The Role of Third-Party Verification in Research Reproducibility

Christophe Pérignon

Column Editor's Note: *Christophe Pérignon, in his contribution to this "Reinforcing Reproducibility and Replicability" column, describes how an innovative institution, called cascad, works. Pérignon, together with collaborator Christophe Hurlin, founded cascad to support researchers in verifying computational reproducibility* before *they would submit to journals. In contrast to Limor Peer, who described in a previous column how her institution provides such services for researchers affiliated with that institution, cascad is offering such reproducibility services to a broad audience in economics and management.*

But cascad also works with journals and data editors (such as me) to expand capacity and, in a particularly innovative twist, to provide specialized fast and knowledgeable access to confidential French administrative data within the Centre d'accès sécurisé aux données (CASD).

In future columns, I will highlight other, emerging methods of providing evidence of computational reproducibility that do not rely on the human-moderated process described by Pérignon. But in the meantime, an ecosystem of like-minded institutions such as cascad *would improve the ability of researchers to demonstrate, and journals to require, reproducibility*.

Keywords: research reproducibility, certification agency, restricted data

---

Background

The quest for a reproducible science requires *three preconditions* to be met, and I believe all three are met today in the field of economics.

The first precondition is to have a good understanding of *what research reproducibility is*. Collectively, the survey of (n.d.-a), the report of the (n.d.-b), and the work of the American Economic Association (AEA; (n.d.-c)(n.d.-d)) brought some much-needed clarity to the different concepts used to describe reanalyses in economics. Currently, the consensus is increasingly favoring the notion that an empirical result is deemed reproducible if it can be recreated by running the original code of the authors on the original data. This type of test contrasts from other forms of reanalyses such as replications, robustness analyses, or extensions (n.d.-e).

The second precondition is *to recognize that the current level of reproducibility is low*. Indeed, there is significant evidence that the success rate of reproducibility studies in economics and finance remains surprisingly low, mainly due to missing

code/data/information and bugs (n.d.-f); (n.d.-g); (n.d.-h); (n.d.-i). Depending on the studies, the success rate ranges between 14% and 52%. This proves the failures of journal policies that only require 'availability on request from authors' or encourage or require authors to post their code and data but do not verify these materials.

The third one is *to acknowledge that this lack of reproducibility is problematic and that we need to act to improve the situation*. Following early decisions by the AEA (n.d.-j) and the Royal Economic Society, most of the other leading scientific associations and academic journals in economics are now strengthening their code and data availability policies, and some regenerate systematically all the results before publication.

Now that these preconditions have been met, the biggest challenge is *implementation*, which is the focus of the rest of this article.

The Advantages of Third-Party Reproducibility Verifications for Journals

As of today, reproducibility verification is mainly conducted by dedicated verification teams working for academic journals or associations, under the supervision of data editors. The verification teams typically include a handful of PhD students and, in some cases, undergraduate students (n.d.-k) or postdocs. It is important to notice that the task of regenerating and checking the results is entirely distinct from the evaluation of a manuscript's scientific merit, conducted by editors and reviewers.

In addition, some journals rely on the help of third-party reproducibility verifiers (e.g., the Certification Agency for Scientific Code and Data [cascad], the Odum Institute at the University of North Carolina). This support can be particularly beneficial in the following situations: (1) when the third party has permanent access to some restricted data (n.d.-l); (2) when the third party can obtain a one-time temporary permission to access some restricted data, sparing the journal's internal team from applying itself; (3) when it owns a license of, or expertise in, a software that the journal does not have; and (4) when the journal lacks staff or computing power to verify all the newly accepted papers.

While it is not currently the case in economics, it is conceivable that in the future, some journals might fully delegate the task of verifying the reproducibility of all accepted papers to an external entity. The main advantage of a centralized system in which a third party works for many journals is the economies of scale generated (see below for collaboration examples). This would lower the average verification cost per paper, which would eventually make verification possible for smaller journals or associations with fewer financial resources.

Either internal or external to a journal, the verifier starts by checking whether the submitted material (all code, data, readme file, manuscript) complies with a set of guidelines. Then, he recreates the same computing environment as the authors (e.g., same operating system, versions of the software, libraries), gets a copy of the various data sets, and attempts to run the code entirely. Like any auditor, the third-party verifier describes in

a report all the steps, actions, and problems faced during the verification process. The report also includes the regenerated results and highlights any discrepancies with those in the manuscript. Specifically, the verifier compares the numerical values of all original and regenerated parameters presented in tables, as well as the level, shapes, and positions of all curves or data points depicted in figures.

Finally, the report is sent to the (data) editor of the journal, who is the one deciding on whether the verification is successful (i.e., the paper can be published) or unsuccessful (i.e., the paper needs to undergo an additional round of verifications). The fact that the data editor makes the final decision is crucial because the third party may not be familiar with the journal's research topics and the research fields' standards. From a legal viewpoint, the third party promises its best efforts, but cannot be held liable for damages if the research turns out not to be computationally reproducible.

The Case of Presubmission Verifications

Regardless of the reproducibility policy of the journals, authors may voluntarily submit their working papers to a third-party verifier before submitting them for publication.

The first reason to conduct a presubmission verification is that it allows the authors to *detect mistakes or inconsistencies* in their analysis. Indeed, when preparing the materials required to request a verification, the authors often identify typos and mistakes, which they can then correct at no cost. Differently, when such mistakes are discovered later in the process, especially after publication, the research community must then discern whether they are honest mistakes or forms of misconduct.

Second, authors often rely on *tacit knowledge* in their research process, not sufficiently encoding and documenting all the steps. The sooner the research is shared with independent parties, who lack this tacit knowledge, the sooner such omissions can be discovered and corrected by the authors.

Third, presubmission verifications help to *build trust*, particularly among coauthors. Indeed, most academic papers have multiple authors, each often specializing in areas where they have comparative advantages. Furthermore, some specialized coauthors may not have the time, nor the skills, to monitor and review tasks outside their area of expertise. In this case, an independent verification provides some reassurance for all the parties involved.

Conducting a presubmission verification does not obviate the necessity for the journal to carry out a prepublication verification. This is because the analysis, code, and data typically undergo significant changes during the review process. However, conducting presubmission verification will greatly ease and speed up the journal's prepublication verification.

The cascad Certification Agency

Christophe Hurlin and I founded cascad (www.cascad.tech) in 2019 with a double objective: (1) to help individual researchers signal the reproducible nature of their research by granting reproducibility certificates and (2) to help other scientific actors (e.g., academic journals, universities, funding agencies, scientific consortia, data providers) verify the reproducibility of the research they publish, fund, or contribute to the production of.

Cascad is a nonprofit research laboratory funded by the French National Center for Scientific Research (CNRS) along with several universities and research institutions. While cascad is based in France, it collaborates with researchers and academic journals from all around the world. Its workforce comprises full-time reproducibility engineers, part-time graduate students, and a group of faculty that oversees the operations and promotes the services offered.

The establishment of cascad was driven by two firm beliefs. First, we believe that for science to be taken seriously, there needs to be a serious commitment to reproducibility. To put it simply, if we want the chain of science to be strong and useful to society, reproducibility should not be its weakest link. Second, we hold the conviction that merely making code and data publicly accessible does not fully address the reproducibility challenge. We have come to this resolute belief after launching and managing RunMyCode (www.runmycode.org), a repository for code and data used by various economics and management journals. In this role, we frequently observed researchers failing to share all the essential components (code, data, explanations) necessary to regenerate their results. This was often due to hurdles such as copyright issues, nondisclosure agreements (NDAs), or concerns related to data privacy. Moreover, even when all components were available, other researchers regularly struggled to execute them, and occasionally failed entirely (for consistent evidence, see (n.d.-m); (n.d.-n); (n.d.-o); (n.d.-p).

Examples of Collaborations

Collaborations with economics journals. Since 2019, cascad has provided verification reports to the data editors of the AEA and the Royal Economic Society. Such verifications concern conditionally accepted articles in one of the 11 journals managed by these two associations (e.g., *American Economic Review*, *American Economic Journal: Macroeconomics*, *Economics Journal*). Initially, the data editor of the AEA contacted cascad to request a verification based on French restricted data, to which he did not have access. However, today cascad often verifies articles based on sharable data. To date, around 80 verifications have been conducted by cascad for these journals.

Collaboration with a restricted data access center. Since 2020, the cascad agency has partnered with the Centre d'accès sécurisé aux données (CASD), a French public research infrastructure that enables researchers to access granular, individual data from the French Institute of Statistics and Economic Studies (INSEE), the Banque de France, and from various French public administrations and ministries. In total, CASD hosts data from 378

sources and offers a data provider service to 742 user institutions. This example allows us to illustrate the economy of scale argument introduced earlier. Indeed, (n.d.-q) found 134 articles on Google Scholar using CASD data, published in 91 different academic journals. To verify the reproducibility of all these articles, each journal would have had to go through a lengthy accreditation process to access the original data. Instead, cascad offers a single point of entry to all academic journals seeking a reproducibility check for articles using restricted data accessed through CASD.

Collaboration with a scientific consortium. In 2021, cascad was tasked with assessing the reproducibility of the empirical results of 168 international research teams, gathered from more than 200 universities, who were participating in the Fincap project (n.d.-r). Each team had to answer the same six research questions using the same data set consisting of 720 million financial transactions. (n.d.-s) showed that running the original researchers' code on the same raw data regenerated exactly the same results only 52% of the time.

The Business Model of Third-Party Verifiers

Launching and operating a third-party reproducibility verification service is costly. (n.d.-t) decomposed the total costs between the fixed costs corresponding to the IT infrastructure (including software) and the variable costs corresponding to labor, computing, and accessing data costs. In a calibration exercise based on the actual number of papers published by 12 leading economics journals, they show that exploiting economies of scale could lower the average verification cost per paper from $763 (separate verification teams) to $330 (one single verification team for the 12 journals).

Our experience at cascad suggests that in addition to accessing restricted data, the most challenging and time-consuming task is to reconstruct the computing environment used by the original authors.[1] Another challenge in practice is to be able to locate the results in the regenerated logfile because a surprisingly large fraction of code still does not automatically generate tables and figures (see (n.d.-u)). These challenges suggest that one way to reduce verification costs is to increase automation in the verification process, raise awareness among researchers, and increase their coding skills.

The question of who should pay for the extra cost associated with reproducibility checks is also key. In the case of voluntary presubmission verifications, it seems natural that the researchers requesting such certification will cover the associated costs. In the case of mandatory prepublication checks, we propose that the cost should be shared between journals and research funding agencies. This subsidy from research funding agencies is justified by the public good and externality effects of producing reproducible research (see (n.d.-v)).

---

[1] The use of containerization, popularly known as *Docker* or *Apptainer*, is not yet widely used in economics (Boettiger, 2015; Clyburne-Sherin et al., 2019).

Conclusion

In this article, we argue that third-party verification services are useful actors in the reproducibility ecosystem. They complement the verification efforts of journals, particularly in research involving restricted data or requiring special skills or computing environments. We claim that for long-term success, third-party verifiers need to automate their labor-intensive processes, exploit economies of scale, and clarify their business models.

Acknowledgments

I thank two reviewers, Olivier Akmansoy, Jean-Edouard Colliard, Christophe Hurlin, Jacques Olivier, and Lars Vilhuber (the Editor) for their comments, suggestions, and support.

Pérignon's original contribution grew out of the oral presentation in Session 6 of the NSF-supported Conference on Reproducibility and Replicability in Economics and the Social Sciences. You can find the presentation on Youtube (archived as (n.d.-w)).

Disclosure Statement

The Conference on Reproducibility and Replicability in Economics and the Social Sciences (CRRESS) webinar series was funded by National Science Foundation Grant #2217493.

References

Boettiger, C. (2015). An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review*, *49*(1), 71–79. https://doi.org/10.1145/2723872.2723882

Chang, A. C., & Li, P. (2017). A preanalysis plan to replicate sixty economics research papers that worked half of the time. *American Economic Review*, *107*(5), 60–64. https://doi.org/10.1257/aer.p20171034

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, *56*(3), 920–980. https://doi.org/10.1257/jel.20171350

Clyburne-Sherin, A., Fei, X., & Green, S. A. (2019). Computational reproducibility via containers in psychology. *Meta-Psychology*, *3*, Article MP.2018.892. https://doi.org/10.15626/MP.2018.892

Colliard, J.-E., Hurlin, C., & Pérignon, C. (2023). *The economics of computational reproducibility* (Working Paper). HEC Paris. https://dx.doi.org/10.2139/ssrn.3418896

Duflo, E., & Hoynes, H. (2018). Report of the search committee to appoint a data editor for the AEA. *AEA Papers and Proceedings*, *108*, 745. https://doi.org/10.1257/pandp.108.745

Gertler, P., Galiani, S., & Romero, M. (2018). How to make replication the norm. *Nature*, *554*(7693), 417–419. https://doi.org/10.1038/d41586-018-02108-9

Herbert, S., Kingi, H., Stanchi, F., & Vilhuber, L. (2023). *The reproducibility of economics research: A case study* (Working Paper No. 853). Banque de France, Working Paper Series. https://www.banque-france.fr/en/publications-and-statistics/publications/reproducibility-economics-research-case-study

Menkveld, A., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Razen, M., Weitzel U., Abad-Diaz, D., Abudy, M., Adrian, T., Ait-Sahalia, Y., Akmanskoy, O., Alcock, J., Alexeev, V., Aloosh, A., Amato, L., Amaya, D., ... Deev, O. (in press). Non-standard errors, *Journal of Finance*.

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press.

Pérignon, C., O. Akmansoy, C. Hurlin, A. Menkveld, Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Razen, M., & Weitzel U. (in press). Computational reproducibility in finance: Evidence from 1,000 tests. *Review of Financial Studies*.

Pérignon, C., Gadouche, K., Hurlin, C., Silberman, R., & Debonnel, E. (2019). Certify reproducibility with confidential data, *Science*, *365*(6449), 127–128. https://doi.org/10.1126/science.aaw2825

Pérignon, C., Greiner, B., Christian, T.-M., & Connolly M. (2023). *Institutional support: How do journal reproducibility verification services work?* CRRESS, Conference on Reproducibility and Replicability in Economics and the Social Sciences. Labor Dynamics Institute. https://doi.org/10.7298/0g2q-d958

Trisovic, A., Lau, M. K., Pasquier T., & Crosas, M. (2022). A large-scale study on research code quality and execution. *Scientific Data*, *9*, Article 60. https://doi.org/10.1038/s41597-022-01143-6

Vilhuber, L. (2020). Reproducibility and replicability in economics. *Harvard Data Science Review*, *2*(4). https://doi.org/10.1162/99608f92.4f6b9e67

Vilhuber, L. (2021). Report by the AEA data editor. *AEA Papers and Proceedings*, *111*, 808–817. https://doi.org/10.1257/pandp.111.808

Vilhuber, L. (2023). Reproducibility and transparency versus privacy and confidentiality: Reflections from a data editor. *Journal of Econometrics*, *235*(2), 2285–2294. https://doi.org/10.1016/j.jeconom.2023.05.001

(n.d.-a).

(n.d.-b).

(n.d.-c).

(n.d.-d).

(n.d.-e).

(n.d.-f).

(n.d.-g).

(n.d.-h).

(n.d.-i).

(n.d.-j).

(n.d.-k).

(n.d.-l).

(n.d.-m).

(n.d.-n).

(n.d.-o).

(n.d.-p).

(n.d.-q).

(n.d.-r).

(n.d.-s).

(n.d.-t).

(n.d.-u).

(n.d.-v).

(n.d.-w).