# *Reproduce to Validate: a Comprehensive Study on the Reproducibility of Economics Research*
# *Online Appendix*

Sylvérie Herbert
*Banque de France*

Hautahi Kingi
*Google*

Flavio Stanchi
*Airbnb*

Lars Vilhuber*
*Labor Dynamics Institute, Cornell University*

Corresponding author: lars.vilhuber@cornell.edu.

.

## Appendix A1: Appendix

### *A1. Appendix Tables*

TABLE A1
Assignments by publication year

| | Assignment Years | | | | |
|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 |
| 2009 | 0 | 23 | 3 | 1 | 0 |
| 2010 | 32 | 6 | 1 | 0 | 0 |
| 2011 | 36 | 9 | 1 | 0 | 0 |
| 2012 | 0 | 41 | 1 | 5 | 0 |
| 2013 | 0 | 12 | 1 | 0 | 0 |
| 2014 | 0 | 0 | 41 | 5 | 0 |
| 2015 | 0 | 0 | 0 | 24 | 0 |
| 2016 | 0 | 0 | 0 | 36 | 0 |
| 2017 | 0 | 0 | 0 | 43 | 0 |
| 2018 | 0 | 0 | 0 | 20 | 1 |

TABLE A2
Assessment of Data Availability, By Year

| Reason | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total | Percent |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Confidential Data | 5 | 10 | 8 | 11 | 1 | 17 | 2 | 12 | 12 | 8 | 86 | 31.39 |
| Data was Provided | 14 | 13 | 24 | 26 | 7 | 23 | 18 | 20 | 26 | 9 | 180 | 65.69 |
| No Data or Reason | 2 | 0 | 0 | 2 | 0 | 0 | 1 | 1 | 1 | 1 | 8 | 2.92 |
| Total | 21 | 23 | 32 | 39 | 8 | 40 | 21 | 33 | 39 | 18 | 274 | 100 |

Notes: Assessments made by replicators using the entry questionnaire, prior to attempting reproduction.

TABLE A3
Reproduction Results

| Year | Successful | Partial | Confidential Data | Other failure | Total |
|---|---|---|---|---|---|
| 2009 | 4 | 4 | 5 | 1 | 14 |
| 2010 | 7 | 3 | 2 | 1 | 13 |
| 2011 | 10 | 2 | 5 | 7 | 24 |
| 2012 | 8 | 13 | 2 | 3 | 26 |
| 2013 | 4 | 3 | | | 7 |
| 2014 | 7 | 11 | 3 | 2 | 23 |
| 2015 | 4 | 12 | 2 | | 18 |
| 2016 | 8 | 7 | 4 | 1 | 20 |
| 2017 | 13 | 8 | 3 | 2 | 26 |
| 2018 | 3 | 3 | 2 | 1 | 9 |
| Total | 68 | 66 | 28 | 18 | 180 |
| Percent | 37.8% | 36.7% | 15.6% | 10% | 100% |

TABLE A4
Probit: Determinants of Reproducibility, Year 0

| | Outcome: Full or Partial Reproduction | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| 'Avg. H-index' | 0.0004 | | 0.081 |
| | (0.021) | | (0.099) |
| 'Max H-index' | | −0.002 | −0.032 |
| | | (0.011) | (0.039) |
| 'Highest Institution Publications' | −0.011 | −0.012 | |
| | (0.009) | (0.009) | |
| 'Min H-index' | | | −0.021 |
| | | | (0.067) |
| 'Institution Publications (top)' | | | −0.382 |
| | | | (0.596) |
| 'Institution Publications (bottom)' | | | 4.240 |
| | | | (349.000) |
| 'Highest Co-author Experience' | −0.024 | −0.023 | −0.026 |
| | (0.019) | (0.018) | (0.023) |
| 'Number of authors' | 0.136 | 0.147 | 0.395 |
| | (0.177) | (0.188) | (0.279) |
| 'Author at US university' | 0.273 | 0.278 | 0.057 |
| | (0.476) | (0.477) | (0.479) |
| 'Solo-authored' | | | 0.856 |
| | | | (0.715) |
| Constant | 1.680*** | 1.670*** | 0.823 |
| | (0.529) | (0.535) | (0.946) |
| $N$ | 113 | 113 | 113 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Sample restricted to articles published in 2012 and later, with attempted reproduction.

TABLE A5
Publication and Author Metrics (2012+)

| | Confidential data | Unsuccessful | Partial | Successful |
|---|---|---|---|---|
| Avg h-index | 12.06 | 14.67 | 11.68 | 14.04 |
| Lowest h-index | 6.21 | 7.33 | 6.14 | 7.4 |
| Number of Authors | 2.27 | 2.44 | 2.32 | 2.6 |
| Citations | 2.87 | 2.11 | 3.58 | 5.47 |
| Highest experience | 20.57 | 24.89 | 19.28 | 21.53 |
| Institutional productivity | 19.63 | 27.85 | 18.48 | 21.47 |
| Percent of authors in US | 82.14 | 77.78 | 71.93 | 82.98 |
| N | 84 | 9 | 57 | 47 |

Notes: All assessed articles published after 2011, except for 1 author dropped for inconsistent OA data.with . Author and institutional characteristics are measured in the year of publication.
Institutional (cumulative) productivity measured in 10,000 publications.

**Appendix A2: Sample directory change**

FIGURE A1 Example of change to author-provided file

```
> svn diff -r961:HEAD $SVNURL/10.1257/app.5.4.92/replication-xxx/Data/
 do_files/main_analysis.do
Index: main_analysis.do
===================================================================
--- main_analysis.do (revision 961)
+++ main_analysis.do (revision 3425)
@@ -6,7 +6,7 @@
     version 11.2

 *place path here:
-global path "C:\"
+global path "\\rschfs1x\usercl\spring\xxx\Data"
 cd            "$path\output"
 use "$path\data\thefts_sales.dta",clear
 cap log close
```

## Appendix A3: Comparison of bibliometric sources

### *A3. Web of Science*

In an earlier version of this paper, we had used data from Web of Science (**?**, WoS). We manually queried the WoS database citations for each article by entering the partial DOI of articles in an issue of AEJ:AE. For later issues, the DOI structure changed, and alternate search criteria were used. For each search results, WoS provided year-by-year citations, as well as total and average citations. We did this in 2015, 2017, and after the conclusion of the exercise, in 2019.

For each of up to five authors per article, we also (contemporaneously) queried WoS, searching for that author, and recording their h-index (**?**) and the underlying number of citations for each author by year, as well as the search criteria used to find the author. In some cases, a simple search by author name does not yield a unique person (e.g., "Smith, Adam"), and sometimes, the metadata in Web of Science contained errors.

We adjusted total citations as reported for years since publication (which differed for each issue), and used that adjusted number in our analysis. Results in **??** relied on these numbers. By design, we excluded several years worth of articles from these analyses, since not enough time had elapsed since the article had been published.

### *A3. OpenAlex*

In response to a referee, we investigated expanding the measurement again, for a larger set of articles over a larger expanse of time. The very manual process combined with an absence of research participants and authors working on this project suggested an alternative approach, which had become feasible in the interim, thanks to the efforts of **?**. We therefore queried the OpenAlex (OA) corpus (**?**) via the openly accessible **API!** (**API!**). Code to do so is included in the replication package.

Generically, OA has more works than WoS. To facilitate the comparison, when computing as-of-year h-index, we remove a few types of works that are not usually encountered on WoS or for that matter on Google Scholar (GS): "`other`", "`paratext`", "`peer-review`", "`reference-entry`".

For the 1114 authors in our sample, we identified 506 . The median institution is recorded as having published 16684 as of 2023. Cornell University, an R1 research university, is recorded as having published 334943 , whereas Wellesley College, a liberal arts college focused on teaching undergraduates, is listed with 9023 works.

### *A3. Comparison*

No two bibliometric data sources are identical, and we observe certain differences in these two databases as well. This is also true of different

snapshots of the same database over time, even for data that ostensibly refers to citations occurring several years back. Thus, variability is to be expected.

In this section, we compare our subsample of the two databases, for the same articles and authors, where possible. Table **??** is for the smaller sample with manually collected WoS data. Table **??** reproduces Table 11 in the main text, matched to the WoS extract, but collected at a later stage.

In general, the absolute level of citations and derivative h-indexes are much higher in OA. In both databases, however, the approximate relative levels are similar.

TABLE A6
Publication and Author Metrics, WoS

| | Unsuccessful | Partial | Successful |
|---|---|---|---|
| Avg h-index | 7.14 | 7.22 | 7.78 |
| Lowest h-index | 5.07 | 4.25 | 4.51 |
| Number of Authors | 2.14 | 2.4 | 2.58 |
| Citations | 4.04 | 3.6 | 5.26 |
| N | 14 | 52 | 45 |

Notes: Articles with attempted reproduction.
Bibliometric data manually queried from WoS.

TABLE A7
Publication and Author Metrics for WoS sample

| | Unsuccessful | Partial | Successful |
|---|---|---|---|
| Avg h-index | 20.76 | 17.84 | 19.65 |
| Lowest h-index | 14.57 | 10.38 | 10.73 |
| Number of Authors | 2.14 | 2.42 | 2.58 |
| Citations | 7.07 | 8.56 | 12.04 |
| Highest experience | 25.79 | 23.4 | 25.27 |
| Institutional productivity | 26.57 | 18.12 | 23.2 |
| Percent of authors in US | 71.43 | 73.08 | 82.22 |
| N | 14 | 52 | 45 |

Notes: Assessed articles matched to Web of Science database extract. Author and institutional characteristics are measured 4 years after publication. Institutional (cumulative) productivity measured in 10,000 publications.

## Appendix A4: Summary statistics

Tables **??** and **??** present the summary statistics for the two main samples used in our regressions. The first one concerns assessed articles, based on complete entry questionnaires, and the second are articles for which reproduction was attempted.

TABLE A8
Summary statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Fully reproduced | 180 | 0.378 | 0.486 | 0 | 1 |
| Full or Partial | 180 | 0.761 | 0.428 | 0 | 1 |
| AsinH(YTD citations) | 273 | 4.040 | 0.886 | 0.000 | 6.540 |
| Avg. H-index | 273 | 16.400 | 11.000 | 1.000 | 63.000 |
| Max H-index | 273 | 24.800 | 20.400 | 1 | 133 |
| Min H-index | 273 | 9.360 | 7.340 | 1 | 52 |

Notes:
Full or partial reproduction are defined in the text.
Results for all articles with complete assessment.

TABLE A9
Summary statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Fully reproduced | 180 | 0.378 | 0.486 | 0 | 1 |
| Full or Partial | 180 | 0.761 | 0.428 | 0 | 1 |
| AsinH(YTD citations) | 180 | 4.060 | 0.875 | 0.881 | 6.540 |
| Avg. H-index | 180 | 16.800 | 11.100 | 1.000 | 63.000 |
| Max H-index | 180 | 25.500 | 20.900 | 1 | 133 |
| Min H-index | 180 | 9.710 | 7.630 | 1 | 52 |

Notes:
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.

## Appendix A5: Alternative specifications

### *A5. Alternative specifications of the link between citation count, reproducibility and hindex*

*A5.1. Levels, log, and Poisson regressions*

TABLE A10
OLS: YTD Citations on Reproduction Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.608 | | | 4.500*** |
| | (0.373) | | | (1.690) |
| 'Max H-index' | | 0.208 | | −1.870** |
| | | (0.194) | | (0.740) |
| 'Full or Partial' | | | | −13.800 |
| | | | | (13.700) |
| 'Min H-index' | | | 0.572 | −1.700 |
| | | | (0.533) | (1.340) |
| 'Fully reproduced' | −11.500 | −11.200 | 8.230 | 0.477 |
| | (11.600) | (10.400) | (10.400) | (13.000) |
| 'Highest Institution Publications' | 0.163 | 0.141 | 0.161 | 0.115 |
| | (0.161) | (0.162) | (0.169) | (0.165) |
| 'Highest Co-author Experience' | −0.098 | 0.028 | 0.196 | −0.217 |
| | (0.338) | (0.332) | (0.339) | (0.339) |
| 'Number of authors' | 9.400** | 7.640** | 10.400*** | 13.100*** |
| | (3.630) | (3.710) | (3.850) | (4.110) |
| 'Author at US university' | 3.680 | 4.440 | 5.190 | 3.150 |
| | (7.700) | (7.730) | (8.000) | (7.720) |
| 'Solo-authored' | 11.500 | 10.900 | 8.600 | 21.000* |
| | (10.100) | (10.100) | (10.500) | (10.800) |
| 'Avg. H-index':'Fully reproduced' | 1.420** | | | 0.686 |
| | (0.574) | | | (0.671) |
| 'Max H-index':'Fully reproduced' | | 0.976*** | | |
| | | (0.326) | | |
| 'Min H-index':'Fully reproduced' | | | 0.466 | |
| | | | (0.846) | |
| 'Max H-index':'Full or Partial' | | | | 0.469 |
| | | | | (0.384) |
| 'Full or Partial':'Min H-index' | | | | −0.452 |
| | | | | (1.070) |
| Constant | −2.850 | 3.060 | −7.530 | −6.180 |
| | (14.200) | (13.900) | (15.000) | (16.600) |
| Observations | 180 | 180 | 180 | 180 |
| Adjusted $R^2$ | 0.159 | 0.149 | 0.089 | 0.172 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.

TABLE A11
OLS: Log Citations on Reproduction Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.023*** | | | 0.104*** |
| | (0.007) | | | (0.039) |
| 'Max H-index' | | 0.009** | | −0.034** |
| | | (0.004) | | (0.015) |
| 'Min H-index' | | | 0.023** | −0.042* |
| | | | (0.010) | (0.025) |
| 'Fully reproduced' | 0.132 | 0.094 | 0.143 | 0.171 |
| | (0.225) | (0.204) | (0.197) | (0.231) |
| 'Highest Institution Publications' | 0.003 | 0.002 | 0.004 | 0.003 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| 'Highest Co-author Experience' | −0.007 | −0.004 | −0.003 | −0.009 |
| | (0.007) | (0.007) | (0.006) | (0.007) |
| 'Number of authors' | 0.185*** | 0.143* | 0.221*** | 0.249*** |
| | (0.070) | (0.073) | (0.073) | (0.079) |
| 'Author at US university' | 0.049 | 0.074 | 0.070 | 0.036 |
| | (0.149) | (0.151) | (0.151) | (0.149) |
| 'Solo-authored' | 0.080 | 0.033 | −0.012 | 0.187 |
| | (0.196) | (0.198) | (0.198) | (0.205) |
| 'Avg. H-index':'Fully reproduced' | 0.004 | | | −0.094 |
| | (0.011) | | | (0.065) |
| 'Max H-index':'Fully reproduced' | | 0.005 | | 0.040 |
| | | (0.006) | | (0.030) |
| 'Min H-index':'Fully reproduced' | | | 0.005 | 0.056 |
| | | | (0.016) | (0.038) |
| Constant | 2.570*** | 2.740*** | 2.530*** | 2.370*** |
| | (0.274) | (0.272) | (0.283) | (0.295) |
| Observations | 180 | 180 | 180 | 180 |
| Adjusted $R^2$ | 0.159 | 0.127 | 0.131 | 0.166 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
Log(YTD citations) is computed as log(YTD Citations + 1).
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.

TABLE A12
Poisson: YTD Citations on Reproduction Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.014*** | | | 0.093*** |
| | (0.001) | | | (0.014) |
| 'Max H-index' | | 0.005 | | −0.032*** |
| | | (0.014) | | (0.004) |
| 'Min H-index' | | | 0.014*** | −0.039*** |
| | | | (0.005) | (0.005) |
| 'Fully reproduced' | −0.084* | −0.110 | 0.264** | 0.045 |
| | (0.046) | (−0.084) | (−0.110) | (0.264) |
| 'Highest Institution Publications' | 0.004*** | 0.003 | 0.004 | 0.003 |
| | (0.001) | (0.004) | (0.003) | (0.197) |
| 'Highest Co-author Experience' | −0.002 | 0.001 | 0.005*** | −0.004 |
| | (0.001) | (−0.002) | (0.001) | (0.141) |
| 'Number of authors' | 0.186*** | 0.153 | 0.197 | 0.242** |
| | (0.012) | (0.186) | (0.153) | (0.098) |
| 'Author at US university' | 0.115*** | 0.136 | 0.141 | 0.105*** |
| | (0.033) | (0.115) | (0.136) | (0.004) |
| 'Solo-authored' | 0.185*** | 0.169 | 0.098 | 0.330 |
| | (0.042) | (0.185) | (0.169) | (0.346) |
| 'Avg. H-index':'Fully reproduced' | 0.020*** | | | −0.041*** |
| | (0.002) | | | (0.008) |
| 'Max H-index':'Fully reproduced' | | 0.014 | | 0.025 |
| | | (0.020) | | (0.262) |
| 'Min H-index':'Fully reproduced' | | | 0.004 | 0.022*** |
| | | | (0.014) | (0.003) |
| Constant | 2.690*** | 2.830 | 2.610 | 2.470 |
| | (0.057) | (2.690) | (2.830) | (2.610) |
| Observations | 180 | 180 | 180 | 180 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.
Heteroskedasticity-robust standard errors.

*A5.1. Full and partial reproducibility as outcome*

TABLE A13
OLS: Arcsin Citations on Reproduction Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.020* | | | 0.244*** |
| | (0.011) | | | (0.090) |
| 'Max H-index' | | 0.006 | | −0.096*** |
| | | (0.006) | | (0.036) |
| 'Min H-index' | | | 0.028* | −0.106* |
| | | | (0.017) | (0.054) |
| 'Full or Partial' | −0.104 | −0.104 | 0.076 | 0.138 |
| | (0.259) | (0.229) | (0.248) | (0.276) |
| 'Highest Institution Publications' | 0.003 | 0.002 | 0.004 | 0.004 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| 'Highest Co-author Experience' | −0.008 | −0.004 | −0.003 | −0.010 |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| 'Number of authors' | 0.198** | 0.154* | 0.233*** | 0.293*** |
| | (0.076) | (0.079) | (0.079) | (0.084) |
| 'Author at US university' | 0.071 | 0.099 | 0.084 | 0.038 |
| | (0.158) | (0.161) | (0.162) | (0.158) |
| 'Solo-authored' | 0.057 | 0.002 | −0.052 | 0.214 |
| | (0.210) | (0.213) | (0.212) | (0.218) |
| 'Avg. H-index':'Full or Partial' | 0.008 | | | −0.184** |
| | (0.012) | | | (0.093) |
| 'Max H-index':'Full or Partial' | | 0.005 | | 0.080** |
| | | (0.007) | | (0.037) |
| 'Min H-index':'Full or Partial' | | | −0.002 | 0.094* |
| | | | (0.019) | (0.057) |
| Constant | 3.300*** | 3.480*** | 3.140*** | 2.790*** |
| | (0.328) | (0.322) | (0.327) | (0.368) |
| Observations | 180 | 180 | 180 | 180 |
| Adjusted $R^2$ | 0.147 | 0.111 | 0.119 | 0.170 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.

TABLE A14
OLS: YTD Citations on Reproduction Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.356 | | | 11.200** |
| | (0.539) | | | (4.430) |
| 'Max H-index' | | 0.036 | | −4.500** |
| | | (0.293) | | (1.770) |
| 'Min H-index' | | | 0.420 | −5.440** |
| | | | (0.840) | (2.660) |
| 'Full or Partial' | −16.200 | −12.800 | −2.190 | −5.700 |
| | (12.900) | (11.300) | (12.500) | (13.600) |
| 'Highest Institution Publications' | 0.114 | 0.099 | 0.159 | 0.138 |
| | (0.166) | (0.168) | (0.172) | (0.166) |
| 'Highest Co-author Experience' | −0.130 | 0.052 | 0.180 | −0.254 |
| | (0.346) | (0.343) | (0.343) | (0.340) |
| 'Number of authors' | 10.000*** | 7.990** | 10.900*** | 15.000*** |
| | (3.770) | (3.890) | (3.950) | (4.130) |
| 'Author at US university' | 5.340 | 6.180 | 6.830 | 3.900 |
| | (7.850) | (7.950) | (8.130) | (7.730) |
| 'Solo-authored' | 13.600 | 11.100 | 8.380 | 24.700** |
| | (10.400) | (10.500) | (10.700) | (10.700) |
| 'Avg. H-index':'Full or Partial' | 1.030* | | | −5.880 |
| | (0.610) | | | (4.560) |
| 'Max H-index':'Full or Partial' | | 0.558* | | |
| | | (0.330) | | |
| 'Min H-index':'Full or Partial' | | | 0.438 | |
| | | | (0.946) | |
| 'Full or Partial':'Max H-index' | | | | 2.870 |
| | | | | (1.840) |
| 'Full or Partial':'Min H-index' | | | | 3.110 |
| | | | | (2.780) |
| Constant | 4.120 | 9.770 | −4.710 | −19.500 |
| | (16.300) | (15.900) | (16.500) | (18.000) |
| Observations | 180 | 180 | 180 | 180 |
| Adjusted $R^2$ | 0.122 | 0.094 | 0.068 | 0.165 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.

TABLE A15
OLS: Log Citations on Reproduction Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.018* | | | 0.235*** |
| | (0.010) | | | (0.085) |
| 'Max H-index' | | 0.006 | | −0.092*** |
| | | (0.006) | | (0.034) |
| 'Min H-index' | | | 0.026 | −0.102** |
| | | | (0.016) | (0.051) |
| 'Full or Partial' | −0.119 | −0.116 | 0.058 | 0.115 |
| | (0.245) | (0.216) | (0.235) | (0.261) |
| 'Highest Institution Publications' | 0.003 | 0.002 | 0.004 | 0.004 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| 'Highest Co-author Experience' | −0.008 | −0.004 | −0.003 | −0.010 |
| | (0.007) | (0.007) | (0.006) | (0.007) |
| 'Number of authors' | 0.192*** | 0.150** | 0.225*** | 0.283*** |
| | (0.072) | (0.075) | (0.074) | (0.080) |
| 'Author at US university' | 0.073 | 0.099 | 0.086 | 0.041 |
| | (0.150) | (0.152) | (0.153) | (0.149) |
| 'Solo-authored' | 0.078 | 0.025 | −0.027 | 0.231 |
| | (0.199) | (0.202) | (0.200) | (0.206) |
| 'Avg. H-index':'Full or Partial' | 0.008 | | | −0.176** |
| | (0.012) | | | (0.088) |
| 'Max H-index':'Full or Partial' | | 0.005 | | 0.077** |
| | | (0.006) | | (0.035) |
| 'Min H-index':'Full or Partial' | | | −0.001 | 0.091* |
| | | | (0.018) | (0.054) |
| Constant | 2.680*** | 2.850*** | 2.520*** | 2.200*** |
| | (0.311) | (0.305) | (0.309) | (0.348) |
| Observations | 180 | 180 | 180 | 180 |
| Adjusted $R^2$ | 0.148 | 0.112 | 0.118 | 0.172 |

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
Log(YTD citations) is computed as log(YTD Citations + 1).
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.

TABLE A16
Poisson: YTD Citations on Reproduction Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.006*** | | | 0.251*** |
| | (0.002) | | | (0.009) |
| 'Max H-index' | | 0.000 | | −0.102*** |
| | | (0.006) | | (0.004) |
| 'Min H-index' | | | 0.009*** | −0.120*** |
| | | | (0.000) | (0.004) |
| 'Full or Partial' | −0.372*** | −0.259 | −0.032 | −0.118*** |
| | (0.050) | (−0.372) | (−0.259) | (−0.032) |
| 'Highest Institution Publications' | 0.003*** | 0.002 | 0.004 | 0.004 |
| | (0.001) | (0.003) | (0.002) | (0.203) |
| 'Highest Co-author Experience' | −0.003* | 0.002 | 0.004** | −0.005 |
| | (0.001) | (−0.003) | (0.002) | (0.186) |
| 'Number of authors' | 0.206*** | 0.161 | 0.203 | 0.303*** |
| | (0.013) | (0.206) | (0.161) | (0.071) |
| 'Author at US university' | 0.157*** | 0.182 | 0.186 | 0.120*** |
| | (0.033) | (0.157) | (0.182) | (0.009) |
| 'Solo-authored' | 0.219*** | 0.144 | 0.071 | 0.449 |
| | (0.042) | (0.219) | (0.144) | (0.351) |
| 'Avg. H-index':'Full or Partial' | 0.021*** | | | −0.164*** |
| | (0.002) | | | (0.014) |
| 'Max H-index':'Full or Partial' | | 0.010 | | 0.077 |
| | | (0.021) | | (0.236) |
| 'Min H-index':'Full or Partial' | | | 0.009 | 0.088*** |
| | | | (0.010) | (0.003) |
| Constant | 2.900*** | 3.010 | 2.730 | 2.350 |
| | (0.063) | (2.900) | (3.010) | (2.730) |
| Observations | 180 | 180 | 180 | 180 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.
Heteroskedasticity-robust standard errors.

TABLE A17
OLS: Arcsin Citations on Reproduction Outcomes, post-2012

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 'Avg. H-index' | 0.029 | | | 0.236 |
| | (0.018) | | | (0.156) |
| 'Max H-index' | | 0.012 | | −0.084 |
| | | (0.010) | | (0.063) |
| 'Min H-index' | | | 0.036 | −0.121 |
| | | | (0.028) | (0.102) |
| 'Full or Partial' | 0.217 | 0.240 | 0.226 | 0.271 |
| | (0.311) | (0.292) | (0.288) | (0.319) |
| 'Highest Institution Publications' | 0.007* | 0.007* | 0.008** | 0.008* |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| 'Highest Co-author Experience' | −0.004 | −0.000 | −0.002 | −0.006 |
| | (0.008) | (0.008) | (0.008) | (0.008) |
| 'Number of authors' | 0.231*** | 0.192** | 0.277*** | 0.292*** |
| | (0.082) | (0.085) | (0.085) | (0.093) |
| 'Author at US university' | −0.035 | 0.003 | −0.020 | −0.023 |
| | (0.176) | (0.178) | (0.179) | (0.181) |
| 'Solo-authored' | −0.004 | −0.073 | −0.081 | 0.059 |
| | (0.241) | (0.243) | (0.241) | (0.250) |
| 'Avg. H-index':'Full or Partial' | −0.001 | | | −0.194 |
| | (0.019) | | | (0.160) |
| 'Max H-index':'Full or Partial' | | −0.002 | | |
| | | (0.011) | | |
| 'Min H-index':'Full or Partial' | | | 0.000 | |
| | | | (0.031) | |
| 'Full or Partial':'Max H-index' | | | | 0.074 |
| | | | | (0.065) |
| 'Full or Partial':'Min H-index' | | | | 0.130 |
| | | | | (0.105) |
| Constant | 2.820*** | 2.950*** | 2.750*** | 2.580*** |
| | (0.393) | (0.393) | (0.372) | (0.416) |
| Observations | 129 | 129 | 129 | 129 |
| Adjusted $R^2$ | 0.237 | 0.208 | 0.224 | 0.231 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Notes:
YTD citations are cumulative citations to the article in Year 4.
H-index measures are computed across all authors of an article, in the previous year.
An author without citations has an h-index of 0.
Full or partial reproduction are defined in the text.
Results for all articles with attempted reproduction.
Sample restricted to articles published in 2012 and later, with attempted reproduction.

**Appendix A6: Assessment Questionnaire**

# Entry_Questionnaire_Draft

Please fill out the form to the best of your abilities.

* Indicates required question

1. NetID *

   (only if Cornell student; enter email otherwise)

2. DOI *

   What is the DOI (not the URL!) of the article you are reviewing? (do not include "doi://", do include the full DOI, e.g. "10.1245/article2345")

3. TypeOfArticle *

   Does the article contain empirical work, simulations, or experimental work?

   *Mark only one oval.*

   - Yes
   - No

## Computer Programs

In this section, we will describe any computer programs that are provided with the article. Please do not describe any data sets or additional online materials, we'll gather this information in the next sections.

4. Programs *

Does the article provide computer programs/code?

*Mark only one oval.*

- ◯ Yes
- ◯ No *Skip to question 14*

5. OnlineProgramsDOI

Please enter the DOI of the downloadable programs (only the "10.1234/article245" part )

_____

6. OnlineProgramsURL

Please enter the URL of the downloadable programs (notation: https:// )

_____

7. OnlineProgramsHDL

Please enter the Handle of the downloadable programs (notation: hdl:// )

_____

8. ProgramFormat

What format are the programs in?

*Check all that apply.*

- ☐ Stata
- ☐ R
- ☐ Matlab
- ☐ SPSS
- ☐ SAS
- ☐ Other: _____

9. ProgramsDocumentation

Are the programs clearly documented? (There are comments throughout the program that briefly describe what is done at each step)

*Mark only one oval.*

- Yes
- No

10. ProgramsHeaderAuthor

Do the programs have a header that identifies the author? (Program metadata)

*Mark only one oval.*

- Yes
- No

11. ProgramsHeaderInfo

Do the programs have a header that identifies when they were created and/or modified (Program metadata)

*Mark only one oval.*

- Yes
- No

12. ProgramsStructureManual

Do the instructions require the user to make manual modifications to data or programs?

*Mark only one oval.*

- Yes
- No

13. ProgramSequence

Does there appear to be enough detail provided (either in the code itself or in accompanying materials) to run all the programs?

*Mark only one oval.*

- ◯ Yes
- ◯ No

## Data

In this section, we will describe any data that is provided with the article as well as any other data sources mentioned in the article.

It is useful to distinguish between two types of data:

1. INPUT DATA: The "underlying source data" as collected by the authors or other agency (eg. the CPS or "my survey" data). This data is used to construct the final analysis data set.

2. ANALYSIS DATA: The post-processed or "cleaned" data set(s) that are direct inputs into the final programs to produce the results reported in the article.

The basic research workflow is as follows: INPUT DATA -> [preparation programs] -> ANALYSIS DATA -> [regression programs] -> results.

[Comment: Authors will usually only supply the analysis data set. The analysis data set is often constructed from numerous sources, which can sometimes be described within the article itself or in an online appendix.]

14. OnlineDataProvided *

Are any datasets provided with the article?

*Mark only one oval.*

- ◯ Yes *Skip to question 18*
- ◯ No

15. DataAbsence *

Why is data not provided? (Eg. because the data are confidential or proprietary, or because there is a public-use download site for the data)

*Check all that apply.*

- [ ] Confidential Data
- [ ] Proprietary Data
- [ ] Missing data (no justification)
- [ ] Licensed data
- [ ] Redistribution not authorized
- [ ] Other data download site provided
- [ ] Other: _____

16. DataAvailabilityAccess

Do the data require users to apply for access, purchase, or otherwise sign or enter into agreements to access the data? This could be a license agreement, or even a click-through acknowledgement. (This should be mentioned in the Readme PDF or in the article) [Answer DK (Don't know) if it is not clear from the article how users can access non-downloadable data.]

*Mark only one oval.*

- ◯ Yes
- ◯ No
- ◯ DK

17. DataAvailabilityExclusive

Are the data accessible only to the authors? [Answer yes if the authors clearly state that the data are only available to them. Answer No if there is clear evidence that others can access to the data, albeit with restrictions. Answer DK if you can't figure it out from the article.]

*Mark only one oval.*

- ◯ Yes
- ◯ No
- ◯ DK

*Skip to question 31*

## Online Data

18. OnlineDataDOI1

Please enter the DOI of the downloadable dataset (only the "10.1234/article245" part )

_____

19. OnlineDataURL1

Please enter the URL of the downloadable dataset. (this may duplicate one or the other of DOI or HDL, but is a more general way to describe it. notation: https://)

_____

20. OnlineDataHandle1

Please enter the Handle of the downloadable dataset (notation: hdl://)

_____

21. OnlineDataFormat1

Online dataset format (Hint: Check the file extensions. .dta=Stata, .mat=Matlab etc.)

*Check all that apply.*

- [ ] Stata
- [ ] CSV
- [ ] R
- [ ] Matlab
- [ ] SPSS
- [ ] SAS
- [ ] Excel
- [ ] Other: _____

22. DataSetClassification1

Please classify the provided data

*Mark only one oval.*

- ◯ Analysis Data
- ◯ Input Data
- ◯ Don't Know

23. OtherNotes1

Any notes for this dataset that was not covered by the questions above.

_____

_____

_____

_____

_____

24. Do you want to describe another dataset provided with the article? *

*Mark only one oval.*

- ◯ Yes
- ◯ No *Skip to question 31*

25. OnlineDataDOI2

Please enter the DOI of the downloadable dataset (only the "10.1234/article245" part )

_____

26. OnlineDataHandle2

Please enter the Handle of the downloadable dataset (notation: hdl://)

_____

27. OnlineDataURL2

Please enter the URL of the downloadable dataset. (this may duplicate one or the other of DOI or HDL, but is a more general way to describe it. notation: https://)

_____

28. OnlineDataFormat2

Online dataset format

*Check all that apply.*

- ☐ Stata
- ☐ CSV
- ☐ R
- ☐ Matlab
- ☐ SPSS
- ☐ SAS
- ☐ Excel
- ☐ Other: _____

29. DataSetClassification2

Please classify the provided data

*Mark only one oval.*

- ◯ Analysis Data
- ◯ Input Data
- ◯ Don't Know

30. OtherNotes2

Any notes for this dataset that was not covered by the questions above.

_____

_____

_____

_____

_____

## Input Data Not Provided

More often than not, data sets that are provided with the article are constructed from other sources, which are sometimes described in the article itself or the accompanying materials.

31. InputData *

Does the article describe input data which is not provided with the article?

*Mark only one oval.*

- ◯ Yes
- ◯ No *Skip to question 47*

32. InputData1

Name the agency/individual from which the input data was sourced

_____

33. InputDataDOI1

Please enter the DOI of the input dataset (only the "10.1234/article245" part )

_____

34. InputDataHandle1

Please enter the Handle of the downloadable dataset (notation: hdl://)

_____

35. InputDataURL1

Please enter the URL of the source data set

_____

36. InputDataFormat1

Input dataset format

*Check all that apply.*

- [ ] Stata
- [ ] CSV
- [ ] R
- [ ] Matlab
- [ ] SPSS
- [ ] SAS
- [ ] Excel
- [ ] Other: _____

37. InputDataAvailability1

Is the input dataset publicly available?

*Mark only one oval.*

- ◯ Yes
- ◯ No
- ◯ Don't know

38. InputDataOtherNotes1

Any notes for this dataset that was not covered by the questions above.

_____

_____

_____

_____

_____

39. *

Do you want to describe another input dataset?

*Mark only one oval.*

- ◯ Yes
- ◯ No *Skip to question 47*

40. InputData2

Name the agency/individual from which the input data was sourced. (Eg. Census Bureau, Robert Shimer's website)

_____

41. InputDataDOI2

Please enter the DOI of the input dataset (only the "10.1234/article245" part )

42. InputDataHandle2

Please enter the Handle of the downloadable dataset (notation: hdl://)

43. InputDataURL2

Please enter the URL of the source data set

44. InputDataFormat2

Input dataset format

*Check all that apply.*

- [ ] Stata
- [ ] CSV
- [ ] R
- [ ] Matlab
- [ ] SPSS
- [ ] SAS
- [ ] Excel
- [ ] Other:

45. InputDataAvailability2

Is the input dataset publicly available?

*Mark only one oval.*

- ◯ Yes
- ◯ No
- ◯ Don't know

46. InputDataOtherNotes2

Any notes for this dataset that was not covered by the questions above.

## Online Material

In this section, we will describe the additional materials provided with the article.

47. OnlineMaterials *

Does the article have additional online materials (eg. Online Appendix, Author Diclosure Statements etc.)?

*Mark only one oval.*

- ◯ Yes
- ◯ No *Skip to question 59*

48. OnlineMaterials1

Name and URL of online material/file. (Right-click on link, copy and paste here)

49. OnlineMaterials1_DOI

DOI of online material/file (only the "10.1234/article245" part )

_____

50. OnlineMaterials1_Description

Brief description (eg. Readme file describing the order in which the programs should be run)

_____

_____

_____

_____

_____

51. Is there some other online material that you wish to describe? *

*Mark only one oval.*

- ◯ Yes
- ◯ No *Skip to question 59*

52. OnlineMaterials2

Name and URL of online material/file. (Right-click on link, copy and paste here)

_____

53. OnlineMaterials2_DOI

DOI of online material/file (only the "10.1234/article245" part )

_____

54. OnlineMaterials2_Description

Brief description (eg. Readme file describing the order in which the programs should be run)

55. Is there some other online material that you wish to describe? *

*Mark only one oval.*

- Yes
- No *Skip to question 59*

56. OnlineMaterials3

Name and URL of online material/file. (Right-click on link, copy and paste here)

57. OnlineMaterials3_DOI

DOI of online material/file (only the "10.1234/article245" part )

58. OnlineMaterials3_Description

Brief description (eg. Readme file describing the order in which the programs should be run)

Final Thoughts

59. GeneralNotes

General notes on this article, that wasn't captured by the questions

60. How difficult do you think replicating the article will be? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| easi | ◯ | ◯ | ◯ | ◯ | ◯ | hardest |

61. Flag for follow-up

*Mark only one oval.*

- ◯ Yes
- ◯ No

**Appendix A7: Exit Questionnaire**

# Exit_Questionnaire_Draft

Please fill out the form to the best of your abilities.

* Indicates required question

1. NetID or email *

   Please enter your NetID (if Cornell) or email (if not Cornell)

2. DOI *

   What is the DOI (not the URL!) of the article you reviewed?

3. Replication_Success *

   Did you successfully replicate the article?

   *Mark only one oval.*

   - ◯ Yes *Skip to question 7*
   - ◯ No *Skip to question 4*
   - ◯ Partial *Skip to question 7*

Main Issue

4. Main_Issue *

What was the main issue preventing the replication of the article?

*Mark only one oval.*

- ◯ Confidential Data *Skip to question 5*
- ◯ Missing Code
- ◯ Code Error
- ◯ Corrupted Data
- ◯ Do not have access to software
- ◯ Other: _____

## Confidential Data Classification

These categories can be somewhat subjective. If you are not sure how to categorize, leave it blank and we will discuss at the next meeting.

5. Data_Type

Please select the type of confidential data used by the authors.

*Mark only one oval.*

- ◯ administrative, national
- ◯ administrative, regional
- ◯ administrative, local
- ◯ private, commercial entity
- ◯ private, other

6. Data_Access_Type

Please select the type of confidential data used by the authors.

*Mark only one oval.*

- ◯ formal access
- ◯ informal access, commitment
- ◯ informal access, no commitment
- ◯ no information

*Skip to question 13*

## Original Program

7. Code_Success *

Did you manage to eventually get all the programs to run successfully?

*Mark only one oval.*

- ◯ Yes
- ◯ No

8. Program_Run_Clean *

Did the programs run "as is" without needing to make ANY changes?

*Mark only one oval.*

- ◯ Yes, no changes were necessary. *Skip to question 11*
- ◯ No, I needed to make changes in the code. *Skip to question 9*

## Changes to Program Code

9. Directory_Change *

Were the changes restricted to simply redirecting file/folder paths?

*Mark only one oval.*

- ◯ Yes
- ◯ No, the changes to the code were more involved.

10. Code_Changes

If the changes were more involved, briefly describe what changes you had to make.

## Program vs Paper Discrepancies

11. Output_Accuracy

Do the numbers produced by your program exactly match their corresponding values in the paper?

*Mark only one oval.*

- ◯ Yes
- ◯ No, some of the numbers are different.

12. Discrepancy_Location

If there are values that do not match, please list their location (ie. table number, column, page).

## Software and Data Classification

13. Software_Type *

Please select the relevant software used by the authors.

*Check all that apply.*

- [ ] Stata
- [ ] Matlab
- [ ] Python
- [ ] R
- [ ] Other: _____

14. Data_Type *

Please select the format of the provided data

*Check all that apply.*

- [ ] csv
- [ ] Excel
- [ ] txt
- [ ] Stata
- [ ] Matlab
- [ ] R
- [ ] Other: _____

## Software Issues

15. Software_Extensions

Did you have to load any software extensions? (Eg. In matlab, the optimization toolkit is required to run the fmincon command. In Stata, outreg2 needs to be installed before running the command.)

*Mark only one oval.*

- Yes
- No
- DK

16. Software_Version

Did the authors use a different version of software (ie. Stata11 instead of Stata13)?

*Mark only one oval.*

- Yes
- No
- DK

17. What Software did you use for the replication (including the version)? *

18. First_Replicator

Are you the first replicator?

*Mark only one oval.*

- Yes *Skip to question 22*
- No

## Previous Replicator Questions

19. Common_Issues

Did you encounter the same issues as the previous replicator?

*Mark only one oval.*

- Yes
- No

20. Overcome_Issues

Were you able to overcome any problems faced by the previous replicator?

*Mark only one oval.*

- Yes
- No
- N/A. The previous replicator had no issues.

21. Replication_Helpfulness

Describe the usefulness of the previous replicator's notes. Did you add to them?

## Original Author

22. How complete was the original author's readme/generic instruction file?

*Mark only one oval.*

- ◯ Complete. Provided all information required to run the programs.
- ◯ Incomplete. Was ambiguous or left out crucial steps.
- ◯ No readme file was provided.

23. What actions could the authors have made to make the replication exercise easier? (Eg. correctly point to folder names)

## Costs (optional)

If costs were indicated or found for (a) access to data (b) purchase of data (c) software you would have needed to buy, or any other monetary costs, please fill in the following two questions:

24. Cost of data access (in approximate $)

25. Cost of software purchase (in approximate $)

## Overall Rating

26. How difficult do you think the replication exercise was? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| easi | ◯ | ◯ | ◯ | ◯ | ◯ | hardest |

27. Comments on difficulty

If this differs from the initial assessment, why?

28. GeneralNotes

General notes on this article/replication, that wasn't captured by the questions

## Appendix A8: Replication Team

The following members of the Replication Lab provided valuable assistance:

Alice Elaine Chou Haeyong Shin, Yaxian Xie, Nathan Allan Bach, Cindy Vincens, Yuxin Chen, Sarah Jane Harrison, Yiwen Jiang, Jack Wendler, Jose Fernandez, Joran Isenberg, Sarah Harrison, Koonj Vekaria, Charley Chen, Yang Guo, Yiwen (Evelyn) Jiang, Noah Kwicklis, Madeline Kwicklis, Koonj Vekaria, Robin Wang, Jack Q. Wendler, Qianyan Yao, Joao Vitor Costa, Evan Shapiro, Yudi (Grace) Wang, Christopher Chang, Chuhan Liu, Daniel Kim, Kassandra Madulka, Robert Goldberg, Xinyi Wan, Siming Zou, Yu Gao, Andrew Wink, Matthew Salazar, Naomi Li, Anderson Park, Carina Chien, Nick Swan, Vendela Norman, Hayley A. Timmons, Jack VanSlyke, Gabriel Bond, Wenxin (Andee) Cao, Mcrid Wang, John Park, Xueshi Su, Sam Mbugua, Jiazhen Tan.

**Appendix A9: Acronyms Used**

**API** application programming interface

## References


Herbert, S., H. Kingi, F. Stanchi, and L. Vilhuber (2021) "The reproducibility of economics research: A case study," Working paper, Banque de France

Hirsch, J. E. (2005) "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences of the United States of America* 102(46), 16569–16572

Kingi, H., L. Vilhuber, S. Herbert, and F. Stanchi (2018) "The reproducibility of economics research: A case study," mimeo, BITSS

OurResearch (2023) "OpenAlex,"

Priem, J., H. Piwowar, and R. Orr (2022) "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts,"

Thomson-Reuters (2016) "Web of science,"