Brodeur, Abel; Carrell, Scott; Figlio, David; Lusher, Lester

Working Paper

Response to de Gendre and Salamanca's Comment on "Unpacking p-Hacking and Publication Bias"

I4R Discussion Paper Series, No. 306

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Brodeur, Abel; Carrell, Scott; Figlio, David; Lusher, Lester (2026) : Response to de Gendre and Salamanca's Comment on "Unpacking p-Hacking and Publication Bias", I4R Discussion Paper Series, No. 306, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/342311

INSTITUTE for REPLICATION

No. 306
DISCUSSION PAPER SERIES

Response to de Gendre and Salamanca's Comment on "Unpacking p-Hacking and Publication Bias"

Abel Brodeur

Scott Carrell

David Figlio

Lester Lusher

This paper responds to:

De Gendre, A., and N. Salamanca. 2026. Reproduction Report on "Unpacking p-Hacking and Publication Bias". *I4R Discussion Paper Series* No. 305. Institute for Replication.

July 2026

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 306

Response to de Gendre and Salamanca's Comment on "Unpacking p-Hacking and Publication Bias"

Abel Brodeur[1], Scott Carrell[2], David Figlio[3], Lester Lusher[4]

[1]University of Ottawa/Canada and Institute for Replication
[2]University of Texas, Austin/USA and NBER, Cambridge/USA and IZA Bonn/Germany
[3]University of Rochester/USA and NBER, Cambridge/USA and IZA Bonn/Germany
[4]University of Pittsburgh/USA and IZA Bonn/Germany

JULY 2026

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Response to de Gendre and Salamanca's Comment on "Unpacking p-Hacking and Publication Bias"

Abel Brodeur, Scott Carrell, David Figlio, and Lester Lusher

June 2026

Abstract

We are grateful to Alexandra de Gendre and Nicolás Salamanca for their careful reproduction, robustness analyses, and extensions of our article, "Unpacking p-Hacking and Publication Bias." Their comment exemplifies the type of scientific engagement that strengthens research and advances knowledge. We are particularly pleased that they were able not only to reproduce many of our findings, but also to probe their robustness, develop additional statistical tests, and propose new methodological approaches. Such efforts are especially valuable given the unusual data constraints surrounding our study. Because the underlying editorial data are highly confidential, substantial restrictions governed data access and sharing. Nevertheless, we worked to make the project as transparent and reproducible as possible, and we are delighted to see that these efforts enabled a thoughtful and productive independent assessment. We believe the revised comment makes a meaningful contribution to the literature on publication bias, p-hacking, peer review, and research transparency. We clarify a minor issue regarding sample sizes, highlight points of agreement between the comment and our original article, and discuss how this exchange illustrates the value of reproduction and critical scrutiny in empirical economics.

Background

In *Unpacking p-Hacking and Publication Bias* (Brodeur, Carrell, Figlio, and Lusher 2024), we use data from submissions to the *Journal of Human Resources* to identify and unpack publication bias and p-hacking throughout the editorial process. The paper examines how the distribution of test statistics evolves from initial submission through editorial review and publication. We document significant bunching among initial submissions, suggesting that the distribution of published test statistics cannot be fully attributed to publication bias arising during peer review. We further show that desk-rejected manuscripts display greater heaping around

conventional significance thresholds than manuscripts sent for review, implying that manuscripts with marginally significant findings are more likely to be desk rejected. Reviewer recommendations, in contrast, are positively associated with statistical significance. Overall, we find that the peer review process has relatively little effect on the distribution of test statistics. Finally, by tracking rejected manuscripts after they leave the journal, we provide evidence suggesting that publication biases may be less severe than is often feared.

This article is a response to the thoughtful comment by de Gendre and Salamanca, who reproduce, reassess, and extend many of our analyses. Their comment revisits our findings using alternative statistical approaches, conducts a series of robustness exercises, develops additional formal tests, and proposes methodological extensions relevant to the study of publication bias and p-hacking. We are grateful for their careful engagement with our work and view their contribution as an important addition to the growing literature on reproducibility, publication bias, research transparency, and the economics of scientific production.

Response

We would first like to thank de Gendre and Salamanca for the considerable time, effort, and care that went into their reproduction report and comment. They provided an initial version of this comment, which we provided feedback on. Having now read the revised version of the comment, as well as their response to our earlier observations, we are pleased with the final product. The revised comment is careful, balanced, and constructive. It acknowledges areas of agreement while also pursuing important methodological and substantive questions that naturally arise from our study. We believe that this is precisely how scientific debate should proceed.

More broadly, we view their contribution as a success story for open science and research transparency. One of the central motivations behind *Unpacking p-Hacking and Publication Bias* was to create a unique dataset that would allow researchers to study publication bias and p-hacking throughout the editorial process. At the time, little evidence existed regarding how these phenomena evolved between initial submission, peer review, revision, rejection, and eventual publication in economics. The collection and assembly of these data required extensive collaboration with the *Journal of Human Resources* and substantial efforts to ensure that confidentiality obligations were respected.

These confidentiality constraints were unusually severe. Indeed, portions of the underlying data were so sensitive that even a member of the research team (AB) could not directly inspect most manuscripts (i.e., AB could only read accepted manuscripts). Access protocols were designed to protect authors, referees, editors, and the journal itself. As a consequence, there were inevitable limits on what could be included in a public replication package. Despite these challenges, we worked to make available as much data as possible. We therefore take particular satisfaction in seeing researchers successfully engage with the project in a meaningful way. The fact that de Gendre and Salamanca were able to reproduce many of our analyses, identify areas deserving additional scrutiny, conduct robustness exercises, and

develop methodological extensions is, from our perspective, a testament to the value of making research materials available whenever possible, even under difficult circumstances.

Importantly, we do not view reproduction efforts merely as a mechanism for detecting errors. Rather, we see them as a way of deepening our understanding of empirical findings. In this respect, the comment goes well beyond a conventional reproducibility exercise. The authors do not simply ask whether our results can be computationally reproduced; they investigate how sensitive those results are to alternative modeling decisions, examine the behavior of different statistical procedures, reconstruct additional formal tests, and propose new approaches for evaluating the hypotheses at the center of our paper. Whether we, the original authors, ultimately agree with every interpretation is, in some sense, secondary. We will let interested scientists read the original article and comment and make up their own minds. We see our article as mainly descriptive and do not have a stake in any of the results or conclusions.

We are also pleased that the revised version of the comment clearly highlights several areas of agreement between our findings and theirs. In particular, we appreciate their explicit recognition that their analyses support two of the central conclusions of our article: first, that the distribution of desk-rejected manuscripts differs from that of manuscripts that proceed to peer review, and second, that the distributions of test statistics in initial submissions and final accepted versions are remarkably similar. The latter result was, in our view, one of the most important findings of the original paper. The observation that distributions look similar suggests that much of the specification searching and result selection documented in economics likely occurs before manuscripts are submitted to journals.

We would also like to clarify a few minor issues. First, we want to clarify a minor issue regarding sample sizes discussed in the comment. The full sample collected for our study contains 543 papers, consisting of 171 desk-rejected manuscripts, 210 manuscripts rejected after referee reports (including three manuscripts rejected after multiple rounds of review), and 162 first drafts of manuscripts that were eventually published. However, our analyses using Elliott et al.'s tests were conducted only on non-ambiguous estimates from initial submissions.

Applying this restriction removes 12 papers because all of their estimates are classified as ambiguous—six desk-rejected manuscripts and six referee-rejected manuscripts. Restricting attention to initial versions further excludes two referee-rejected papers. Consequently, the Elliott et al. exercises are conducted on a sample of 529 papers rather than the full sample of 543 papers reported in the summary statistics. In short, the summary statistics table includes all papers collected for analysis, whereas our analyses using Elliott et al.'s tests focus on a slightly smaller sample restricted to non-ambiguous estimates from first drafts. We thank the authors for drawing attention to this issue and hope this clarification will be useful for future researchers working with the replication materials.

Second, we thank de Gendre and Salamanca for pointing out and fixing our code for Table 2, row 7, and Table A3, Row 4.

More generally, we believe the comment advances the literature in several important ways. First, it provides an independent assessment of a dataset that is unique within economics. Second, it highlights the strengths and limitations of alternative approaches for studying publication bias and p-hacking. Third, it develops additional statistical tools that future researchers may find useful when analyzing publication processes and editorial decision-making. Fourth, it contributes to the broader conversation about reproducibility and robustness in economics, a conversation that has become increasingly important in recent years. Finally, it demonstrates the scientific value of replication exercises themselves. The report illustrates that reproductions need not be adversarial. They can instead serve as vehicles for methodological innovation, clarification, and cumulative knowledge production.

We therefore hope that researchers interested in publication bias, p-hacking, peer review, research transparency, and reproducibility will read the comment alongside our original article. The two papers should be viewed as complementary contributions. Our article introduced a new dataset and presented the first descriptive evidence on how publication bias and p-hacking evolve throughout the editorial process. The comment carefully reproduces many of those findings, evaluates their robustness, and extends the analysis in new directions. Together, the papers provide a richer and more nuanced understanding of the publication process than either could provide alone.

For this reason, we would encourage future scholars who cite our article to also engage with and cite the accompanying comment. The comment contains important methodological contributions, valuable robustness analyses, and useful extensions that deserve recognition in their own right. We expect that future research in this area will benefit substantially from considering both pieces together.

In closing, we would like to reiterate our appreciation to Alexandra de Gendre and Nicolás Salamanca for their excellent work.