subscription fees, and a reproducibility level. We show that reproducibility is suboptimally low in equilibrium. The reason is that journals are competing fiercely to attract authors: as each paper can only be submitted to one journal, authors derive market power from the uniqueness of their papers and form a "competitive

[2] *cascad* stands for Certification Agency for Scientific Code And Data and its website is `www.cascad.tech`. It helps (1) individual researchers to signal the reproducible nature of their research by delivering reproducibility certificates and (2) academic journals (e.g. *American Economic Review*, *Economic Journal*) by regenerating all computational results before publication. *cascad* is funded by the French National Center for Scientific Research (CNRS) along with several research institutions and universities. The service is currently free of charge for users.

bottleneck" (Armstrong, 2006). In contrast, readers can subscribe to both journals and competition to attract readers is less fierce. As reproducibility is costly to authors, lowering reproducibility requirements is a way to lure authors from competing journals, which distorts the equilibrium level of reproducibility downwards.[3] Hence, we argue that the low level of reproducibility observed in most economic journals cannot be assumed to optimally balance the supply of and the demand for reproducibility.

An implication of our analysis is that increasing reproducibility in economics research, while desirable, is hard to achieve, since low reproducibility may be an equilibrium outcome. To counter the supply and competition effects, two types of policy actions can be envisioned. Supply policies should aim at optimizing the verification process in order to reduce reproducibility costs for journals. Competition policies may rely on reproducibility standards imposed by journals with sufficient market power, scientific associations, or public funding agencies.

1 Benefits and Costs of Reproducibility

In order to apprehend the economics of reproducibility, we rely on the literature conceptualizing academic journals as platforms intermediating between authors and readers, in particular Jeon and Rochet (2010). We start by considering the case of a single journal, characterized by submission fees on authors p_A, subscription fees on readers p_R, and a level of reproducibility $q \in \mathbb{R}^+$. The three stages of reproducibility defined above can be thought of as three

[3] In Armstrong (2006), competing platforms subsidize the side on which there is a "competitive bottleneck" through prices. In our setup, reproducibility acts as a "quality" variable which gives a second instrument to subsidize one side of the market. To our knowledge, this setup has not been considered in the theoretical literature.

discrete values of q, but we allow for a continuum between those stages.

We assume the journal publishes n_A articles indexed by i, each article being characterized by a vector of characteristics X_i (e.g., topic, quality, authors, etc.). The journal has n_R readers indexed by j, with different characteristics Y_j (e.g., academic or non-academic readers, different tastes or topics of interest, etc.). The point of this section is to summarize the insights the existing literature gives about how reproducibility affects the utility function of both readers and authors.

1.1 Reproducibility for Readers

We denote $u^R(q, X_i, Y_j)$ the utility a reader j derives from having access to article i. The reasons why the level of reproducibility q enters this utility function were already articulated in Ragnar Frisch's 1933 *Econometrica* editorial: to "*stimulate criticism, control, and further studies*".

To fix ideas, we propose a simple specification of u^R that encompasses Frisch's three benefits of reproducibility. Assume that a given paper is correct with probability π and wrong with probability $1 - \pi$. The paper must satisfy the journal's reproducibility policy. This is always the case for correct papers, which are then published. A wrong paper instead does not satisfy the policy with probability $p(q)$, and $p' \geq 0$. With probability $1 - p(q)$ a wrong paper still satisfies the reproducibility policy and gets published. Finally, published papers can later be replicated on different datasets or with different methodologies, which happens with probability $\rho(q)$. Replication attempts always succeed for correct paper and always fail for wrong papers.

We assume that a reader gets a utility $v(q)$ from reading a correct paper, and $-w$ from reading a wrong paper. The reader rationally discards papers that have been shown to be wrong, either because they failed the reproducibility test or because they failed the replication test, but reads the others.[4] In this simple model, the reader will read all the correct papers, but will also read those wrong papers that still passed the reproducibility test and were not considered for replication. Mathematically, the expected utility of the reader for a given paper is:

$$\pi \underbrace{v(q)}_{\text{Further studies}} - (1-\pi)\underbrace{(1-p(q))}_{\text{Control}}\underbrace{(1-\rho(q))}_{\text{Stimulate}} w. \tag{1}$$

This formula encompasses the three dimensions mentioned by Frisch:

Control: Reproducibility serves as a control that the results reported in an article come from the methodology described, which decreases the probability of reading a wrong paper by a factor $1 - p(q)$. In their survey of transparency and reproducibility in economics research, Christensen and Miguel (2018) rightfully claim that this basic standard should be expected of all published economics research, as it is the first step toward a more thorough assessment of the validity of a scientific claim. In particular, reproducibility allows to conduct an in-depth analysis of the code and data, which allows one to spot coding errors (see Herndon *et al.* (2013) on Reinhart and Rogoff (2010)), cases of specification searching and p-hacking (Brodeur *et al.* (2016); Christensen *et al.* (2019b); Brodeur *et al.* (2020); DellaVigna and Linos (2021)), or possible cases of data fraud (Simonsohn (2013)). Moreover, given the high reputation cost for a researcher whose publication is found to be erroneous, requiring data and code to be publicly available should encourage researchers to exert more effort in detecting such errors in the first place (π may thus increase in q).

[4] All the parameters could in principle be functions of X_i and Y_j, but we omit these extra parameters for brevity.

Stimulate: As reproducible research calls for the disclosure of code, it stimulates the criticism of existing results by simplifying the conduct of replication studies running the original methodology on another dataset. Replication decreases the probability of reading a wrong paper by $1 - \rho(q)$, where ρ should be seen as increasing in q. Considering a sample of articles published in top-50 economics journals, Mueller-Langer *et al.* (2019) show that a mandatory data-disclosure policy has a positive effect on the replication probability by six percentage points. They conclude that replication efforts could be incentivized by promoting data disclosure and hence reducing the cost of replication.

Further studies: reproducibility can encourage further studies by communicating more information to the academic community on how exactly to conduct a given analysis. This generates economies of scale as different authors working on the same data do not have to repeat time-consuming procedures, for instance those necessary to clean up the data. Reading a paper known to be correct can hence be more valuable because of reproducibility, or mathematically $v(q)$ is increasing in q.

1.2 Reproducibility for Authors

1.2.1 Theory

The authors of article i derive a certain utility from publishing their article in the academic journal. A quite general formulation for this utility is:

$$B - C^A(q, X_i) + \sum_{j=1}^{n_R} u^A(q, Y_j, X_i) - p^A. \tag{2}$$

B is the benefit of publishing in the journal independently of who reads the published papers (the line on the CV). $C^A(q, X_i)$ is the cost of reaching reproducibility level q for the authors, and is independent of who reads the journal. Finally, u^A is a utility the author derives from the paper being read by the subscribers to the journal (the future citations). Our focus is on how C^A and u^A depend on the level of reproducibility q.

The first and perhaps main cost that $C^A(q, X_i)$ captures is the opportunity cost of the time spent cleaning data, documenting code, and providing technical support to other researchers using the shared material (Miguel (2021)). Indeed, the pressure to publish is indeed ranked first among all the reasons put forward by scientists when surveyed by *Nature* about the impediments to reproducible research (Baker (2016)).

A second important cost is that sharing one's code and data will make the authors face more competition when writing follow-up papers. These costs depend again on an article's characteristics X_i. Some types of research may be particularly discouraged by a strict reproducibility policy, in particular research using proprietary datasets (Harvey, 2014). The costs could also be larger for more productive authors and/or more innovative papers.

A third cost is reputational risk. Making one's code and data available makes it easier for others to spot errors (coding or other), which is socially beneficial but privately costly (Azoulay *et al.* (2017); Jin *et al.* (2019)).

Reproducibility may also have benefits for researchers, that enter authors' utility through the sum of the u^A terms. A first benefit is that journals with a stricter reproducibility policy may attract more readers, and hence publishing in such journals may attract more citations (McCabe and Mueller-Langer (2019); Christensen *et al.* (2019a)). In our framework, journals with a higher q may endogenously have a higher n^R and this benefits authors.

A second benefit is that a high level of reproducibility may signal the high quality of a paper and of its authors to the relevant audience, e.g. peers, universities, and research funding agencies. Lerner and Tirole (2002) argue that this type of signaling motive is an important driver of the open-source software development community. Thus, u^A could be increasing in q.

1.2.2 Empirics

Stodden (2010) reports that almost half of the respondents in a survey state that the lack of incentives and direct benefits is an important reason for researchers not to make their computer code publicly available. However, we are not aware of empirical evidence other than surveys on the benefits for authors of making their research reproducible. To fill this gap, we analyze authors' decisions to voluntarily make their paper reproducible, and how this decision correlates with various authors' characteristics. Our empirical analysis focuses on all papers published in the *Journal of Financial Economics* (JFE) between January 2010 and September 2020, which corresponds to all issues between the first one of volume 95 and the third one of volume 137.[5] We end up with a total number of 1,347 papers written by 2,231 authors.

The JFE is an ideal laboratory to estimate authors' preferences regarding reproducibility. Over the past decade, this journal has encouraged authors to share code and data associated with their papers, but never made it mandatory. This has two important implications. First, there is significant within-journal variation in the reproducibility of papers, so that we can compare reproducible and non-reproducible papers, holding the journal constant. Second, since the policy was not mandatory, it is unlikely that authors with high reproducibility costs self-selected out of the journal.

[5]With a 5.731 impact factor, JFE is ranked in the top-3 in finance and in the top-10 in economics (2019 SCImago Journal Rank). Schwert (2021) reports that 88% of the papers published during our sample period were empirical. Moreover, even theoretical papers commonly use computer code to solve numerical problems or produce theoretical results.

More formally, we can use the framework of the previous subsection. Assume the authors of an accepted paper i with a vector of characteristics X_i can choose $q = 1$ ("open" paper) or $q = 0$ ("closed") paper. According to equation (2), the authors' utility in both cases can be written as:

$$\begin{aligned} U_i^{open} &= B - C^A(1, X_i) + \sum_{j=1}^{n_R} u^A(1, Y_j, X_i) - p^A, & (3) \\ U_i^{closed} &= B - C^A(0, X_i) + \sum_{j=1}^{n_R} u^A(0, Y_j, X_i) - p^A. & (4) \end{aligned}$$

Assume that C^A and u^A are such that the dependence of utilities on authors' observable characteristics is linear, plus a noise term reflecting unobservable characteristics, so that:

$$\begin{aligned} U_i^{open} &= B - p^A + \alpha_O(Y_j) + \beta_O(Y_j)'X_i + \epsilon_{Oi}, & (5) \\ U_i^{closed} &= B - p^A + \alpha_C(Y_j) + \beta_C(Y_j)'X_i + \epsilon_{Ci}. & (6) \end{aligned}$$

As is well known, if ϵ_{Oi} and ϵ_{Ci} follow i.i.d. extreme value distributions, then the probability p_i^{open} that $U_i^{open} \geq U_i^{closed}$, and that the authors hence choose to make their article open, is given by:

$$\ln\left(\frac{p_i^{open}}{1 - p_i^{open}}\right) = [\alpha_O(Y_j) - \alpha_C(Y_j)] + [\beta_O(Y_j) - \beta_C(Y_j)]'X_i \qquad (7)$$

and the coefficients $\alpha = \alpha_O(Y_j) - \alpha_C(Y_j)$ and $\beta = \beta_O(Y_j) - \beta_C(Y_j)$ can be estimated using a logistic regression. Since authors can choose whether to make their article open in the same journal, B, p^A, and Y_j are the same in U_i^{open} and U_i^{closed}, and this allows us to estimate how authors' preferences depend on their observable characteristics.

We collected information about available code and data from the JFE data and program webpage.[6] Over our sample period, 67 published papers or 4.97% of all published papers are open, i.e., having code or data, or both, available for download from the JFE website.[7] This low percentage is in line with previous evidence showing that sharing code and data is not widespread in economics (see Section 2.1). Alternatively, we consider a subsample only including issues with at least one paper that has code and/or data available. In this subsample, there are 544 papers and the fraction of open papers is 12.13%.

We estimate the following logistic regression model:

$$\ln\left(\frac{p_i^{open}}{1-p_i^{open}}\right) = \alpha+\beta_1\cdot Top10_i+\beta_2\cdot International_i+\beta_3\cdot Seniority_i+\beta_4\cdot Citations_i+\varepsilon_i \quad (8)$$

We define the different variables in Table 1. Panel A gives summary statistics for the four explanatory variables, and Panel B reports the results of different regression specifications.

We draw several insights from our analysis. First, as the vast majority of authors choose not to share their code, it seems that for these authors the costs of sharing exceed the benefits. However, there is some heterogeneity, and for 4.97% of paper instead the authors seem to consider that the benefits outweigh the costs. Second, the heterogeneity in this cost-benefit analysis is difficult to explain with proxies for the academic "quality" of authors. Indeed, the decision to share one's code does not seem to depend on the authors'

[6] http://jfe.rochester.edu/data.htm

[7] Information about code and data were retrieved on October 7, 2020. We excluded ten papers with sharable material but published prior to 2010 and did not include ten forthcoming papers with available sharable material but scheduled to be published after the end of our sample period. As we did not check the personal website of all 2,231 authors or other data repositories for downloadable material associated with JFE papers, the reported 4.97% frequency should be seen as a lower bound.

number of citations, their seniority, or affiliation with a top 10 institution.[8] [9] Third, we find that authors affiliated with universities located outside of North-America are significantly more likely to share code or data.

The costs and benefits mentioned in Section 1.2.1 should vary across authors. In particular, one may expect the opportunity costs and reputational costs of sharing data to be larger for more senior and more productive authors. Under this assumption, the fact that the variables $Top10$, $Citations$, and $Seniority$ are not statistically significant could indicate that these costs are not first-order in the decision to share code. Conversely, the impact of $International$ could be consistent with the signaling benefit playing a role: new entrants (i.e., international researchers) send an additional costly signal to a predominantly North-American research community.[10] There are of course alternative interpretations, as $International$ may be correlated with other unobservable author characteristics.

In any case, the lack of statistical significance of variables related to seniority is striking. According to Harvey (2014), finance journals were reluctant to adopt mandatory data sharing policies by fear of losing submissions by the most senior and/or productive authors. Our evidence does not support this view. In columns (5) and (6), we check directly whether the best-cited authors are more reluctant to share their data or code, and again find no effect.

[8]The results on seniority and citations are in line with the survey evidence reported in Swanson *et al.* (2020).

[9]Instead of the mean across coauthors, we also used the median, minimum, and maximum values. We also contrasted researchers with tenure (i.e., $Seniority > 6$ years) and without tenure. The *Citation* variable was used alternatively with and without log transformation. We estimated the regression for North-American researchers only. In all cases, results remained qualitatively unchanged.

[10]Schwert (2021) indicates that over the past decade the percentage of US authors (respectively referees) at JFE was greater than 65% (85%). He also shows in the context of a logit model that US authors have, all else equal, a higher acceptance rate than their international peers.

Table 1: **The Drivers of Code and Data Sharing: Evidence from the *Journal of Financial Economics*.**

Panel A displays the summary statistics of the four explanatory variables. Panel B displays the estimated coefficients and the associated p-values of the logistic regression (8), estimated both with and without year fixed effects (FE). The explained variable *Open* is a binary variable taking the value of one for any paper with code and/or data available, and zero otherwise. *Top10* takes the value of one if at least one coauthor is affiliated with a top-10 university (see Heckman and Moktan (2020)) and *International* equals one if at least one coauthor is affiliated with a university outside North-America. *Seniority* is equal to the time between the year of a given JFE publication and the year when the author got his or her first citation in the *Web of Science* database. *Citations* is the annual number of citations of a given author in *Web of Science*, measured in the year of a given JFE publication. In Panel B, we take the log of 1+Citations. For papers with more than one author, we use the average *Seniority* and *Citations* across all coauthors. The two continuous variables have been manually constructed for all the authors of the 544 papers included in the subsample.

Panel A: Summary Statistics

	Mean	Std-dev	Min	Q1	Median	Q3	Max	Obs.
International	0.318	0.466	0	0	0	1	1	1,347
Top10	0.232	0.422	0	0	0	0	1	1,347
Citations	98.70	226.67	0	17.00	41.42	112.33	3,756.50	544
Seniority	9.14	6.62	-4	4.67	8.50	12.67	45.50	544

Panel B: Logistic Regression

	(1)	(2)	(3)	(4)	(5)	(6)
International	0.5089	0.5306	0.4470	0.4609		
	(0.0658)	(0.0579)	(0.0857)	(0.0817)		
Top10	0.0197	0.0015	-0.0386	-0.0237		
	(0.9527)	(0.9965)	(0.9023)	(0.9406)		
Citations	0.1306	0.1210			0.0793	0.0797
	(0.3334)	(0.3769)			(0.3958)	(0.3951)
Seniority	-0.0098	-0.0061				
	(0.7354)	(0.8348)				
Year FE	No	Yes	No	Yes	No	Yes
Obs.	544	544	1,347	1,347	544	544

2 The Unverified Reproducibility Equilibrium

Having shed some light on the demand for reproducibility by readers and the supply of reproducibility by authors, we turn to the role of journals in matching the two. We summarize the existing evidence about the level of reproducibility in economic journals. Until the recent introduction of systematic pre-publication reproduction of the results by some journals (Vilhuber, 2019), the level of reproducibility has generally been low in economics. We then show theoretically that competition between journals can be expected to lead to a suboptimally low level of reproducibility.

2.1 Empirical Evidence

When they exist, data policies have been only partially enforced.[11] The study of McCullough *et al.* (2008) on several economics journals with compulsory data-sharing policies reveals that the fraction of the papers actually having a data file available in the journal archive ranges from 12% for the *Economic Journal* to close to 100% for the *Journal of Applied Econometrics.* More recently, Vlaeminck and Podkrajac (2017) report a 43.7% average compliance rate for economics papers using non-restricted data. Furthermore, compliance rates also vary through time: for 100% at the *Federal Reserve Bank of St. Louis Review* when the policy was first introduced in 1993 to 50% ten years later. Differently, as of today, all papers published in the *Economic Journal* have their code and data (at least synthetic) publicly available.

Is the unverified-reproducibility policy sufficient to guarantee reproducibility, or do journals need to move to stage 3, i.e., verified reproducibility? This can be tested by checking which fraction of articles in stage-2 journals can

[11] Höffler (2017) reports that 54% of a sample of 343 economics journals have a DCAP.

actually be reproduced from the numerical resources available in the journal's archives. McCullough *et al.* (2006) analyzed the data archive of the *Journal of Money, Credit and Banking* between 1996 and 2003.[12] Out of 193 empirical papers subject to the journal's data policy, only 69 had their code and data publicly available for download. Only 14 papers, or 7.25%, shared material of sufficient quality for McCullough *et al.* (2006) to successfully reproduce the results.[13] In a case study focusing on the *AEJ: Applied Economics* over the 2009-2018 period, Herbert *et al.* (2021) report a 38% replication success rate. Conditional on the data being available, 42% of articles were successfully reproduced, with an additional 43% only partially reproduced.

Is the situation any better for the top-5 economics journals? In 2008, the *American Economic Review* launched an audit to assess the quality of the data and code contained in its online data archive. A replication team selected a sample of 39 empirical articles, out of the 135 published articles subject to the data policy between 2006 and 2008. The results, published by Glandon (2011), indicate an 80% compliance rate with the policy: out of the selected 39 articles, 11 were based on proprietary data and 20 had the appropriate code and data posted. McCullough (2018) draws a less positive conclusion from this audit, noting that Glandon's study actually verifies the results of nine papers, of which only five have been fully reproduced. Chang and Li (2017) study a broader set of 67 articles published in top economics journals. They were able to reproduce the results for one-third of these papers from the code and data available on the journals' repositories. Gertler *et al.* (2018) aim to regenerate the results of 203 empirical papers published in nine leading

[12] See Dewald *et al.* (1986) for an earlier attempt to reproduce empirical articles published in the *JMCB* in which authors were contacted by the replication team to obtain code/data.

[13] See Gertler *et al.* (2018) and Mueller-Langer *et al.* (2019) for more recent estimates of the compliance rate with DCAPs, across a broader set of journals.

economics journals and which did not use any restricted data. They were able to go from the raw data to the final tables/figures for only 14% of the studies.

All these results point toward an unverified reproducibility policy being insufficient. A natural question is whether economics journals should increase their level of reproducibility by moving to Stage 3, and whether they have an incentive to do so. We address this question in the following section.

2.2 Is Low Reproducibility Socially Optimal?

The literature documenting a low level of reproducibility in economics typically implies that this level is inefficient. However, since reproducibility is costly to authors, it is possible that competition between journals leads to the level of reproducibility that equates readers' demand with authors' supply. Observing a low level then simply means that reproducibility is costly for authors.

In this section, we challenge this optimistic view of the low level of reproducibility. To emphasize this point, we show a theoretical example in which journals are competitive and choose a suboptimally low level of reproducibility. Hence, the low level of reproducibility observed in economics may be due to a market failure and call for corrective actions. The market failure at play is that a paper can be published in only one journal, whereas readers can read multiple journals. Authors thus have market power, and competition between academic journals leads them to cater more to the preferences of authors, which they do by lowering the level of reproducibility.

We enrich the setup of Section 1 with a model of competition between two academic journals, indexed by $k \in \{1, 2\}$. A journal serves as a platform for authors and readers, and attracts an endogenous number n_k^A of articles and n_k^R of readers. Reaching reproducibility level q_k has a cost $C^J(q_k, n_k^A)$ for a journal, with $\frac{\partial C^J}{\partial q_k} \geq 0$. Each journal aims at maximizing its "impact", or the number of citations per article,[14] which we assume is proportional to the number of readers n_k^R.[15]

Each journal k simultaneously chooses the authors' fee p_k^A, readers' fee $p_k^R \geq 0$, and reproducibility level $q_k \geq 0$. As in Jeon and Rochet (2010), we can have $p_k^A \leq 0$ (authors being paid to publish) but not $p_k^R < 0$.[16] Each journal faces a break-even constraint:

$$n_k^A p_k^A + n_k^R p_k^R - C^J(q_k, n_k^A) \geq 0. \tag{9}$$

The authors and readers observe (p_1^A, p_1^R, q_1) and (p_2^A, p_2^R, q_2). Each author chooses whether to submit to journal 1, journal 2, or to no journal (a paper cannot be submitted to two journals). For simplicity, we assume that all articles are accepted. Equivalently, we could assume that all articles are of the same quality ex ante and hence have the same probability of acceptance.[17] Each reader chooses whether to subscribe to journal 1, journal 2, both journals, or no journal. All players make their decisions simultaneously.

To keep the model tractable, and in line with the literature on two-sided markets,[18] we reduce the articles' characteristics to a single dimension $x_i \hookrightarrow$

[14]See Card and DellaVigna (2020) for evidence on editors' objectives.

[15]In our setup maximizing the number of readers is equivalent to maximizing the welfare of the journal's readers, see the Appendix A.1. While other specifications are possible, we are adopting the one that seems the least likely to bias the outcome towards a low level of reproducibility.

[16]A journal may attract more readers by subsidizing them. However, the journal cannot control that a reader who gets a subsidy indeed reads the journal, so that negative subscription fees would lead to having "fake" readers.

[17]The growing theoretical literature on academic journals has focused on the screening function of journals, see McCabe and Snyder (2005), Jeon and Rochet (2010), Wang (2018), and Gehrig and Stenbacka (2021). While screening and reproducibility can both be seen as a quality variable, a crucial difference is that a high level of reproducibility imposes a cost on authors and not only on journals.

[18]Most notably Armstrong (2006) (see also Armstrong (2015) for an application to aca-

$\mathcal{U}([0,1])$. More specifically, we assume that the authors of an article i published in journal k obtain:

$$u_1^A(n_1^R, q_1, x_i) - p_1^A = B + \alpha n_1^R - aq_1 - \frac{t}{2}x_i - p_1^A \text{ if } k = 1, \quad (10)$$

$$u_2^A(n_2^R, q_2, x_i) - p_2^A = B + \alpha n_2^R - aq_2 - \frac{t}{2}(1 - x_i) - p_2^A \text{ if } k = 2. \quad (11)$$

This assumption corresponds to a linear Hotelling specification. This is a horizontal differentiation setup, with some authors having a preference for journal 1 and others for journal 2.[19] A higher t implies that the two journals are more differentiated and have more market power over authors.

Symmetrically, readers' characteristics are reduced to a single dimension $y_j \in \mathbb{R}^+$, distributed such that y readers have $y_j \leq y$. Subscribing to journal k gives reader j the payoff:

$$u^R(n_k^A, q_k, y_j) - p_k^R = \beta n_k^A + bq_k - y_j - p_k^R. \quad (12)$$

Note that authors care only about the number of readers, but not about their characteristics, while symilarly readers care only about the number of authors.[20] The parameters α and β measure the sensitivity of each side to the number of agents on the other side. The parameter $a > 0$ measures the cost

demic journals). This section can be seen as an extension of Section 5 in his paper, with reproducibility as an additional "quality" variable chosen by journals. If quality were a characteristic of authors instead, journals would face the problem of excluding some types (as in Hagiu (2009)) or sorting them with prices (as in Damiano and Li (2008)).

[19]We model competition between journals at a similar "level", e.g., the top-5 economics journals. An interesting variant would be to consider a model of vertical differentiation, with one journal more prestigious (higher B) than the other.

[20]The critical assumption here is that the submission decisions of authors are not driven by characteristics that readers also care about. Note that our empirical analysis in Section 1.2.2 does not reject this assumption. Assuming instead for instance that authors with potentially more cited papers are also more sensitive to the level of reproducibility should intuitively reinforce the market failure exhibited in this section.

of reproducibility for authors, and the parameter $b > 0$ the gain for readers.

Finally, we use a simple quadratic specification for the reproducibility cost faced by journals:

$$C^J(q, n_A) = n_A \frac{\kappa}{2} q^2. \tag{13}$$

We make the following assumptions on the parameters:

Assumption 1. $t > \alpha\beta$

Assumption 2. $\beta < \frac{a}{b} < \alpha$

Assumption 3. $\kappa > 2b^2$

Assumption 1 is a stability condition, standard in the literature, that ensures that both journals are sufficiently differentiated to coexist in equilibrium. Assumption 2 means that reproducibility is desirable at least if the journal's costs are null, but not "too desirable". Assumption 3 means that the journal's costs are sufficiently high relative to the benefits of reproducibility for readers. Assumptions 2 and 3 reduce the number of equilibrium configurations to consider.

We look for a subgame perfect Nash equilibrium. Authors and readers choose a journal so as to maximize their utility, rationally anticipating the behavior of other players. The two journals choose their fees and reproducibility levels to maximize their readership, rationally anticipating the future behavior of authors and readers. In addition, journals have to break even. Finally, we restrict our attention to symmetric equilibria with full coverage: in such an equilibrium $(p_1^A, p_1^R, q_1) = (p_2^A, p_2^R, q_2)$ and all authors submit to a journal.

A critical feature of this market is that a given article can only be published in one journal at most ("single-homing"). Hence, an article i is published in journal 1 if and only if $u_1^A(n_1^R, q_1, x_i) - p_1^A \geq \max(0, u_2^A(n_2^R, q_2, x_i) - p_2^A)$, and symmetrically for journal 2. In contrast, readers can subscribe to different journals ("multi-homing"). Reader j subscribes to journal k if and only if $u^R(n_k^A, q_k, y_j) - p_k^R \geq 0$. Since we focus on equilibria with full coverage, the number of articles submitted to each journal for given prices and reproducibility levels is determined by solving for the cutoff type x_i indifferent between both journals. In the Appendix, we solve in closed-form the equilibrium cutoff type, the numbers of readers and authors at both journals, and then the equilibrium choice of prices and reproducibility levels by journals:[21]

Proposition 1. *A symmetric equilibrium with full coverage exists if and only if $t \leq \bar{t}$, where the value of $\bar{t}$ is given in the Appendix. In such an equilibrium both journals choose (p_A^*, p_R^*, q^*), with $p_R^* = 0$, $p_A^* = C^J(q^*, 1)$, and*

$$q^* = \frac{2b(t - \alpha\beta) + \beta(\alpha b - a)}{\beta\kappa}.$$

Is the equilibrium level of reproducibility socially optimal? To answer this question, we consider the program of a social planner who would choose p_k^A, p_k^R, q_k. The planner's objective is to maximize the total number of readers across both journals (which is equivalent to maximizing the total welfare of readers, see the Appendix A.1), under the constraints that both journals break even and all authors submit their paper (full coverage). We obtain the following solution:

Proposition 2. *For any $t \leq \bar{t}$, the social planner implements $(p_A^{**}, p_R^{**}, q^{**})$ in*

[21] In this equilibrium, readers can subscribe to each journal for free ("open access"), and the costs of reproducibility are fully borne by the authors. This result is due to the assumption that $\beta < a/b$ and is not generic. McCullough (2009b) observes that open access journals are less advanced than others in promoting reproducibility. Our model highlights that this may be due to the greater necessity for these journals to attract authors.

both journals, with $p_R^{**} = 0$, $p_A^{**} = C^J(q^{**}, 1)$, *and*

$$q^{**} = \frac{\alpha b - a + \sqrt{(\alpha b - a)^2 + \kappa(2B - t + \alpha\beta)}}{\kappa}. \tag{14}$$

We can now compare the level of reproducibility achieved under competition and with a social planner:

Proposition 3. *For any* $t \leq \bar{t}$, q^* *increases in* t *and* q^{**} *decreases in* t. *Moreover,* $q^* \leq q^{**}$ *with an equality in* $t = \bar{t}$.

Hence, we obtain that the social planner always chooses a higher level of reproducibility than the one we obtain under competition, as illustrated by Figure 2. The intuition is the following. Because the authors are "single-homing", they form what Armstrong (2006) calls a "competitive bottleneck": the journals are competing over attracting the marginal article, whereas for a given number of readers the demand of readers for a journal does not depend on the strategy of the other journal. Since reproducibility is costly to authors, the journals reduce their reproducibility level towards the level favored by authors, even though their objective is to maximize readers' welfare. The social planner instead does not face the bottleneck problem and does not have to leave any surplus to the author with $x_i = 1/2$.[22] As t increases the social planner needs to choose a lower level of reproducibility in order to keep all authors submitting, hence q decreases. On the contrary, under competition as t increases the journals are more and more differentiated and can choose a higher q. As t approaches $\bar{t}$, competition between journals gives close to zero surplus to the author with $x_i = 1/2$. The bottleneck effect disappears and the outcome of competition converges to the social planner's solution.

[22] Similarly, competition between platforms owned by associations in Rochet and Tirole (2003) does not lead to the first-best, due to a business-stealing effect.

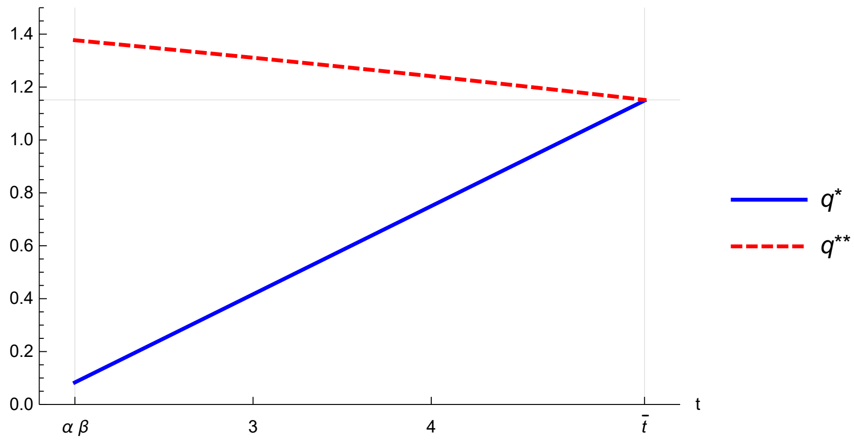

Figure 2: **Reproducibility level under competition (q^*) and under a social planner (q^{**})**

The effect of competition between journals is similar to situations in which competition between producers leads to a suboptimally low quality (e.g., Kranton (2003)). Classical solutions to this problem are the establishment and enforcement of industry standards, in this case a common reproducibility policy across journals, or an initiative to increase reproducibility taken by a journal with sufficient market power (e.g., if one journal had a higher B). As we will now discuss, the top-3 finance journals illustrate the first possibility, and the *American Economic Review* the second.

As explained in Harvey (2014), an initiative to increase reproducibility at the top-3 finance journals emerged in 2010 but was ultimately not adopted. One of the reasons was the need for top 3 finance journals to attract the best finance papers, that could alternatively be submitted to top economics journals. In other words, top finance journals would have lacked the market power to enforce a higher level of reproducibility on authors. In top-5 economics journals instead, the marginal benefit of publishing a paper is so large (Heckman and Moktan (2020), Ductor *et al.* (2020)) that researchers have strong incentives to comply with any standard or disclosing requirement imposed by these journals. Despite the existing competition among top-5 journals, the market power of the *AER* seems to have been high enough to keep its ambitious 2004 data policy, and even taking the next step of verified reproducibility in 2018 for conditionally accepted papers (as shown in Figure 2).[23] Moreover, the 2004 data policy became a standard that has been adopted by other journals. Indeed, all top-5 journals now have a similar data policy, and two of them (*JPE* and *QJE*) explicitly mention that they adopted the AER's 2004 policy.

3 Models and Costs of Reproducibility Verification

The analysis developed so far indicates that the level of the reproducibility in economics can be suboptimally low because of competition between academic journals. As changing the nature of competition between journals is challenging, we show that a more practical way of moving from unverified to verified reproducibility is to optimize the verification process.

3.1 Models of Reproducibility Verification

Stage 3 reproducibility relies on a verification process that takes place before the final acceptance of a manuscript and follows two steps. After verifying that the submitted material complies with a set of guidelines, a reproducibility reviewer regenerates all the results from the code and data of the authors,

[23] Card and DellaVigna (2014) make a similar argument regarding the page limit policy introduced by the *American Economic Review* in 2008 and the *Journal of the European Economic Association* in 2009. The AER's market power was sufficient for authors to choose to comply with the policy, whereas authors became less likely to submit to the JEEA.

verifies that they correspond to the results in the paper, and submits a reproducibility report.[24] This process is easier to achieve if it is conducted by people or organizations with the right expertise and incentives. In theory, three different models could be envisioned:

A first possibility is to *add verification to the tasks of the regular editors and referees* of the journal. However, editors and referees may not have the time, expertise, and data access to check the reproducibility of all accepted papers. Importantly, they also do not have the right incentives. In the suboptimal equilibrium discussed in Section 2.2, even if both journals announce a minimum level of reproducibility $\underline{q}$, each journal has an incentive to renege on this level ex post. Concretely, one can imagine the situation of an editor and referees who have accepted a promising paper for publication after multiple rounds of revision. If at this stage, the editorial team discovers that the data policy is not strictly adhered to by the authors, there seems to be a large cost and little benefit to stop the publication process. On the contrary, the editorial team may rightly consider that the benefit for the journal of publishing an impactful paper will be larger than the cost of not fully enforcing the data policy.

A second possibility is for the journal to *appoint a dedicated editor* in charge of implementing the verification policy, thus avoiding the conflict of objectives that arises when the same editor is in charge of selecting impactful papers and verifying reproducibility.[25] For instance, in 2018 the AEA appointed Lars Vilhuber from Cornell University as the data editor for all the journals operated by the Association. The AEA data editor oversees the Replication Lab at

[24]For such guidelines, see the AEA data standards in Vilhuber (2019) or the constantly updated ones of the Social Science Data Editors (2021).

[25]To the best of our knowledge, the first journal to follow such a policy was *Biostatistics* (Peng (2009); Peng (2011)). For more examples of academic journals verifying computational reproducibility, see Willis and Stodden (2020).

Cornell University, which checks the reproducibility of the results of all conditionally accepted papers. Since then, similar positions have been created at *Review of Economics Studies*, *Economic Journal*, and *Management Science*.

A third possibility for a journal is to use the services of a *trusted third party* dedicated to verifying research reproducibility. The latter can either complement a journal's internal replication team for some types of verification or replace it. An example is the *cascad* certification agency, which conducted 21 verifications for journals managed by the *American Economic Association* in 2020 (Vilhuber, 2021). Another interesting example is the partnership between the *American Journal of Political Science* and the University of North Carolina's Odum Institute (Christian *et al.*, 2018).

3.2 The Case of Non-Shareable Data

The use of non-shareable data is often mentioned as a major impediment to the implementation of reproducible research (Christensen and Miguel, 2018). To summarize the situation, we provide in Figure 3 a flowchart explaining which types of datasets can be shared. In particular, some data directly collected by the researchers and most data obtained from third parties (e.g., confidential administrative data, proprietary data obtained from companies) cannot be put in the public domain without breaching contracts or violating the law.

Without a solution to handle papers using non-shareable data, only two outcomes are possible. A first outcome is *exclusion*: the journal publishes only papers based on non-confidential data, and may have to pass on many interesting papers. For instance, the DCAP of the review *PLOS* states that whenever the data cannot be accessed by other researchers, the manuscript must include an additional analysis based on public data that validates the

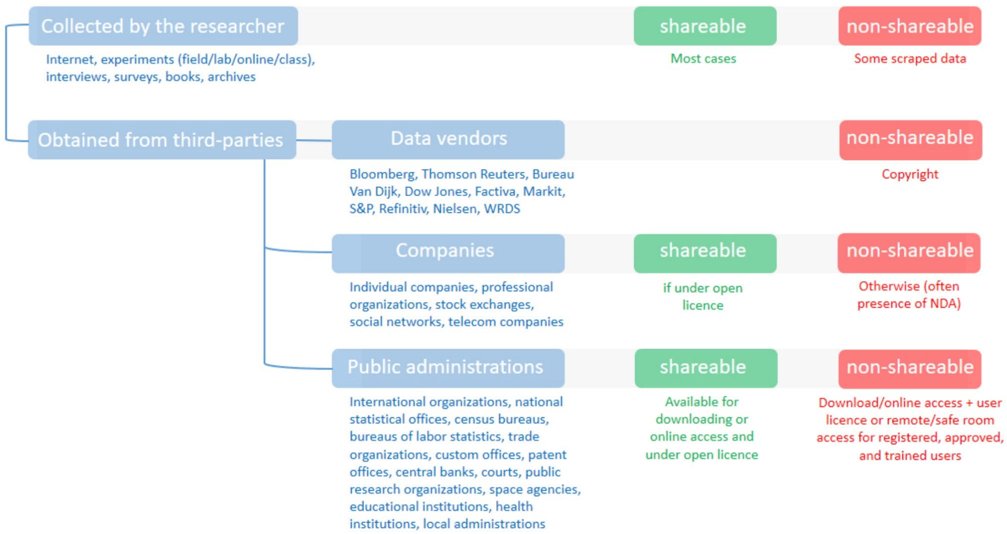

Figure 3: **Data sources and data shareability**

conclusions so that others can reproduce the results.[26]

A second outcome is *exemption*: papers using confidential data are exempted from the DCAP. Nowadays, exemption is the most common policy among economics journals. Among the 49 DCAP considered by Vlaeminck and Herrmann (2015), 34 offer exemptions to the policy for confidential datasets. Christensen and Miguel (2018) show that the share of empirical papers published in the *AER* that fall under these exemptions rose sharply from 10% to around 40% between 2005 and 2015 (also see Vilhuber (2020)).

These numbers show that exclusion is not a realistic possibility in economics. Confidential data on consumers or firms allow researchers to address new research questions or provide innovative answers to traditional ones. It provides those who can access such data with a comparative advantage and increases chances to publish in top journals. Moreover, the extension of the legal frameworks protecting privacy implies that a growing fraction of the data

[26] https://journals.plos.org/plosone/s/data-availability.

used in economics now has to be treated as sensitive.[27]

If the exclusion of papers using confidential data imposes too high a cost on the progress of economics, there is a need to find a reproducibility solution for papers using confidential data. In principle, journal editors or referees could access such data for purpose of reproducibility by following a specific accreditation process for each confidential-data provider. The reason why journals typically do not use this option is that each journal would have to follow an often long and tedious accreditation process for each provider.

Here the use of a third party can prove particularly efficient: once accredited by a confidential-data provider, the third party can use this accreditation for all papers using the data, regardless of the journal they are published in. As an example, *cascad* recently partnered with the Centre d'Accès Sécurisé aux Données (CASD), a public research infrastructure enabling researchers to access individual data from the French Institute of Statistics and Economic Studies (INSEE), and from various French public administrations and ministries (Pérignon *et al.*, 2019). In total, CASD hosts data from 378 sources and offers a data provider service to 742 user institutions. Over 2016-2020, we found on Google Scholar 134 articles acknowledging using CASD data and being published in 91 different academic journals. To verify the reproducibility of all these articles, each of these 91 different journals would have had to go through a lengthy accreditation process to access the same original data.

[27] A recent example is the EU General Data Protection Regulation enforced in 2018. Unlike the European Union, the US does not have a single law on data protection but instead a system of federal and state laws and regulations, including among others the Federal Trade Commission Act, the Financial Services Modernization Act, the Health Insurance Portability and Accountability Act, the Electronic Communications Privacy Act, etc.

3.3 A Quantification of Verification Costs

An obvious way of favoring the move towards verified reproducibility is to decrease its cost, represented by $C^J(q, n_A)$ in our framework. In this section, we summarize some quantitative information regarding this cost and discuss the implications for how to best organize the verification of reproducibility.

We consider a given level of reproducibility $\bar{q}$, corresponding to Stage 3 on the reproducibility scale displayed in Figure 1. We take as given the number n_A of articles to verify, and the number n_D of distinct non-shareable databases to access. The total cost can be represented as:

$$\bar{C}(n_A, n_D) = c_F + n_A(c_L + c_C) + n_D c_D \tag{15}$$

We briefly discuss and give a tentative estimate of each cost. Our estimates are based on the actual experience of the *cascad* verification agency. However, these estimates are necessarily quite rough and are only provided to give an order of magnitude of the costs of verifying reproducibility.

- *Fixed costs* (c_F) reflect the cost of setting up a Swiss-army-knife IT infrastructure allowing the replicating team to run any code provided by an author. This cost includes expenses related to dedicated hardware, storage capacity, cloud resources, software, etc. In addition, the costs include building an online platform allowing data editors to manage manuscripts and to communicate with reviewers and researchers, and covering legal and administrative costs, as well as a fraction of the salary of the two reproducibility editors. Based on the actual expenses faced by *cascad*, we set c_F to 50,000 euros.

- *Labor costs* (c_L) reflect the compensation of the technical staff in charge of checking the compliance of the submitted material to the guidelines, running the code, comparing the results with the ones in the paper, and writing a reproducibility report to be provided to the data editor. By looking at the actual time spent by the reproducibility reviewers in 2020, we set the average number of hours per verification to 10 hours.[28] Given the salaries actually paid by *cascad* in 2020, we use an hourly rate of 15 EUR. Hence c_L is approximately equal to 150 EUR on average.

- *Computing costs* (c_C) need to be paid when the code is run on a commercial cloud. Our estimate of C_C is 11 EUR per article. We estimate the average computing cost by multiplying the average computing time reported in the large-scale reproducibility study conducted by *cascad* on Menkveld *et al.* (2021) by the actual cost per hour of the same machine. The average computing time per paper is 660.53 minutes, or approximately 11 hours, and the cost per hour for a 8vCPU, 64 GB RAM, 1388 GB Temporary storage virtual machine is 1.112 USD, or approximately 1 EUR.[29]

- *Costs of accessing data* (c_D) vary a lot across databases. Many commercial databases are already available "for free" via a campus license. Finding the fee for other commercial databases is easy, but providing the fair monetary cost of establishing a partnership with a restricted-data access center like CASD is almost impossible.[30] Yet, not including it would lead to massively underestimating the cost of setting up a verification service. Averaging across commercial and administrative databases, we use $c_D = 5,000$ EUR as a rough

[28]Our estimate is higher than the 5 hours reported by Vilhuber (2019) at Cornell University. We believe the reason the reviewing time is on the high side at *cascad* is that (1) the proportion of papers using confidential data is larger at *cascad* and (2) the level of compliance of the submitted material with the guideline remains moderate.

[29]Numbers as of November 25, 2021. Source: `https://azure.microsoft.com/en-us/pricing/calculator`.

[30]In the case of CASD, access to the data requires formal approval from the French Statistical Secrecy Committee, which is a 3-6 month process. In some cases, direct access to the data by academic journals would have been simply impossible, as access is restricted to users based in France (e.g., data from the French Tax authority).

and conservative estimate.

We now have estimates for all the parameters of the function $\bar{C}(n_A, n_D)$. Given the large share of data costs in the total reproducibility costs, a critical issue is how many new databases become necessary as the number of papers to be verified grows. The marginal cost of a new paper using a dataset that is not currently available to reviewers is much higher than the cost of a paper using already available datasets. Furthermore, bringing a new data source enriches the data portfolio of the reviewing team, which makes less likely the need to access an additional data source for the next papers to be verified (as the pool of available data sources is now larger).

To estimate how the number of necessary datasets grows with the number of papers, we use the following model. Assume there is a maximum number $\overline{n_D}$ of non-shareable datasets that exist at a given time and can be used in economic research. A fraction $1-\theta$ of papers use shareable data or no data, and a fraction θ pick one dataset at random among the $\overline{n_D}$ of non-shareable datasets. We show in the Appendix that the expected number of distinct non-shareable datasets that will be used by n_A papers is equal to:

$$n_D^*(n_A, \overline{n_D}) = \overline{n_D} \times \left[1 - \left(\frac{\overline{n_D}-1}{\overline{n_D}}\right)^{\theta n_A}\right]. \quad (16)$$

In our quantification exercise below, we use $\theta = 0.4$ (estimate of Christensen and Miguel (2018) for the *American Economic Review*) and tentatively set $\overline{n_D} = 50$. We end up with the following cost function:

$$C^J(\bar{q}, n_A) = \bar{C}(n_A, n_D^*(n_A, \overline{n_D})) = c_F + n_A(c_L + c_C) + c_D n_D^*(n_A, \overline{n_D}). \quad (17)$$

Finally, we take into account that there might be multiple verification teams.

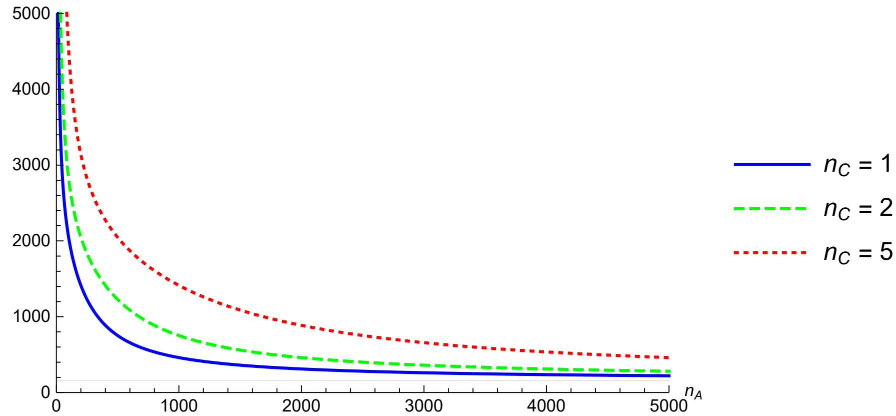

Figure 4: **Average costs per article, as a function of the number of articles n_A and number of verification teams n_C**

If there are n_C such teams and they share the n_A articles to be verified equally, we can compute the average cost per article $AC(n_A, n_C)$ as:

$$AC(n_A, n_C) = \frac{C^J(\bar{q}, n_A/n_C)}{n_A/n_C}. \tag{18}$$

Figure 4 shows the average cost per article for different values of n_C, using our numerical estimates for the parameters. The graph makes it clear that economies of scale are very large, driven both by the fixed cost c_F and the data costs c_D. The economic feasibility of reproducibility thus critically depends on the number of articles n_A.

To give a concrete example, assume we consider verifying all the articles published by the AEA Journals (except JEP, JEL, AEA P&P) and amortize the fixed costs cover three years of activity. According to the 2020 editorial reports, these journals collectively published 344 papers in 2019. Multiplying by three, we reach a total of $n_A = 1,032$ over three years. At this level of activity, the average cost per article is 452 EUR for $n_C = 1$, and 735 EUR for $n_C = 2$, hence economies of scale are still very significant. Assume we go further and the verification team also verifies the papers from the other four top-5 journals. Together, these four journals published 261 papers in 2019, which leads to a total of $n_A = 1,032 + 3 \times 261 = 1,815$ for three years. At this level, the average cost per paper falls to 326 EUR for $n_C = 1$, and 491 EUR for $n_C = 2$. Obviously the costs fall further if one adds other journals or the fixed costs can be amortized over more years, and the average cost reaches 155 EUR per article in the limit.

Assuming there is only one verification team, how large is the cost of 326 EUR $\simeq$ 365 USD (2019) per article? One way to answer this question is to estimate by how much different sources of income for the AEA would have to be increased to compensate the costs of verifying the reproducibility of all articles published by AEA journals.[31] One possibility is to ask the authors of the papers reproduced to pay the cost. This would increase the submission fees for authors of accepted papers from 200 USD to 565 USD, a 183% increase. A second possibility would be to increase the submission fees for all papers submitted. Given a 7% acceptance rate for AEA journals on average, one would need to raise submission fees by $0.07 \times 365 = 25.55$ USD, a 12.8% increase from the current submission fees. Finally, a third possibility would be to raise the costs on readers. According to the financial statements of the AEA, the AEA earned 5.931 million USD in licensing fees and subscription fees for its journals in 2019. In order to absorb an extra cost of $344 \times 365 = 125,560$ USD, these fees would have to increase by 2.12%.

The lesson we draw from this quantitative exercise is that the costs of verified reproducibility are far from negligible, but still manageable if one strives to reduce implementation costs. Our estimates mean that the total cost of verify-

[31] We use numbers from the AEA Financial Statements 2019, available here: https://benny.aeaweb.org/content/file?id=12235.

ing the reproducibility of all articles published in the top-5 economics journals and other AEA journals would be around $(261 + 344) \times 365 \simeq 221,000$ USD per year, close to the average annual salary of one full professor in economics (Scott and Siegfried, 2021). However, this cost assumes operating at scale. This is certainly a factor in the success of the multi-journal policy adopted by the AEA journals. Achieving the same outcome will be more difficult for standalone journals, unless they are able to pool resources or resort to third-party verification.

4 Discussion

4.1 Externalities and Government Intervention

In our framework, the journals internalize the demand for reproducibility coming from "readers", which could be the subscribers of the journals and/or academics who cite the articles published by the journal in their research. However, an article i published in a journal may have an additional social utility, denoted $u^S(q, X_i)$. Even a monopoly journal maximizing the utility of readers will fail to take this additional social utility into account. If u^S is increasing in q, then this is another channel through which the equilibrium level of reproducibility might be suboptimally low.

A traditional externality of research is its application to create new products or processes and generate economic growth. Reproducibility of research may strengthen such spillovers: the availability of the code and data used by academics may greatly facilitate the transformation of scientific discoveries into economic forces. This echoes the "stimulation" function of reproducibility of Ragnar Frisch described in Section 1.1.

In recent years, the demand for evidence-based policymaking has also greatly increased (Foundations for Evidence-Based Policymaking Act of 2018; Cole *et al.* (2020)). As governments and policy institutions increasingly want to base their decisions on academic research, there is logically a demand for this research to be as trustworthy as possible. This echoes the "control" function of reproducibility of Ragnar Frisch discussed in Section 1.1. This function was very prominent during the Covid-19 crisis, with the emphasis put by various public decision-makers on the use of peer-reviewed medical research.

Such externalities may rationalize the use of government intervention to increase the level of reproducibility. While the government has no regulatory power over academic journals, it can considerably influence research policy via its control of public universities and various funding sources. For instance, public research funding agencies now impose open research standards and data management plans for the projects they fund.[32]

How well-intended these policies might be, centralized authorities are likely to be in a worse position than academics themselves to evaluate the costs of reproducibility on researchers and strike the right trade-off. Scientific associations and journals may be seen as "self-regulatory organizations" that set the rules of research independently of the government, but with government intervention remaining a last resort possibility. This implies that academics have a collective interest in setting an appropriate level of reproducibility in research, in order to make additional government intervention unnecessary.

[32] In 2014, the NSF proposed a framework to improve the reproducibility and replicability in funded research, including data sharing policy and data management plans (National Science Foundation, 2014). In the European Union's Horizon 2020 funding program, research data underlying a publication has to be made available, in addition to the requirement to create a data management plan (European Commission, 2020).

4.2 Refereeing

Our finding that competition between journals may lead to a suboptimally low level of reproducibility stands in sharp contrast with the view expressed by Ellison (2002a) and Ellison (2002b) that there exists a "race to the top" in the requirements imposed by referees on authors, leading to longer delays in publication, more robustness checks, longer appendices, etc. However, there is no contradiction between these two mechanisms. The "race to the top" in Ellison (2002a) stems from the behavioral biases of referees, who mistakenly infer from their own submission history that journals require a very high quality of execution. The requirements on reproducibility are for the moment not in the hands of referees but of editors, and it is hard to see how the mechanism on which Ellison (2002a) relies could be transposed at this level.

Conversely, our "competitive bottleneck" mechanism could imply that editors may care less about quality of execution (which may be another interpretation of q in the model) and more about the idea of a paper.

Conclusion

Writing empirical papers in economics takes a great deal of time - literally years of data cleaning, merging, coding, debugging, analyzing and re-analyzing - yet, for decades, data and code played no role in the peer review process. When receiving a new manuscript to review, editors and referees had to assume that the results outlined in the paper were actually resulting from running the researchers' computer code on their data. Over the past 15 years, economics journals have introduced reproducibility policies, but until recently without verification. There is now growing empirical evidence to show that this unverified reproducibility stage led to low compliance rate with DCAP, low quality of the shared resources, and in turn low reproducibility rates.

In this paper, we show that this situation can be the suboptimal outcome of two-sided competition between economic journals. If so, it is possible to improve the situation by increasing the level of reproducibility. Finally, we show that one way to mitigate this market failure is to conduct a systematic verification of the results prior to publication. Such pre-publication verification could be conducted either internally by journals or outsourced to trusted third parties. Interestingly, this is the strategy followed by the *American Economic Association*, as well as by a handful of other leading actors of the scientific publishing industry.

References

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics*, **37** (3), 668–691.

— (2015). Opening access to research. *The Economic Journal*, **125** (586), F1–F30.

Azoulay, P., Bonatti, A. and Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy*, **46** (9), 1552–1569.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, **533** (7604), 452–455.

Barba, L. A. (2018). *Terminologies for reproducible research*. Working paper.

Bernanke, B. S. (2004). Editorial statement. *The American Economic Review*, **94** (1), 404–404.

Brodeur, A., Cook, N. and Heyes, A. (2020). Methods matter: p-hacking and publication bias in causal analysis in economics. *The American Economic Review*, **110** (11), 3634–60.

—, Le, M., Sangnier, M. and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, **8** (1), 1–32.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T. *et al.* (2016). Evaluating replicability of laboratory experiments in economics. *Science*, **351** (6280), 1433–1436.

Card, D. and DellaVigna, S. (2014). Page limits on economics articles: Evidence from two journals. *Journal of Economic Perspectives*, **28** (3), 149–68.

— and — (2020). What Do Editors Maximize? Evidence from Four Economics Journals. *The Review of Economics and Statistics*, **102** (1), 195–217.

Chang, A. C. and Li, P. (2017). A preanalysis plan to replicate sixty economics research papers that worked half of the time. *The American Economic Review*, **107** (5), 60–64.

Christensen, G., Dafoe, A., Miguel, E., Moore, D. A. and Rose, A. K. (2019a). A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLOS ONE*, **14** (12).

—, Freese, J. and Miguel, E. (2019b). *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press.

— and Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, **56** (3), 920–80.

Christian, T.-M. L., Lafferty-Hess, S., Jacoby, W. G. and Carsey, T. (2018). Operationalizing the replication standard. *International Journal of Digital Curation*, **13** (1), 114–124.

Cole, S., Dhaliwal, I., Sautmann, A. and Vilhuber, L. (2020). *Handbook on Using Administrative Data for Research and Evidence-based Policy.*

Damiano, E. and Li, H. (2008). Competing matchmaking. *Journal of the European Economic Association*, **6** (4), 789–818.

DellaVigna, S. and Linos, E. (2021). Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica*, **Forthcoming**.

— and Pope, D. (2021). Stability of experimental results: Forecasts and evidence. *American Economic Journal: Microeconomics*, **Forthcoming**.

Dewald, W. G., Thursby, J. G. and Anderson, R. G. (1986). Replication in empirical economics: The journal of money, credit and banking project. *The American Economic Review*, **76** (4), 587–603.

Drazen, A., Dreber, A., Ozbay, E. Y. and Snowberg, E. (2021). Journal-based replication of experiments: An application to "being chosen to lead". *Journal of Public Economics*, **202**, 104482.

Ductor, L., Goyal, S., van der Leij, M. and Paez, G. N. (2020). *On the Influence of Top Journals*. Working paper.

Duvendack, M., Palmer-Jones, R. W. and Reed, W. R. (2017). What is meant by "replication" and why does it encounter resistance in economics? *The American Economic Review*, **107** (5), 46–51.

Ellison, G. (2002a). Evolving standards for academic publishing: A q-r theory. *Journal of Political Economy*, **110** (5), 994–1034.

— (2002b). The slowdown of the economics publishing process. *Journal of Political Economy*, **110** (5), 947–993.

European Commission (2020). *Reproducibility of scientific results in the EU: scoping report*. Working paper.

Frisch, R. (1933). Editor's note. *Econometrica*, **1** (1), 1–4.

Gehrig, T. and Stenbacka, R. (2021). Journal competition and the quality of published research: Simultaneous versus sequential screening. *International Journal of Industrial Organization*, **76**, 102718.

Gertler, P., Galiani, S. and Romero, M. (2018). How to make replication the norm. *Nature*, **554** (7693).

Glandon, P. J. (2011). Appendix to the report of the editor: Report on the american economic review data availability compliance project. *The American Economic Review*, **101** (3), 696–699.

Hagiu, A. (2009). *Quantity vs. Quality and Exclusion by Two-Sided Platforms*. Working paper.

Harvey, C. R. (2014). Reflections on editing the Journal of Finance, 2006 to 2012. In M. Szenberg and L. Ramrattan (eds.), *Secrets of Economics Editors*, Cambridge: MIT Press.

Heckman, J. J. and Moktan, S. (2020). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, **58** (2), 419–70.

Herbert, S., Kingi, H., Stanchi, F. and Vilhuber, L. (2021). *The Reproducibility of Economics Research: A Case Study*. Working paper.

Herndon, T., Ash, M. and Pollin, R. (2013). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, **38** (2), 257–279.

Höffler, J. H. (2017). Replication and economics journal policies. *The American Economic Review*, **107** (5), 52–55.

Hou, K., Xue, C. and Zhang, L. (2018). Replicating anomalies. *The Review of Financial Studies*, **33** (5), 2019–2133.

Jeon, D.-S. and Rochet, J.-C. (2010). The pricing of academic journals: A two-sided market perspective. *American Economic Journal: Microeconomics*, **2** (2), 222–55.

Jin, G. Z., Jones, B., Lu, S. F. and Uzzi, B. (2019). The Reverse Matthew Effect: Consequences of Retraction in Scientific Teams. *The Review of Economics and Statistics*, **101** (3), 492–506.

Kitzes, J., Turek, D. and Deniz, F. (2017). *The practice of reproducible research: Case studies and lessons from the data-intensive sciences*. University of California Press.

Kranton, R. E. (2003). Competition and the incentive to produce high quality. *Economica*, **70** (279), 385–404.

Lerner, J. and Tirole, J. (2002). Some simple economics of open source. *The Journal of Industrial Economics*, **50** (2), 197–234.

McCabe, M. J. and Mueller-Langer, F. (2019). *Does Data Disclosure Increase Citations? Empirical Evidence from a Natural Experiment in Leading Economics Journals*. Working paper.

— and Snyder, C. M. (2005). Open access and academic journal quality. *The American Economic Review*, **95** (2), 453–458.

McCullough, B. (2009a). Open access economics journals and the market for reproducible economic research. *Economic Analysis and Policy*, **39** (1), 117 – 126.

— (2009b). Open access economics journals and the market for reproducible economic research. *Economic Analysis and Policy*, **39** (1), 117–126.

— (2018). Quis custodiet ipsos custodes? Despite evidence to the contrary, the American Economic Review concluded that all was well with its archive. *Economics : the Open-Access, Open-Assessment e-Journal*, **12**, 1–13.

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

—, — and — (2008). Do economics journal archives promote replicable research? *The Canadian Journal of Economics / Revue canadienne d'Economique*, **41** (4), 1406–1420.

— and Vinod, H. D. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review*, **93** (3), 873–892.

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Neusüss, S., Razen, M. and Weitzel, U. (2021). *Non-standard errors*. Working paper.

Miguel, E. (2021). Evidence on research transparency in economics. *Journal of Economic Perspectives*, **35** (3), 193–214.

Mitton, T. (2021). Methodological Variation in Empirical Corporate Finance. *The Review of Financial Studies*, **Forthcoming**.

Mueller-Langer, F., Fecher, B., Harhoff, D. and Wagner, G. G. (2019). Replication studies in economics - How many and which papers are chosen for replication, and why? *Research Policy*, **48** (1), 62–83.

National Science Foundation (2014). *A Framework for Ongoing and Future National Science Foundation Activities to Improve Reproducibility, Replicability, and Robustness in Funded Research*. Working paper.

Peng, R. D. (2009). Reproducible research and biostatistics. *Biostatistics*, **10** (3), 405–408.

— (2011). Reproducible research in computational science. *Science*, **334** (6060), 1226–1227.

—, Dominici, F. and Zeger, S. L. (2006). Reproducible epidemiologic research. *American Journal of Epidemiology*, **163** (9), 783–789.

Pérignon, C., Gadouche, K., Hurlin, C., Silberman, R. and Debonnel, E. (2019). Certify reproducibility with confidential data. *Science*, **365** (6449), 127–128.

Reinhart, C. M. and Rogoff, K. S. (2010). Growth in a time of debt. *The American Economic Review*, **100** (2), 573–578.

Rochet, J.-C. and Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, **1** (4), 990–1029.

Schwert, G. W. (2021). The remarkable growth in financial economics, 1974-2020. *Journal of Financial Economics*, **140** (3), 1008–1046.

Scott, C. E. and Siegfried, J. J. (2021). American economic association 2020-2021 universal academic questionnaire summary statistics. *AEA Papers and Proceedings*, **111**, 647–49.

Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological science*, **24** (10), 1875–1888.

Social Science Data Editors (2021). Unofficial guidance on various topics by social science data editors. http://social-science-data-editors.github.io/guidance/, Accessed: 2020-11-30.

Stodden, V. (2010). Data sharing in social science repositories: Facilitating reproducible computational research. Communication at the NIPS workshop: Computational Science and the Wisdom of Crowds.

Swanson, N., Christensen, G., Littman, R., Birke, D., Miguel, E., Paluck, E. L. and Wang, Z. (2020). Research transparency is on the rise in economics. *AEA Papers and Proceedings*, **110**, 61–65.

Vilhuber, L. (2019). Report by the AEA Data Editor. *AEA Papers and Proceedings*, **109**, 718–29.

— (2020). Reproducibility and replicability in economics. *Harvard Data Science Review*, **2** (4).

— (2021). Report by the AEA Data Editor. *AEA Papers and Proceedings*, **111**, 808–17.

Vlaeminck, S. and Herrmann, L.-K. (2015). *Data Policies and Data Archives: A New Paradigm for Academic Publishing in Economic Sciences?* Working paper.

— and Podkrajac, F. (2017). Journals in economic sciences: Paying lip service to reproducible research? *IASSIST Quarterly*, **41** (1-4), 16.

Wang, J. (2018). Quality screening and information disclosure in two-sided markets. *Economics Letters*, **171**, 183–188.

Willis, C. and Stodden, V. (2020). Trust but verify: How to leverage policies, workflows, and infrastructure to ensure computational reproducibility in publication. *Harvard Data Science Review*, **2** (4).

A Online Appendix

A.1 Proof that maximizing readers also maximizes welfare

For a given journal k, the number of readers is defined by:

$$n_k^R = \beta n_k^A + bq_k - p_k^R. \tag{A.1}$$

The total reader welfare generated by this journal is thus:

$$\int_0^{n_k^R} u^R(n_k^A, q_k, y)dy \tag{A.2}$$
$$= \int_0^{n_k^R} [\beta n_k^A + bq_k - p_k^R - y]dy \tag{A.3}$$
$$= \int_0^{n_k^R} [n_k^R - y]dy = \frac{(n_k^R)^2}{2}. \tag{A.4}$$

Hence, maximizing the number of readers of a given journal is equivalent to maximizing the reader welfare generated by this journal.

A.2 Proof of Proposition 1

We first solve for the equilibrium numbers of authors and readers on each journal for given fees and quality levels. Given our assumptions, these numbers have to satisfy the following system:

$$n_1^A = \frac{1}{2} + \frac{\alpha(n_1^R - n_2^R) - a(q_1 - q_2) - (p_1^A - p_2^A)}{2t} \tag{A.5}$$
$$n_2^A = 1 - n_1^A \tag{A.6}$$
$$n_1^R = \beta n_1^A + bq_1 - p_1^R \tag{A.7}$$
$$n_2^R = \beta n_2^A + bq_2 - p_2^R \tag{A.8}$$

Solving for this system yields the following equilibrium allocation of authors and readers, for given prices and reproducibility levels:

$$\begin{aligned}
n_1^A &= \frac{1}{2} - \frac{\Delta}{2(t-\alpha\beta)} && \text{(A.9)}\\
n_2^A &= \frac{1}{2} + \frac{\Delta}{2(t-\alpha\beta)} && \text{(A.10)}\\
n_1^R &= \left(\frac{\beta}{2} + bq_1 - p_1^R\right) - \frac{\beta\Delta}{2(t-\alpha\beta)} && \text{(A.11)}\\
n_2^R &= \left(\frac{\beta}{2} + bq_2 - p_2^R\right) + \frac{\beta\Delta}{2(t-\alpha\beta)} && \text{(A.12)}\\
\text{with } \Delta &= p_1^A - p_2^A + \alpha(p_1^R - p_2^R) - (\alpha\beta - a)(q_1 - q_2) && \text{(A.13)}
\end{aligned}$$

Note that Assumption 1 ensures that n_k^A and n_k^R are both decreasing in p_k^A and p_k^R, while Assumption 2 ensures that all else equal n_k^A and n_k^R are both increasing in q_k.

We can now solve for the equilibrium in Step 1. We denote $\tilde{C}^J(q) = C^J(q, 1)$ for convenience. Taking (p_2^A, p_2^R, q_2) as given, we write the following Lagrangian for journal 1:

$$\mathcal{L} = n_1^R + \lambda[n_1^A(p_1^A - \tilde{C}^J(q_1)) + n_1^R p_1^R] + \mu p_1^R + \nu q_1 \tag{A.14}$$

We then differentiate with respect to p_1^A, p_1^R, and q_1:

$$\begin{aligned}
\frac{d\mathcal{L}}{dp_1^A} &= \frac{dn_1^R}{dp_1^A} + \lambda\left[\frac{dn_1^A}{dp_1^A}(p_1^A - \tilde{C}^J(q_1)) + n_1^A + \frac{dn_1^R}{dp_1^A}p_1^R\right] = 0 && \text{(A.15)}\\
\frac{d\mathcal{L}}{dp_1^R} &= \frac{dn_1^R}{dp_1^R} + \lambda\left[\frac{dn_1^A}{dp_1^R}(p_1^A - \tilde{C}^J(q_1)) + n_1^R + \frac{dn_1^R}{dp_1^R}p_1^R\right] + \mu = 0 && \text{(A.16)}\\
\frac{d\mathcal{L}}{dq_1} &= \frac{dn_1^R}{dq_1} + \lambda\left[\frac{dn_1^A}{dq_1}(p_1^A - \tilde{C}^J(q_1)) - n_1^A\tilde{C}^{J\prime}(q_1) + \frac{dn_1^R}{dq_1}p_1^R\right] + \nu = 0 && \text{(A.17)}
\end{aligned}$$

In a symmetric equilibrium, these derivatives have to be zero at the equilibrium prices and reproducibility levels $p_1^A = p_2^A = p^A$, $p_1^R = p_2^R = p^R$, and $q_1 = q_2 = q$. We obtain:

$$
\begin{aligned}
0 &= -\frac{\beta}{2(t-\alpha\beta)} + \lambda\left[\frac{1}{2} - \frac{1}{2(t-\alpha\beta)}(p^A - \tilde{C}^J(q)) - \frac{\beta p^R}{2(t-\alpha\beta)}\right] && \text{(A.18)}\\
0 &= -1 - \frac{\alpha\beta}{2(t-\alpha\beta)} \\
&+ \lambda\left[-\frac{\alpha(p^A - \tilde{C}^J(q))}{2(t-\alpha\beta)} + p^R\left(-1 - \frac{\alpha\beta}{2(t-\alpha\beta)}\right) + \frac{\beta}{2} + bq - p_R\right] + \mu && \text{(A.19)}\\
0 &= b + \frac{\beta(\alpha b - a)}{2(t-\alpha\beta)} \\
&+ \lambda\left[-\frac{\gamma q}{2} + \frac{\alpha b - a}{2(t-\alpha\beta)}(p^A - \tilde{C}^J(q)) + \left(b + \frac{\beta(\alpha b - a)}{2(t-\alpha\beta)}\right)p^R\right] + \nu && \text{(A.20)}
\end{aligned}
$$

In addition, we have the constraints $\lambda \geq 0$, $\mu \geq 0$, $\nu \geq 0$, $n^A(p^A - \tilde{C}^J(q)) + n^R p^R \geq 0$, $p^R \geq 0$, $q^R \geq 0$, $\lambda[n^A(p^A - \tilde{C}^J(q)) + n^R p^R] = 0$, $\mu p^R = 0$ and $\nu q = 0$. Finally, we need to check that the solution satisfies the assumption of full coverage, meaning that an article with $x_i = \frac{1}{2}$ gives positive surplus to its authors, which gives:

$$B + \alpha n^R - aq - p^A - \frac{t}{2} \geq 0. \tag{A.21}$$

We are going to show that under our assumptions any solution to this problem has $\lambda > 0$, $\mu > 0$, and $\nu = 0$. We then solve analytically for this equilibrium, derive the expression of $\bar{t}$ and show that our equilibrium holds if and only if $t \leq \bar{t}$.

Note first that immediately follows from (A.18) that the budget constraint is binding and $\lambda > 0$.

Step 1 - $\nu = 0$: Assume $\nu > 0$ and hence $q = 0$. Then the budget constraint becomes $n^A p^A + n^R p^R = 0$, which gives:

$$p^A = p^R\left[p^R - \frac{\beta}{2}\right]. \tag{A.22}$$

Replacing p^A in (A.20) gives:

$$\nu = -b(1 + \lambda p^R) - \frac{\alpha\beta - a}{2(t-\alpha\beta)}\left(\beta + \lambda p^R(p^R + (\beta/2))\right). \tag{A.23}$$

This quantity is necessarily negative, which is a contradiction. Hence, we cannot have

$q = 0$.

Step 2 - $\mu > 0$: Assume a solution with $\mu = 0$ and $\nu = 0$. We then solve for the system of equations formed by (A.18), (A.19), (A.20), and the binding budget constraint, to be solved in p^A, p^R, q, λ. In particular, we obtain:[33]

$$
\begin{aligned}
q &= \frac{-a(\kappa - b^2)^2 + b\sqrt{X}}{\kappa(\kappa - b^2)(\kappa - 2b^2)} && \text{(A.24)} \\
p^R &= \frac{(\beta\kappa - ab)(\kappa - b^2) - \sqrt{X}}{2\kappa(\kappa - b^2)} && \text{(A.25)} \\
\text{with } X &= (\kappa - b^2)^2[a^2b^2 + \kappa(\kappa - 2b^2)(2t - \alpha\beta)]. && \text{(A.26)}
\end{aligned}
$$

We need both q and p^R to be positive. Given Assumption 3 this is equivalent to having:

$$\frac{a}{b}(\kappa - b^2)^2 \leq \sqrt{X} \leq (\kappa - b^2)(\beta\kappa - ab). \tag{A.27}$$

However, it is easily shown that the left-hand side term is lower than the right-hand side term if and only if $\beta b \geq a$, which violates Assumption 2. Hence, under our parametric assumptions we cannot have a solution with $\mu = 0$.

Step 3 - Candidate solution: we now consider the only remaining candidate solution, which is to have $\lambda > 0$, $\mu > 0$, and $\nu = 0$. We set $p^R = 0$ and $\nu = 0$. The budget constraint then gives us $p^A = \tilde{C}^J(q)$. Replacing p^A with $\tilde{C}^J(q)$, we then solve for the system of equations formed by (A.18), (A.19), (A.20), to be solved in q, λ, μ.

(A.18) immediately gives:

$$\lambda = \frac{\beta}{t - \alpha\beta} > 0. \tag{A.28}$$

We then plug this value of λ into (A.20) and obtain:

[33] There is another solution to the system, in which q has the same expression with a negative coefficient in front of $\sqrt{X}$. q is then obviously negative, so that this solution can be discarded.

$$q = \frac{2b(t - \alpha\beta) + \beta(\alpha b - a)}{\beta\kappa} > 0. \tag{A.29}$$

Finally, we replace λ and q in (A.19) to obtain:

$$\mu = \frac{2t(\kappa - 2b^2) + 2\beta b(a + \alpha\beta) - \kappa\beta(\alpha + \beta)}{2\kappa(t - \alpha\beta)}. \tag{A.30}$$

We need $\mu > 0$, which is equivalent to $t > \underline{t}$, with:

$$\underline{t} = \alpha\beta - \frac{\beta}{2}[(\alpha - \beta)(\kappa - 2b^2) + 2b(a - \beta b)]. \tag{A.31}$$

The term in brackets is positive due to Assumptions 2 and 3. Hence, $\underline{t} < \alpha\beta$. Assumption 1 thus guarantees that $t > \underline{t}$.

Step 4 - Full coverage: The last point to check is the assumption of full coverage. We need to have $t \leq \bar{t}$, where $\bar{t}$ is such that the author of an article with $x_i = 1/2$ makes zero surplus by submitting to a journal. This gives:

$$\begin{aligned}
\frac{\bar{t}}{2} &= B + \alpha\left(\frac{\beta}{2} + bq\right) - aq - p^A && \text{(A.32)}\\
&= B + \frac{\alpha\beta}{2} + (\alpha b - a)\frac{2b(\bar{t} - \alpha\beta) + \beta(\alpha b - a)}{\beta\kappa} - \frac{\kappa}{2}\left(\frac{2b(\bar{t} - \alpha\beta) + \beta(\alpha b - a)}{\beta\kappa}\right)^2. && \text{(A.33)}
\end{aligned}$$

The unique positive root to this equation gives us:

$$\bar{t} = \alpha\beta + \beta\frac{\sqrt{(4b(a - b\alpha))^2 + 32b^2\kappa B + \beta^2\kappa^2} - \beta\kappa}{8b^2} > \alpha\beta. \tag{A.34}$$

In particular, we have $\bar{t} > \alpha\beta$ so that the range of values t such that a symmetric equilibrium with full coverage exists is always non empty.

A.3 Proof of Proposition 2

The social planner chooses p_A, p_R, q symmetrically for both journals under the constraint that an article with $x_i = 1/2$ gets submitted. Since the two journals are symmetric, we can write the planner's Lagrangian as:

$$\begin{aligned}
\mathcal{L} &= n^R + \lambda[n^A(p^A - \tilde{C}^J(q)) + n^R p^R] + \mu p^R + \nu q && \text{(A.35)}\\
&+ \rho[B + \alpha n^R - aq - p^A - (t/2)] && \text{(A.36)}\\
\text{with } n^A &= \frac{1}{2} && \text{(A.37)}\\
n^R &= \frac{\beta}{2} + bq - p^R && \text{(A.38)}
\end{aligned}$$

We then differentiate with respect to p^A, p^R, and q to get:

$$\begin{aligned}
\frac{\lambda}{2} - \rho &= 0 && \text{(A.39)}\\
-1 + \lambda\left[-p^R + \frac{\beta}{2} + bq - p^R\right] + \mu - \alpha\rho &= 0 && \text{(A.40)}\\
b + \lambda\left[-\frac{\kappa q}{2} + bp^R\right] + \nu + \rho(\alpha b - a) &= 0 && \text{(A.41)}
\end{aligned}$$

Assumption 2 and condition (A.41) immediately give us $q > 0$ and hence $\nu = 0$. Using (A.41) again gives us $\lambda > 0$ and hence $\rho > 0$ using (A.39).

We first show that we cannot have a solution with $p^R > 0$. Assume this were the case. Then we have $\mu = 0$. We then solve the system formed by (A.39), (A.40), (A.41), the budget constraint, and the constraint that all articles are submitted, where we need to solve for $p^A, p^R, q, \lambda, \rho$. In particular, we get the following solution for p^R:

$$p^R = \frac{-1}{4\sqrt{\kappa}(\kappa - b^2)}\left(\sqrt{\kappa}[(\alpha-\beta)(\kappa-2b^2) + 2b(a-\beta b)] + (\kappa - 2b^2)\sqrt{(\alpha+\beta)^2 + 4(2B-t)(\kappa-b^2) + 4a(a - b(\alpha+\beta))}\right). \tag{A.42}$$

Under this form, we immediately see that Assumptions 2 and 3 imply that $p^R < 0$. A contradiction.

The only possible solution to the planner's program thus involves $\mu > 0$ and $p^R = 0$. We then need to solve (A.39), (A.40), (A.41), the budget constraint, and the constraint that all articles are submitted, in $p^A, q, \lambda, \mu, \rho$. We obtain:

$$
\begin{aligned}
q &= \frac{\alpha b - a + \sqrt{X}}{\kappa} && \text{(A.43)} \\
p_A &= \frac{\kappa}{2}\left(\frac{\alpha b - a + \sqrt{X}}{\kappa}\right)^2 && \text{(A.44)} \\
\lambda &= \frac{2b}{\sqrt{X}} && \text{(A.45)} \\
\rho &= \frac{b}{\sqrt{X}} && \text{(A.46)} \\
\mu &= \frac{(\kappa - 2b^2)\sqrt{X} + b[(\alpha - \beta)(\kappa - 2b^2) + 2b(a - \beta b)]}{\kappa\sqrt{X}} && \text{(A.47)} \\
\text{with } X &= (\alpha b - a)^2 + \kappa(2B - t + \alpha\beta) && \text{(A.48)}
\end{aligned}
$$

Let us show that this is always a solution when $t \leq \bar{t}$. We first prove that $X > 0$. To see this, denote $\bar{q}$ the value of q^* obtained in Proposition 1 when $t = \bar{t}$. By definition, at this point the author with $x_i = 1/2$ is indifferent between submitting and not submitting, which gives:

$$
2B + \alpha(\beta + 2b\bar{q}) - 2a\bar{q} - \kappa\bar{q}^2 = \bar{t}. \tag{A.49}
$$

Since $t \leq \bar{t}$, we can write:

$$
\begin{aligned}
X &\geq (\alpha b - a)^2 + \kappa(2B - \bar{t} + \alpha\beta) && \text{(A.50)} \\
\Leftrightarrow X &\geq (\alpha b - a)^2 + \kappa(\kappa\bar{q}^2 - 2\bar{q}(\alpha b - a)) && \text{(A.51)} \\
\Leftrightarrow X &\geq (\alpha b - a - \kappa\bar{q})^2 > 0. && \text{(A.52)}
\end{aligned}
$$

Using this result and Assumptions 2 and 3 implies that $\lambda, \mu, \rho, p^A, q$ are all well-defined and positive.

A.4 Proof of Proposition 3

It immediately follows from the analytical expressions of q^* and q^{**} that the former increases in t and the latter decreases in t. For any t, q^{**} is uniquely determined by the condition that an author with $x_i = 1/2$ has zero surplus, with $p^A = \tilde{C}^J(q)$ and $p^R = 0$. Under competition, $\bar{t}$ is defined as the level of t such that $q^*, p^A = \tilde{C}^J(q^*), p^R = 0$ give zero surplus to an author with $x_i = 1/2$. Hence, in that point q^* and q^{**} need to satisfy the same condition, which can admit only one solution. This shows that $q^* = q^{**}$.

A.5 Proof of Equation (16)

For a given n_A, denote X the number of different databases that the n_A papers will use. Denote L_i the event "database i is used by at least one paper". We have $\mathbb{E}[X] = \overline{n_D}\mathbb{E}[L_1]$ and:

$$\mathbb{E}[L_1] = 1 - \Pr(\text{“no paper uses database 1”}) = 1 - \left(\frac{\overline{n_D} - 1}{\overline{n_D}}\right)^{\theta n_A} \tag{A.53}$$

We then obtain equation (16).