Reproducibility in Management Science

Fišar, Miloš; Greiner, Ben; Huber, Christoph; Katok, Elena; Ozkes, Ali I.; The Management Science Repoducibility Collaboration; Grad, Tom; Hünermund, Paul; Marchesini, Giacomo; Van der Borgh, Michel

Document Version
Accepted author manuscript

Published in:
Management Science

DOI:
10.1287/mnsc.2023.03556

Publication date:
2024

License
Unspecified

Citation for published version (APA):
Fišar, M., Greiner, B., Huber, C., Katok, E., Ozkes, A. I., The Management Science Repoducibility Collaboration, Grad, T., Hünermund, P., Marchesini, G., & Van der Borgh, M. (2024). Reproducibility in *Management Science*. *Management Science*, *70*(3), 1343-1356. https://doi.org/10.1287/mnsc.2023.03556

Reproducibility in Management Science*

Miloš Fišar, Ben Greiner, Christoph Huber, Elena Katok, Ali I. Ozkes, and the Management Science Reproducibility Collaboration†

December 21, 2023

Abstract

With the help of more than 700 reviewers we assess the reproducibility of nearly 500 articles published in the journal *Management Science* before and after the introduction of a new Data and Code Disclosure policy in 2019. When considering only articles for which data accessibility and hard- and software requirements were not an obstacle for reviewers, the results of more than 95% of articles under the new disclosure policy could be fully or largely computationally reproduced. However, for 29% of articles at least part of the dataset was not accessible to the reviewer. Considering all articles in our sample reduces the share of reproduced articles to 68%. These figures represent a significant increase compared to the period before the introduction of the disclosure policy, where only 12% of articles voluntarily provided replication materials, out of which 55% could be (largely) reproduced. Substantial heterogeneity in reproducibility rates across different fields is mainly driven by differences in dataset accessibility. Other reasons for unsuccessful reproduction attempts include missing code, unresolvable code errors, weak or missing documentation, but also soft- and hardware requirements and code complexity. Our findings highlight the importance of journal code and data disclosure policies, and suggest potential avenues for enhancing their effectiveness.

Keywords: reproducibility, replication, crowd science

---

*We thank the members of the Management Science Reproducibility Collaboration for their contributions, Matthew D. Houston and Lucas Unterweger for research support, and Anna Dreber, Susann Fiedler, Lars Vilhuber, two anonymous reviewers, and the Associate and Department editor for helpful comments.

†Fišar: Masaryk University, e-mail: milos.fisar AT econ.muni.cz.
Greiner: Wirtschaftsuniversität Wien, e-mail: bgreiner AT wu.ac.at, and University of New South Wales.
Huber: Wirtschaftsuniversität Wien, e-mail: christoph.huber AT wu.ac.at.
Katok: University of Texas at Dallas, e-mail: ekatok AT utdallas.edu.
Ozkes: SKEMA Business School, Université Côte d'Azur (GREDEG), e-mail: ali.ozkes AT skema.edu, and Université Paris-Dauphine - PSL (LAMSADE).
A complete list of the members of the Management Science Reproducibility Collaboration is included in Appendix A.

I Introduction

To be relevant and credible, scientific results have to be verifiable. The integrity of academic endeavors rests upon reproducibility, wherein independent researchers obtain consistent results using the same methodology and data, and replicability, which involves the application of similar procedures to new data.

The significance of these twin principles for scientific research is commonly agreed upon. Yet, recent assessments of empirical studies in the social sciences suggest a concerning rate of non-reproducibility or non-replicability (e.g., Ioannidis, 2005; Ioannidis and Doucouliagos, 2013; Open Science Collaboration, 2015). A replicability crisis does not only erode the confidence in individual studies, but casts a shadow over entire fields and literatures, and may potentially compromise business and policy decisions based on these findings. Assessing and addressing these issues is imperative to maintain the credibility of social science research, including management, psychology, economics, sociology, and political science, and its subsequent applications in economic policies and management strategies, guiding societal progress.

Several reasons are cited in the literature as contributing to reduced replicability, such as publication bias (De Long and Lang, 1992), undisclosed analysis flexibility (Simmons et al., 2011), *p*-hacking (Brodeur et al., 2016), and plain fraud (John et al., 2012; List et al., 2001). Ensuring that published results can be reliably reproduced is a necessary foundation for addressing these issues. While tackling the underlying reasons for limited replicability may be difficult, the ability to reproduce results based on the original data and analyses can be seen as a minimum criterion for scientific credibility to be expected from all published research (Christensen and Miguel, 2018; Nagel, 2018; Welch, 2019). Indeed, if published results cannot be reproduced because data are unavailable, or code used for data or numerical analysis is missing, poorly documented, or error-ridden, then the replicability crisis is partly also a reproducibility crisis.

In this study, we directly assess the reproducibility of results reported in nearly 500 research articles published in *Management Science*, a premier general interest academic journal that comprises of 14 departments covering a broad variety of areas in business and management. In 2019, the journal introduced a new Policy for Data and Code Disclosure,[1] which stipulates that "Authors of accepted papers ... must provide ... the data, programs, and other details of the experiment and computations sufficient to permit replication." While our focus is primarily on assessing the reproducibility of work published since the disclosure policy went into effect, we also analyze articles accepted before May 2019, for comparison.

In order to reproduce results in articles from a variety of sub-fields of the journal such as finance, accounting, marketing, operations management, organizations, strategy, and behavioral economics, we use a crowd-science approach (Nosek et al., 2012; Uhlmann et al., 2019) to leverage the expertise of many researchers in these different sub-fields. Overall, 733 volunteers joined the *Management*

[1] Retrieved on August 22, 2023, from `https://pubsonline.informs.org/page/mnsc/datapolicy`.

Science Reproducibility Collaboration as reproducibility reviewers (see Appendix A for all names and affiliations), who together reportedly spent more than 6,500 hours on attempting to reproduce the results reported in the articles, using the replication materials and information provided by the article authors.

For articles subject to the 2019 disclosure policy, we find that when the reviewers obtained all necessary data (because they were included, could be accessed elsewhere, or no data were needed) and managed to meet the soft- and hardware requirements of the analysis, then results in the vast majority of articles (95%) were fully or largely reproduced.[2] However, in approximately 29% of the articles, datasets were unavailable either because they were proprietary or under a non-disclosure agreement (NDA), or because they originated in subscription data services to which reviewers did not have access. If we consider all assessed articles under the disclosure policy, then about 68% could be at least largely reproduced. Since data availability was by far the largest obstacle to reproducing results, the methodology used in an article is strongly correlated to its reproducibility. Namely, computational and simulation studies as well as online and laboratory experiments are more likely to be reproducible than field experiments, surveys, and other empirical studies. These differences in methodology and data availability are also the main drivers for substantial heterogeneity in reproducibility across the 14 departments of the journal.

Comparing these results to the period before the introduction of the mandatory disclosure policy, we observe a substantial increase in reproducibility. When code and data disclosure was voluntary, only 12% of article authors provided replication materials. Out of these selected articles, 55% could be (largely) reproduced.

The share of fully and largely reproduced results in our study appears high, in particular considering that the Code and Data Editorial team at the journal primarily assesses the completeness of replication materials, but does not attempt reproduction of the results themselves. That said, in addition to limited data availability, some replication materials suffered from insufficient documentation, missing code, or errors in the code, making reproduction impossible. For some studies, reviewers obtained different results and were not able to make out the reasons for the discrepancies. This implies that there is still room for improvement. We discuss implications for disclosure policies and procedures at *Management Science* and other journals in Section IV of this paper.

Our results complement findings in a recent literature on reproducibility and replicability in the social sciences. The definitions of these terms vary somewhat across studies, with some overlaps in their meaning (e.g., Christensen and Miguel, 2018; Dreber and Johannesson, 2023; Pérignon et al., 2023; Welch, 2019). "Replication" typically refers to verifying the results of a study using different datasets and different methods, thus exploring the robustness of results. The term "computational reproducibility" comes closest to the scope of our study, and is defined as the extent to which results

[2]We use the term "largely reproduced" when only minor issues were found and the conclusions from the analysis were not affected.

in studies can be reproduced based on the same data and analysis as the original study.[3] Other types of reproducibility may consider recreation of analysis and data, or explore robustness to alternative analytical decisions (see also Dreber and Johannesson, 2023, for an in-depth discussion).

Recent systematic replication attempts of published results in the social sciences yielded replication rates of 36% in psychology (Open Science Collaboration, 2015, $N = 100$), 61% in laboratory experiments in economics (Camerer et al., 2016, $N = 18$), 62% in social science experiments published in *Nature* and *Science* (Camerer et al., 2018, $N = 21$), and 80% in behavioral operations management studies published in *Management Science* (Davis et al., 2023, $N = 10$).

In the field of economics, a number of studies targeting different sub-fields have set out to evaluate the computational reproducibility of results. The *Journal of Money, Credit and Banking* (JMCB) was one of the first journals to introduce a "data availability policy", and one of the first ones to be evaluated. Dewald et al. (1986) assess the first 54 studies subject to the policy. Only eight studies (14.8%) submitted materials that were deemed sufficient to attempt a reproduction, and only four of these studies could be reproduced without major issues. As the authors put it, "inadvertent errors ... are a commonplace rather than a rare occurrence" (Dewald et al., 1986, p. 587). McCullough et al. (2006) examine JMCB articles published between 1996 and 2002, and successfully reproduce 22.6% of 62 examined works with a code and data archive, and only 7.5% considering all 186 relevant empirical articles in the journal. McCullough et al. (2008) report that for articles published between 1993 and 2003 in the Federal Reserve Bank of St. Louis Review, only 9 out of 125 studies (7.2%) with an archive could be successfully reproduced.

One of the top journals in economics, the *American Economic Review*, introduced a data and code availability policy in 2004, and other top journals followed. In examining this policy for studies published between 2006 and 2008, Glandon (2011) reports that among the studies with sufficient data archives, five out of nine studies (55.6%) could be reproduced without major issues. Overall, however, only 20 out of 39 sampled studies (51.3%) contained a complete archive, and for eight studies (20.5%) a reproduction was not feasible without contacting the authors.

More recently, Chang and Li (2017) attempt to reproduce articles in macroeconomics published between 2008 and 2013 across several leading journals, and successfully reproduce 22 out of 67 studies (32.8%). Gertler et al. (2018) examine the reproducibility of 203 empirical studies published in 2016 that did not contain proprietary or otherwise restricted data, and are able to reproduce 37% of them (but only 14% from the raw data). For 72% of the studies in the sample, code was provided, but executed without errors in only 40% of the attempts. Herbert et al. (2023) ask undergraduate economics students to attempt to reproduce 303 studies published in the *American Economic Journal: Applied Economics* between 2009 and 2018. Only 162 studies contained non-confidential and non-proprietary data. For these, 68 reproduction attempts (42.0%) were successful and another 69 (42.6%) were deemed partially successful. Pérignon et al. (2023) leverage a set of 168 replication packages produced in the

[3] Other scholars refer to computational reproduction also as verification (Clemens, 2017), verifiability (Freese and Peterson, 2017), or pure replication (Hamermesh, 2007; for an overview see also Ankel-Peters et al., 2023).

context of an open science multi-analyst study in empirical finance (see Menkveld et al., 2023). Out of 1,008 hypothesis tests across all materials, 524 (52.0%) were fully reproducible, with another 114 (11.3%) yielding only small differences to the original results.

Reproducibility studies in other related fields show similarly limited reproducibility. For a sample of 24 studies subject to the *Quarterly Journal of Political Science*'s data and code review, Eubank (2016) finds that only 4 (16.7%) did not require any modification in order to reproduce the results. In genetics, Ioannidis et al. (2009) report that only 8 out of 18 microarray gene expression analyses (44.4%) were reproducible. An analysis of biomedical randomized controlled trials yields 14 out of 37 (37.8%) successfully reproduced studies (Naudet et al., 2018). Artner et al. (2021) attempt to reproduce the main results from 46 published articles in psychology with the underlying data but no code, and were successful in 163 out of 232 statistical tests (70.3%). Xiong and Cribben (2023) examine the reproducibility of 93 articles using fMRI published in prominent statistics journals between 2010 and 2021, of which only 23 (24.7%) included the actual dataset, and 14 (15.1%) could be fully reproduced.

A comparison of reproducibility rates across different studies is difficult. Different studies often apply different definitions and standards of reproducibility, and reasons for non-reproducibility may differ between different journals due to different policies and enforcement procedures, and different methods and data availability conditions in their fields. For example, our share of 95% of (largely) reproduced articles (conditional on data being available to the reviewer and hard- and software requirements being met) appears to be in a similar ballpark as the 85% of at least partially successful reproductions at the *AEJ: Applied Economics*. However, while both journals have similar disclosure policies, in the respective time periods replication materials of articles at *AEJ:AE* only underwent a cursory review while the Code and Data Editorial Team at *Management Science* checked all replication packages for completeness.

In recent years, there have been significant developments in the institutional arrangements for reproducibility of journal articles. For economics, Vlaeminck (2021) reports that in a sample of 327 journals, 59% have data availability policies, a significant increase compared to 21% in the year 2014. Similar developments are present in the fields of business and management. For example, several other journals published by INFORMS have adopted similar code and data disclosure policies after *Management Science* took the lead in 2019. At the time of writing this paper, 20 out of the 24 journals used for the UT Dallas Business School rankings have a code/data disclosure policy, but only 10 made code/data sharing compulsory, and only two have a code and data editor enforcing the policy.[4] Colliard et al. (2023) discuss journals' incentives with respect to reproducibility, and Höffler (2017) provides evidence that in economics, journals with disclosure policies are more often cited than journals without such policies.

[4]For comparison, out of the top 25 journals in the 2022 Scimago ranking in Economics and Econometrics, 23 have code/data policies, 17 require that code/data are shared, and 6 have code/data editors. There is some overlap of this set of journals with the UT Dallas list.

The ability to reproduce results reported in published articles by executing the code on the data, both provided by the authors, does not, by itself, guarantee that results are replicable. But it does provide a useful baseline. It increases confidence that reported results could, in principle, be replicated. Allowing access to original code and data also makes it possible for independent research teams to scrutinize robustness, conduct their own analysis including meta-analytical work spanning multiple studies and datasets, reuse code in other research, and either build on the results or design studies to show the limitations of original results. The ability to do this promotes scientific discourse, and importantly, also decreases incentives for academic fraud and data falsification.

II Study design and procedures

II.A Procedures

Prior to 2019, *Management Science* encouraged but did not require the disclosure of data for submitted/accepted manuscripts. In June 2019, a new policy was established, which applied to all newly submitted manuscripts and is still in effect at the time of this writing. The policy requires that all code and data associated with accepted manuscripts at *Management Science* have to be provided before the manuscript goes into production, but it also allows some exceptions, in particular licensed data (Compustat, CRSP, Factset, WRDS, etc.), proprietary data, or confidential data under a NDA. In these cases, detailed descriptions of data provenance and dataset creation are expected. The journal established the position of a Code and Data Editor (CDE) and consequently positions of Code and Data Associate Editors (CDAEs), who review all replication packages for completeness before an article goes into production. However, the CDE and CDAEs are volunteer positions, so there are limits to a complete check of the packages of all accepted articles for reproduction.[5]

Our study, pre-registered at the Open Science Framework,[6] attempts to assess the reproducibility of articles published in *Management Science* before and after the introduction of the 2019 policy, based on the materials provided by the authors. For the period after the policy change, our initial sample consists of 447 articles[7] that fell under the disclosure policy introduced in June 2019, had been reviewed by the CDE team through January 2023, and were published (with their compulsory replication package) on the journal's website. As a comparison sample we chose all 334 articles that were accepted at the journal between January 2018 and April 2019, and would have fallen under the disclosure policy (i.e., include code or data) but were accepted before the announcement of the policy and were thus not

[5]If code and data are included, the CDE team also attempts to run the code, but without verifying outputs. As a contrasting example, the American Economic Association employs a different model with a paid Data Editor position including a budget for administrative and research assistants, where all replication packages for all AEA journals are fully reproduced before a final acceptance decision is made.

[6]The pre-registration can be found at URL `https://osf.io/mjqg5`. Unless otherwise noted, we followed our pre-registered procedures.

[7]In our pre-registration we mention 450 articles, but during the review phase we noted that 3 of these articles did not fall under the disclosure policy, reducing the initial sample to 447.

subject to the policy (which only applied to articles initially submitted after June 1, 2019).[8] Out of those 334 articles, for 42 the authors had voluntarily provided a replication package, which entered our project reviews. Thus, the size of our initial sample of replication packages to be reproduced is 489.

On January 12, 2023, the Editor-in-Chief of *Management Science* wrote an email to all 9,762 reviewers who provided a review to the journal in the past 5 years, introducing the project and inviting them to serve as reproducibility reviewers (see Appendix E.1). In addition, the invitation to participate in the project was sent via professional mailing lists (e.g., Behavioral Economics, Finance, Marketing). In total, 927 researchers completed an initial reviewer survey asking for their research fields (namely, to which *Management Science* departments they would typically submit their manuscripts) and their familiarity with different analysis software/frameworks and databases (see Appendix E.2).

The assignment of articles to reviewers proceeded over two main assignment rounds and a consecutive third round. In the first assignment round at the beginning of February 2023, we attempted to find a reviewer for each of the 489 packages out of the 927 reviewers. We applied the Hungarian method (Kuhn, 1955) that tries to maximize the match with penalties for mismatches in department, software skills, and database access, and random resolution of ties (see Hornik, 2005, for the R implementation). These matches were then manually assessed for potential conflicts of interest (e.g., reviewer and author in the same department), in which case article and reviewer were removed from the match and re-entered the "pools" of articles and reviewers. Once the match was completed, all reviewers received an email informing them of their assignment, with links to the article, the supplementary materials page, and to guidelines for reviewers. Reviewers were also asked to either confirm their assignment, or to contact us to indicate any conflicts of interests or other reasons that they could not provide a report for the assigned article. These cases were also added back to the pool.

After two weeks, we ran a second assignment round. For articles, the sample consisted of previously unmatched articles (which received priority) and a second set of all articles (to find a second reviewer for many of them). All reviewers with no assignment yet entered the match. We once again used the Hungarian method with moderate penalties for department and software mismatches and prohibitive penalties for assignments of the same article or previous assignments, and random resolution of ties. The resulting match was screened for conflicts of interests. As before, reviewers received their assignment by email, and any reported mismatches or conflicts were tracked. A few dropouts of reviewers were recorded, otherwise articles and reviewers re-entered the "pool". Reviewers who did not confirm their assignment in the first or second round received a reminder email at the end of February.

The third round of assignments, from the beginning of March 2023, was run continuously in several waves and mostly manually. Once a sufficient mass of articles (rejections of assignments, leftover articles who have not received their second assignment yet) and reviewers (unmatched reviewers, or reviewers available for another report) was reached, for each article a list of all possible compatible

[8] Note that we thus deliberately did not include articles in our study that were accepted after the introduction of the 2019 policy but were not subject to it because they were originally submitted before the introduction. For these articles, their authors could have falsely assumed that the new disclosure policy applies while it did not, thus biasing our assessment of the effect of the policy.

reviewer matches was compiled, and out of this one reviewer was assigned. As before, reviewers were informed about their match and asked to confirm their assignment.

Reviewers were asked to make an honest attempt to a reproduction of the article's main results (figures, tables, and other results in the main manuscript) solely based on the provided replication materials (and not to contact the original authors of the articles, see also McCullough et al. 2006, for similar approaches) and to provide their report within about 5 weeks (though we also accepted late entries). Reviewers submitted their report through a structured survey implemented in Qualtrics (see Appendix E.3). They also received detailed guidelines (see Appendix E.4), providing definitions for different reproducibility assessment outcomes and explanations for all survey fields. The survey asked for an overall assessment, information about the content of the replication package (readme, data, code, etc.) and their quality, individual reproducibility assessment of all results tables and figures as well as other results reported in the manuscript, as well as assessments of time spent, of their own expertise in research field and analysis methods, and of their expectation of the replicability (as opposed to reproducibility) of the article. Reviewers were also asked to provide evidence of their reproduction attempts in the form of log files or screenshots.

During the whole review period, we answered any questions by reviewers by email. Once a significant number of reviews had been collected, we checked them for completeness and consistency. Where necessary, we followed up with reviewers to clarify questions and resolve inconsistencies.[9] All in all, we followed up on about 13% of all reports.

In late September 2023, we wrote emails to all corresponding authors of the articles for which we obtained reports, and provided them with the reports (redacted for anonymity). Authors could submit a short comment of up to 2,000 characters on each report, which was then included in our dataset.[10] 115 authors or author teams made use of this possibility and submitted comments.

II.B Final Sample

In total, we received 753 reports from 675 reviewers and reviewer teams, who spent in total more than 6,500 hours on this project.[11] We allowed reviewers to enlist the help of a colleague as a secondary reviewer, so for 61 reports reviewers are actually teams of two persons. While 599 reviewers provided one report each, 74 reviewers provided reports for two different articles, and two reviewers for three articles.

[9] E.g., a reviewer may indicate that log files are provided, but did not verify whether they are consistent with the results. In other cases, the overall assessment of a replication package may not have been consistent with the individual assessments of tables and figures. Some reviewers could initially not find the replication package because the respective link was missing on the journal's webpage, and we provided them with the correct links.

[10] In addition, the journal allows authors to submit an improved replication package, which will replace the previous (reviewed) replication package on the journal's replication server. We note, however, that our analysis is only based on the original replication materials.

[11] Two reviewers entered unrealistically high numbers of more than 160 hours (4 working weeks); we set these observations to "missing" in our dataset. The median reviewer spent 4 hours.

Table 1 shows that a majority of reviewers are at an intermediate stage in their academic career, at the Associate Professor, Assistant Professor, or Postdoc level. About one in seven reviewers was a full professor, and about the same number are PhD students. In addition, there are reviewers working in other roles at research and professional institutions. Across these career levels, reviewers differ in their frequency of enlisting a secondary reviewer (with Full or Associate Professors being more likely to do so, while almost all PhD students worked alone) and the time spent (differences there are mainly driven by whether it was a team or not). However, they do not differ much in their self-assessed expertise in the method or topic of the article. In our analysis below, we also did not find any systematic differences across reviewer characteristics in terms of assessment outcomes or other report characteristics.

TABLE 1: Reviewer characteristics

$N = 675$	Share	Enlisted 2nd reviewer	Avg. Hours Spent	Avg. Expertise Method (0-100)	Avg. Expertise Topic (0-100)
Professor	14%	21%	13.1	84.3	60.8
Associate Professor	26%	11%	8.3	83.2	61.5
Assistant Professor/Postdoc	40%	6%	8.4	84.1	58.7
PhD student	16%	1%	9.0	83.8	59.2
Other	4%	3%	6.1	82.8	52.7

Table 2 gives an overview of our final sample of assessed articles. Out of the 781 articles, 292 from before the introduction of the 2019 policy had no replication package, so are not assessed. For 30 articles with replication packages, we could not find a suitable reviewer, and thus cannot report any reproducibility results.[12]

TABLE 2: Initial and final sample of articles and reports

	Before 2019 policy	After 2019 policy	Total
Initial sample of articles	334	447	781
Articles with replication package available	42	447	489
Articles with package and report(s)	40	419	459
1 report	16	149	165
2 reports	24	270	294

[12]These 30 articles are not part of the analysis. We observe little evidence of selection issues. Table B.1 in Appendix B compares the software requirements of the 30 articles without a report and the 459 articles with at least one report. It seems that articles where we could not find a suitable reviewer were less likely to use the most common software Stata and more likely to use one of the less often used software. Still, these differences are statistically not significant at the 5%-level (Fisher Exact test, two-sided, on the frequency of Stata and frequency of "Other" software).

TABLE 3: Fields of assessed articles and reviewers

Management Science Department	Abbr.	Share of Articles ($N = 489$)	Share of Reviewers ($N = 675$)
Finance	FIN	27.4%	24.3%
Behavioral Economics and Decision Analysis	BDE	18.4%	30.1%
Accounting	ACC	12.5%	8.2%
Operations Management	OPM	9.2%	7.1%
Marketing	MKG	5.7%	6.5%
Revenue Management and Market Analytics	RMA	4.7%	0.7%
Information Systems	INS	4.3%	4.0%
Business Strategy	BST	3.3%	4.6%
Healthcare Management	HCM	3.3%	1.9%
Big Data Analytics/Data Science	BDA	3.1%	3.4%
Organizations	ORG	3.1%	3.6%
Entrepreneurship and Innovation	ENI	2.3%	4.0%
Optimization	OPT	1.4%	1.2%
Stochastic Models and Simulations	SMS	1.4%	0.4%

In Table 3 we list the *Management Science* departments where the articles in our final sample appeared.[13] This distribution is representative for articles in the journal, with Finance, Behavioral Economics and Decision Analysis, Accounting, and Operations Management being the largest fields. To facilitate the matching of reviewers and articles, upon registration we asked reviewers to which department(s) they would most likely send one of their articles. Table 3 shows the distribution of the first-named department. This distribution follows largely the distribution of articles, with the exception that researchers from Behavioral Economics and Decision Analysis contribute disproportionately.[14] During code and data review the CDE team usually classifies articles into one of five categories according to their main methods. While about one-fifth of the articles in the sample mainly use simulations or computations (and thus often do not rely on data), almost 60% of the articles in our sample are based on empirical data (primary or secondary datasets that do not originate from experiments or surveys), with the remaining articles discussing laboratory or online experiments (15%), field experimental data (4%), or data from surveys (3%).

II.C Reviewer consistency and aggregation

In order to obtain information on potential variability in reproducibility assessments, we aimed to get not just one but two reports for as many articles/replication packages as possible. We succeeded in obtaining two reproducibility reports for 294 articles. For 59% of these articles, both reviewers chose

[13] There have been some changes in the structure of departments at the journal over the past years. In case departments were changed or merged, we classified articles by the current (successor) department.

[14] One reason for this might be a higher awareness for the issues of reproducibility and replicability in this field. Another reason could be that most of the primary authors of this reproducibility study come from this research area.

the exact same overall assessment. When only considering whether a reviewer classified an article as at least largely reproducible, or not, then the agreement rate is 86%. For the overall assessment of reproducibility, reviewers seem to mostly differ on whether some minor issues are worth mentioning (in generally reproducible studies), and whether a few results that can be recovered are sufficient to deem a study "Largely reproduced" rather than "Not reproduced." Otherwise, differences may result from whether reviewers obtained access to datasets, managed to run the code in the appropriate software environment, or how much effort they put into the reproduction.[15]

In our analysis presented in the next section, we aggregated assessments at the article level. Specifically, if both reviewers chose the same overall assessment, we select one report randomly. If we have two reports for an article, we select the report with the higher reproducibility assessment. This is based on the expected error structure in assessments. When one reviewer could obtain the data or run the software but the other reviewer could not, then the former's more informed reproducibility judgement should be at least as positive as the latter's. Similarly, while random reviewer errors in assessing the results may lead to a lower reproducibility classification, it is unlikely that those errors yielded exactly the results also obtained by the original authors. And since reviewers had to document their reproducibility efforts and upload log files or screenshots, it seems unlikely that they would have incentives to overstate an assessment result.

We note that our approach in using the higher assessment of multiple reviews is in line with other reproducibility studies, e.g., Herbert et al. (2023). At the end of the next section we discuss the robustness of our results to using other aggregation rules or analyzing the data at the level of individual figures and tables, with detailed results included in Appendix C.

III Results

III.A Main results

In addition to individual reproducibility assessments of tables, figures, and other results, we asked reviewers for an overall assessment of their reproduction attempt. The guidelines given to reviewers stated the following assessment classifications:

- An assessment of "Fully reproduced" means that the output of the reproduction analysis shows the exact same results as reported in the article, for all results reported in the main manuscript.
- "Largely reproduced, with minor issues" means that there may be small differences in the reproduction output compared to the results in the original article, but the article's conclusions and learnings stay the same.

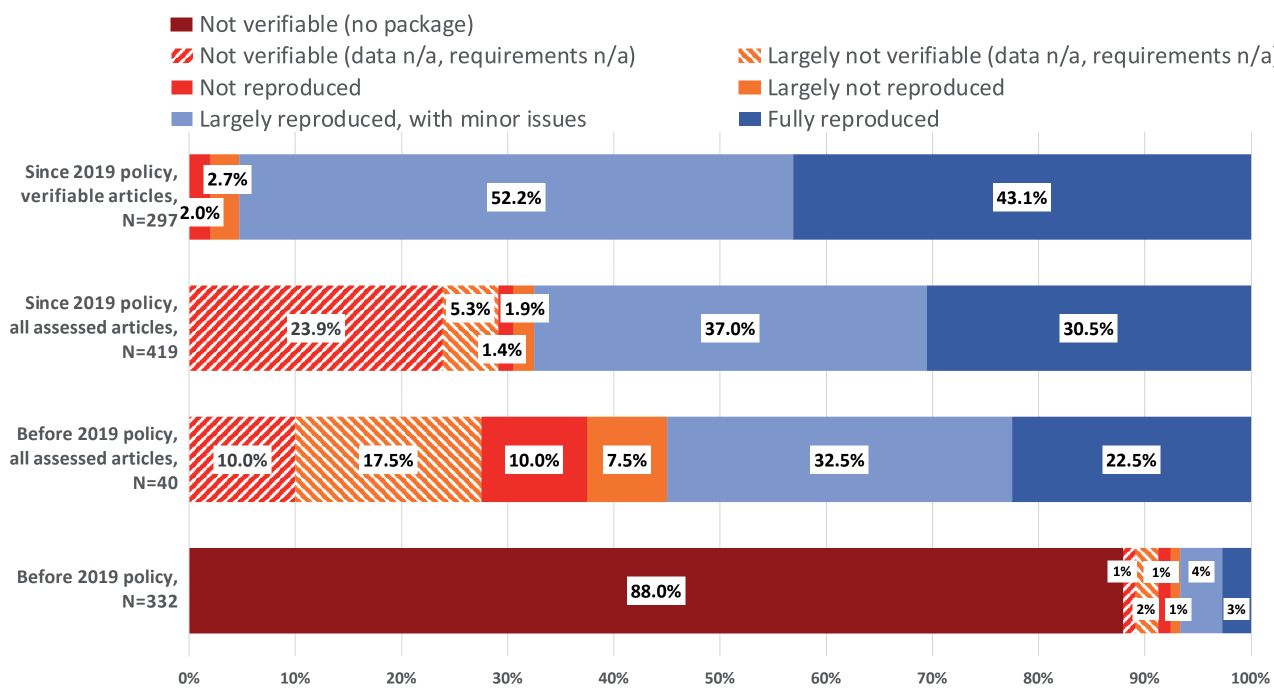
[15] In Appendix D we provide more details on variability in reviewer assessments.

- "Largely not reproduced, with major issues" means that there are major differences in the output compared to the results in the article, such that the reproduction results could not be used to support the conclusions of the original article.

- An assessment of "Not reproduced" means that the results from the reproduction cannot support the conclusions drawn in the paper, either because the output is different, or because the results cannot be produced at all because of missing data or non-recoverable code.

We note, however, that equipped with these guidelines, the eventual categorization of the article remains subjective to the reviewer. For all overall assessments of "Largely not reproduced" and "Not reproduced", we reviewed the individual reports to distill the main reasons for limited reproducibility. Consequently, cases where the reviewer was not able to get access to a required dataset or could not meet the software and hardware requirements of the analysis were labeled "Not verifiable" and "Largely not verifiable" rather than "Not reproduced" and "Largely not reproduced", respectively.[16]

FIGURE 1: OVERALL ARTICLE REPRODUCIBILITY ASSESSMENTS, BY POLICY



Based on these classifications, Figure 1 presents our main outcomes. The upper two panels show reproducibility assessments for articles that were subject to the disclosure policy introduced in 2019, while the lower two panels pertain to articles that were accepted before that policy. The first panel shows the distribution of assessments conditional on reproducibility being verifiable. Among these articles, 95.3% could be classified as fully reproduced or largely reproduced. However, for 29% of

[16]We note that this qualification of assessments was not yet anticipated in our pre-registration.

assessed articles, reviewers could not obtain the dataset, and in 1% the hard- and software requirements could not be met (e.g., software could not be installed, or the code would run for an untenable amount of time). Also in these cases, reviewers were not able to reproduce the results. The second panel in Figure 1 includes these cases, displaying results for all assessed articles. The share of articles that our reviewers were able to fully or largely reproduce is 67.5%.

The third panel of Figure 1 shows the overall assessments for the 40 articles from the time before the 2019 disclosure policy was introduced, for which replication materials were available. Our reviewers could reproduce or largely reproduce the results of 55% of these articles.[17] In the fourth panel of Figure 1, we include all 332 articles from our sample of articles accepted before the 2019 disclosure policy. Considering those articles that do not voluntarily provide replication materials as not reproducible reduces the share of at least largely reproduced articles to 6.6%.[18]

TABLE 4: Regressing reproducibility on disclosure policy existence

Model	(1)		(2)		(3)	
Sample of articles	All incl. no package		All with package		All verifiable	
	Coeff	StdErr	Coeff	StdErr	Coeff	StdErr
Constant	0.066***	(0.021)	0.550***	(0.075)	0.759***	(0.045)
Disclosure Policy	0.609***	(0.028)	0.125	(0.078)	0.194***	(0.047)
Observations	751		459		326	
R^2	0.379		0.006		0.051	

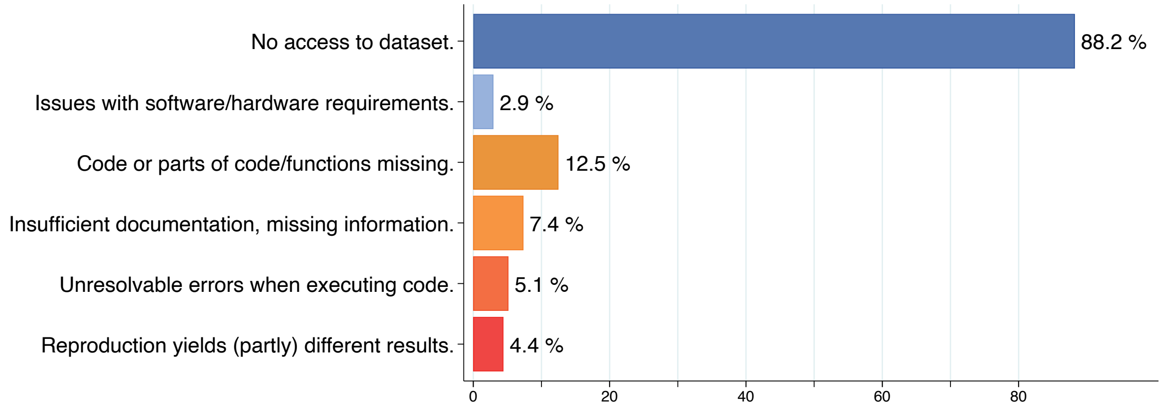
Notes: The dependent variable is a binary indicator whether the article was classified as "fully reproduced" or "largely reproduced", or not. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

Results from linear probability models, displayed in Table 4, lend statistical support to the positive change since the introduction of the data and code disclosure policy. In Model 1 we regress whether an article could be at least largely reproduced or not on the policy dummy for all articles in our sample (i.e., we are comparing the second and the fourth panels in Figure 1), indicating that after the introduction of the policy, a randomly chosen article is 61% more likely to be reproduced. In Model 2 we restrict our attention to the sample of articles for which a replication package was provided (i.e., comparing the second and the third panel in Figure 1). In this regression, the coefficient for the policy is positive but statistically not significant ($p = 0.109$). Finally, Model 3 focuses on all articles which are considered verifiable (i.e., comparing the second and the third panel in Figure 1 but without the

[17] We note, however, that these 40 out of 332 articles are heavily selected: authors voluntarily provided a replication package while being encouraged but not required by the journal. More than 50% of these articles were published in the BDE department, and none of them belonged to the Finance department, indicating selection also on availability of data.

[18] One may argue that when replication materials are not voluntarily provided to the journal, they may still be hosted on authors' personal websites or in other archives. For a random sample of 50 out of 292 articles without replication package, we searched all author websites as well as repositories for replication materials, and we found none.

FIGURE 2: Reasons for non-reproducibility for articles since 2019 policy



non-verifiable articles). The policy coefficient indicates that conditional on data being available and hard- and software requirements being met, articles are 19% more likely to be reproducible after the introduction of the disclosure policy.[19]

The unavailability of data is one of the major impediments for reviewers to reproduce an article. A dataset may be unavailable, for example, because the reviewer does not have a subscription to the commercial provider, because the dataset was collected under NDA with the involved company, or because the dataset contains sensitive information (e.g., on personal health or illegal activity). For the sample of 136 reviewed articles falling under the disclosure policy that were classified as either "Not reproduced" or "Largely not reproduced", Figure 2 displays the main reasons we identified for the reviewers' failure to reproduce.[20]

Limited access to the dataset was a reproducibility barrier for 88% of non-reproducible articles, and the time needed to run the code, complexity of the code, or issues with installing the software environment were the reason for non-reproducibility of another 3%. Other reasons included the non-availability of code or functions (13%), insufficient or missing documentation (7%), or unresolvable errors when executing the code (5%). For 4% of the non-reproducible or largely not reproducible articles, the main reason for this assessment was that the reproduction yielded partly different results than reported in the article.[21]

[19] We obtain the same conclusions employing corresponding Probit/Logit models or Fisher Exact tests. We note that strictly speaking, our data does not allow to imply a causal effect of the disclosure policy. Authors' attitudes towards making their research reproducible may have independently changed over time, just as the intensity of policy enforcement at the journal may have varied. Older replication packages may be less reproducible due to software changes. The introduction of the policy does not have features of a natural experiment, and our sample only spans a relatively short (and interrupted, see Footnote 8) time period.

[20] Note that multiple issues may apply to the same article.

[21] In Table B.2 in Appendix B we contrast these numbers with the reasons for non-reproducibility for articles which voluntarily provided replication packages before the 2019 disclosure policy took effect. Although the sample size for this period is low ($N = 18$), it appears that reasons for non-reproducibility of voluntarily provided packages are less likely to be missing data and more likely to be issues with missing or non-working code. Reproducibility for older materials may also be affected by limited backward compatibility of statistical software, sometimes producing different results. The reviewers in our study did not report such issues, but they may be more relevant when comparing more distant time frames.

Since many authors cannot include the original data in their replication packages for various reasons, in such cases the Code and Data Editor at the journal started to encourage the provision of log files that can show that the analysis code works and produces the desired results. Correspondingly, about 52% of the articles classified as "Not verifiable" or "Largely not verifiable" included log files for all results in the replication package, and further 24% included log files for at least some results. Consequently, 60% of (largely) not verifiable articles were assessed as "Not reproduced but consistent with log files" (84% of those that provided all log files, and 66% of those that provided at least some logs).

III.B Variation in reproducibility

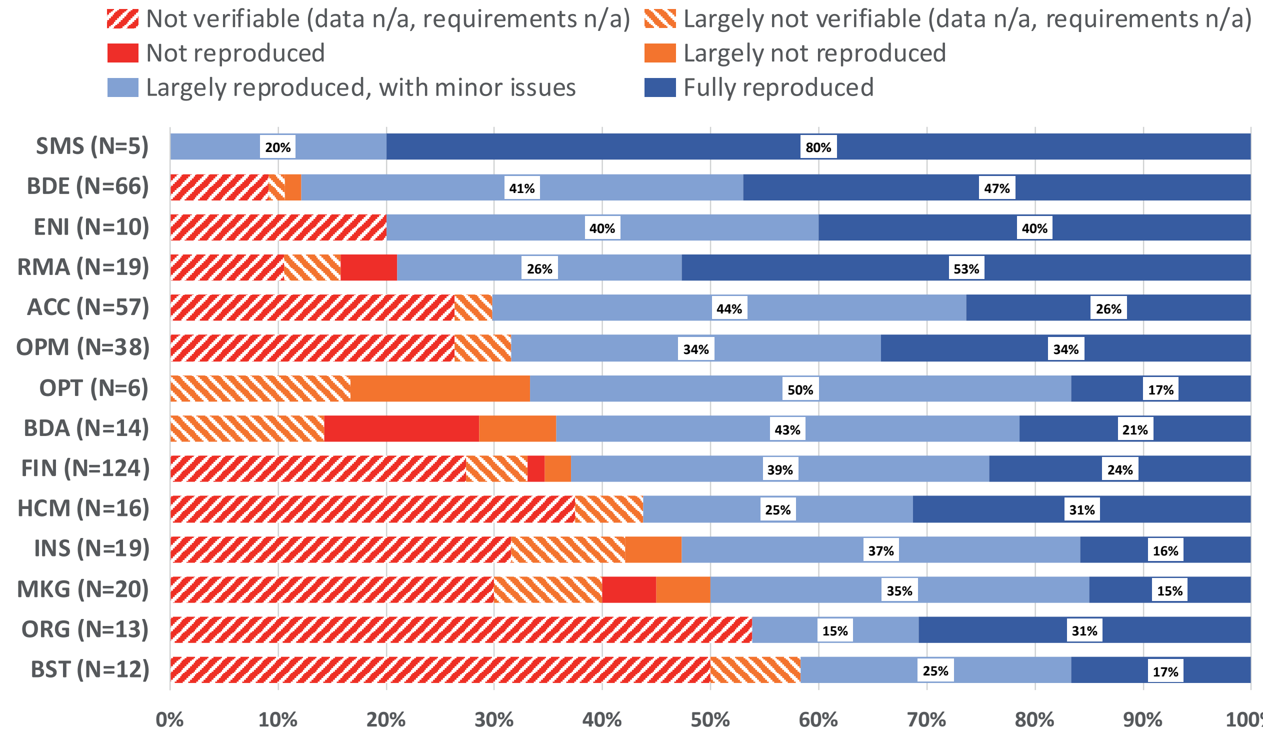
Our data allows us to break down the reproducibility of articles published under the disclosure policy to the level of research fields and types of research. Figure 3 shows the reproducibility assessments across the 14 *Management Science* departments. We observe considerable heterogeneity in the share of reproduced or largely reproduced articles across the different fields, ranging from 42% to 100%. Note, however, that there are substantial differences in the number of published articles across departments. Also, data availability may vary drastically between different fields.

While many studies in the department Behavioral Economics and Decision Analysis (BDE) rely on primary data from experiments, other fields often use proprietary data from subscription databases (e.g., Compustat, CRSP, WRDS), or confidential and sensitive data that cannot be shared with other researchers (e.g., field experiments with companies, health care data, or sensitive surveys). In Figure 4, we distinguish reproducibility outcomes by the primary type/method of the article, as classified during the journal's code and data review. We indeed observe significant differences in the reproducibility outcomes across articles employing different methods. All studies reporting on laboratory and online experiments include their dataset, making them highly reproducible. Most studies running simulations or other computations, mostly embedded in theoretical articles, do not rely on datasets, making them highly reproducible. On the other hand, many empirical studies with primary or secondary datasets rely on proprietary or subscription data, making them less reproducible if reviewers have no access to these datasets. Field experiments in business fields often run under NDAs, and survey studies may include sensitive data that cannot be shared (sometimes even ethics committees restrict the publication of datasets).[22]

In Table 5 we report three linear probability models in which we assess this heterogeneity statistically. The outcome variable in all three models is a dummy indicating whether an article is classified as fully or largely reproduced, or not. In Model (1), we regress reproducibility on department fixed effects, with the baseline being the Finance department (FIN), with a sizable sample size and close to the average reproducibility level. We observe that the SMS and BDE departments have significantly higher reproducibility rates than the Finance department, while the other departments do not differ significantly from Finance. In Model (2), we regress the same outcome on article type fixed effects,

[22]Table B.3 in Appendix B demonstrates the variation of paper types/methods across the different departments of the journal. In the table, we ordered departments and methods by their reproducibility to highlight the correlation.

FIGURE 3: Overall reproducibility assessments by journal department



Note: Department acronyms are SMS: Stochastic Models and Simulations, BDE: Behavioral Economics and Decision Analysis, ENI: Entrepreneurship and Innovation, RMA: Revenue Management and Market Analytics, ACC: Accounting, OPM: Operations Management, OPT: Optimization, BDA: Big Data Analytics/Data Science, FIN: Finance, HCM: Healthcare Management, INS: Information Systems, MKG: Marketing, ORG: Organizations, BST: Business Strategy.

FIGURE 4: Overall reproducibility assessments by article type/method

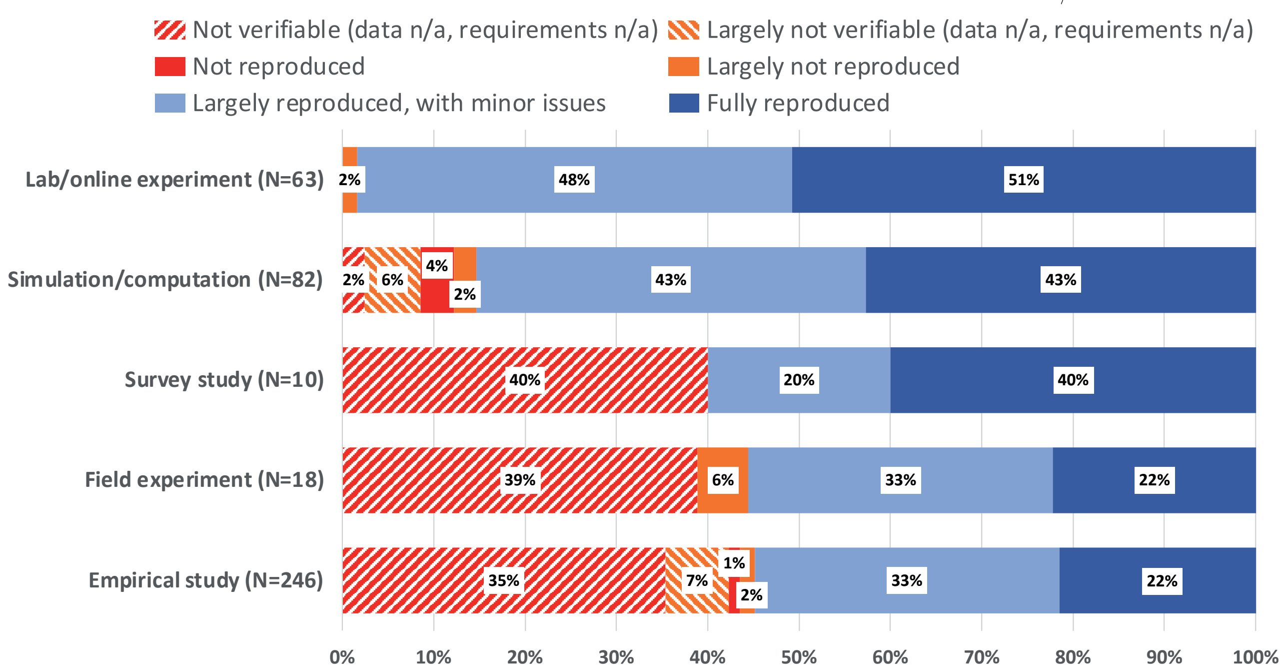


TABLE 5: Regressing reproducibility on journal department and article type

Model	(1)		(2)		(3)	
	Coeff	StdErr	Coeff	StdErr	Coeff	StdErr
Constant	0.629***	(0.041)	0.600***	(0.138)	0.630***	(0.146)
SMS	0.371*	(0.209)			0.034	(0.207)
BDE	0.250***	(0.070)			0.019	(0.087)
ENI	0.171	(0.151)			0.215	(0.143)
RMA	0.160	(0.113)			−0.110	(0.118)
ACC	0.073	(0.073)			0.128*	(0.070)
OPM	0.055	(0.085)			−0.049	(0.083)
OPT	0.038	(0.192)			−0.299	(0.191)
BDA	0.014	(0.129)			−0.323**	(0.137)
HCM	−0.067	(0.122)			−0.059	(0.115)
INS	−0.103	(0.113)			−0.073	(0.108)
MKG	−0.129	(0.111)			−0.118	(0.106)
ORG	−0.167	(0.134)			−0.120	(0.127)
BST	−0.212	(0.139)			−0.188	(0.134)
Lab/Online Experiments			0.384**	(0.149)	0.336**	(0.153)
Simulation/Computation			0.254*	(0.146)	0.336**	(0.155)
Field experiment			−0.044	(0.172)	−0.009	(0.173)
Empirical study			−0.051	(0.141)	−0.087	(0.143)
Observations	419		419		419	
R^2	0.072		0.140		0.180	

Notes: The dependent variable is a binary indicator whether the article was classified as "fully reproduced" or "largely reproduced", or not. Baseline is the Finance department, and survey studies. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively. Department acronyms are SMS: Stochastic Models and Simulations, BDE: Behavioral Economics and Decision Analysis, ENI: Entrepreneurship and Innovation, RMA: Revenue Management and Market Analytics, ACC: Accounting, OPM: Operations Management, OPT: Optimization, BDA: Big Data Analytics/Data Science, FIN: Finance, HCM: Healthcare Management, INS: Information Systems, MKG: Marketing, ORG: Organizations, BST: Business Strategy.

with articles based on surveys as the baseline. We find that while field experiments and empirical studies (other than experiments or surveys) do not differ from survey studies in their reproducibility, lab/online experiments and articles featuring simulation/computation are significantly more likely to be reproducible. Finally, in Model (3), we include both department and article type fixed effects. The coefficients for article type are not much affected by including department fixed effects, while vice versa there are some sizable changes. Once accounting for the article type/method used, articles in departments SMS and BDE are not significantly more reproducible anymore compared to other departments, namely Finance. On the other hand, controlling for methods, articles in the Accounting (ACC) department are significantly more reproducible than articles in Finance (more often including the data set), and articles in the field of Big Data Analytics (BDA) are less reproducible (as datasets are often not included or accessible).

III.C Robustness

In the analysis above we only considered reproducibility assessments at the article level, taking the higher assessment if two reports were available for an article. To examine the robustness of our results, we also examine the reproducibility for different aggregation rules, at the level of individual reports, and at the level of tables, figures, and other results.

In Appendix C, Table C.1 reports distributions of overall assessments when choosing the report with the lower assessment whenever there are multiple reports for an article, and when randomly selecting one of two reports (with 10000 repetitions). Since in our aggregation above we selected the report with the higher reproducibility assessment, these data show somewhat lower reproducibility levels. However, the differences are rather small. E.g., compared to the 95.3% (largely or fully) reproduced results for verifiable articles reported above, we observe 91.4% when taking the lower assessment of multiple reports, and 93.8% when randomizing which of two assessments is considered.

The regressions reported in Table C.2 are based on all reports rather than just one report per article, clustering standard errors at the article level. Their results mirror the results on policy effects reported in Table 4 above. Overall, the same reproducibility patterns emerge: the main reason for non-reproducibility is data access, departments differ widely in their reproduction rates, but that is to a large extent driven by different methods being used across departments.

Appendix C also reports and discusses the assessment results for individual tables, figures, and other results (e.g., statistical tests reported in the manuscript texts). As to be expected, these individual results are highly correlated with the overall assessments. For example, in reports that reached an overall assessment of "Fully reproduced", 99.1% of individual tables and 99.7% of individual figures were classified as largely or fully reproduced. When the overall assessment was "Not reproduced", only 2.7% of tables and 7.5% of figures could be reproduced, on average.

IV Discussion and Conclusion

In this study we undertake a comprehensive assessment of the reproducibility of results in *Management Science*. With the collaborative efforts of over 700 reviewers we examine nearly 500 articles to assess the computational reproducibility of their results. For articles published since the introduction of the 2019 disclosure policy, the good news is that more than 95% of articles could be fully or largely computationally reproduced, when data accessibility and hardware/software requirements were not obstacles for reviewers. This appears commendable. However, reviewers faced data accessibility challenges for approximately 29% of the articles in our sample, and the overall rate of successful reproduction is reduced to 68% when considering such articles as non-reproducible. Relatedly, differences in methods and dataset accessibility also drive heterogeneity in reproducibility rates across different fields.

This makes data availability a central issue in reproducibility. To improve the credibility of research within business and management, efforts should be directed toward facilitating data access and sharing. Strictly restricting a journal in the area of business, economics, and management to only articles that can freely share their data seems unrealistic and would exclude valuable research from being published. Instead, other arrangements may need to be found for such cases. Approaches could include, among others,

- the inclusion of de-identified data in the replication package, only useful for reproduction but not for new original research;
- agreements with subscription databases for access for reproduction purposes via the journal;
- providing access to datasets through special infrastructure that limits use to specific purposes (similar to platforms used by government agencies to provide micro data); or
- sharing data only with a journal's code and data editor or with a third-party agency which then certifies reproducibility.

In addition, human subjects ethics committees may need to be sensitized to also consider the ethics of research transparency in their deliberations, to find compromises that at the same time ensure human participant privacy and allow for the full reproduction of research results. Data access limitations also touch upon important questions of fairness and bias: with proprietary, non-open datasets, certain research results may only be obtained by privileged researchers, with the data provider serving as a gatekeeper with potential conflicts of interest.

Our study underscores the value of large-scale reproducibility assessment projects. We provide an assessment of the current state of affairs in the field of business and management, and thus contribute to drawing a realistic picture of the overall credibility of research in the field. Repeating such assessments will serve as a form of quality control for newly developed journal policies and procedures. The project showcases best practices and may help developing standards for replication materials, but also identifies major gaps and weaknesses in current policies that need to be addressed. Our results can influence journal and funding agency policy decisions. The active participation of more than 700 reviewers who invested significant time and effort in reproducing results highlights the commitment in the community to improving scientific rigor. In an ex-post survey, quite a few of our reviewers reported that their participation was a great learning experience, in particular with respect to preparing their own future replication packages. Informed about the assessments of their articles, most authors appreciated the reviewers' comments, and many voluntarily provided improved versions of their replication packages that address the reviewer comments. Thus, this project also raised awareness of reproducibility issues, furthering a culture of open science, and potentially also the quality of (existing and future) replication materials.

That said, our study also sheds light on the significance of journal code and data review procedures. We observe that the introduction of the 2019 disclosure policy is associated with a significant increase in the reproducibility of articles in *Management Science*. When code and data disclosure was voluntary, only 12% of authors submitted replication materials (out of which 55% could be at least largely reproduced). This suggests that the policy's effect is largely driven by increasing the mere *verifiability* of articles. However, there is still room for significant improvement. Smaller scale changes could be targeted towards improving the current process, such as increasing incentives for authors to provide proper replication packages right away by making the acceptance decision conditional on replication package approval; or integrating the code and data review process into the manuscript handling system to make it more efficient and transparent.

A more comprehensive reevaluation of code and data review procedures, however, may foster the pivotal role that code and data review plays in ensuring research reproducibility more effectively. In particular, large-scale reproducibility projects such as the present study may become obsolete if the journal puts resources and processes into verifying reproducibility already upon publication of an article. In the current institutional setup, the Code and Data Editor at *Management Science* and his team of Associate Editors are volunteers with naturally limited capacity to conduct comprehensive reproduction. To that end, different institutional arrangements may be advisable:

- Similar to the institutional setup at the American Economic Association (see Vilhuber, 2019), code and data review could be professionalized by introducing the position of a (half- or full-time) paid Code and Data Editor, with appropriate budget for assistance and software and data access.
- Code and data review, and reproducibility certification could be delegated to a third-party agency that conducts these activities for a fee (such as, for example, the Odum Institute used by the *American Journal of Political Science*, or CASCaD, see Pérignon et al., 2019).
- The fact that more than 700 reviewers participated in this project indicates that there is sufficient expertise in the community to integrate the code and data review into the peer review cycle of a manuscript, with low direct costs. E.g., in a last minor revision round, one reviewer could be assigned by the Department or Associate Editor to review the replication materials and certify reproducibility. However, while the willingness to participate in this project may have been driven by its novelty, one might have to consider other incentives for reviewers when establishing such reproducibility assessments as a regular procedure.

The scope of Code and Data policies extends beyond just enabling computational reproduction; their broader aim is to facilitate the replication of research results in order to assert their robustness and generalizability. Reproducibility does not imply replicability. There may be instances where a study is reproducible but not replicable (e.g., the results can be obtained with the same dataset but not with a new dataset generated in a different context). Conversely, a study might not be reproducible but replicable (e.g., the original dataset may be unavailable so the code cannot be applied, but results with data collected from a different source show the same effects).

We contend, however, that reproducibility serves as a vital foundation for evaluating replicability. A reproducible study boosts confidence in its results, making it meaningful to further examine its robustness and generalizability. The provision of datasets allows for the detection of anomalies and fraud. Materials provided for the reproduction of a study often facilitate its replication as well, by allowing researchers to better understand the structure of data and to apply the same analysis code to new datasets. In addition, in order to support replication studies, materials required to be provided under most code and data policies extend beyond those purely needed for reproduction. Even if datasets are not available and reproducibility thus not achievable, the packages nevertheless contain detailed descriptions of data provenance and variable dictionaries, aiding replication researchers in gathering new data. For surveys, materials include complete questionnaires or their software implementations, while for experimental studies, they encompass experiment instructions, software code, and other resources critical for running a replication study.

In conclusion, our study illuminates the critical importance of reproducibility in maintaining the integrity and credibility of scientific research in Management Science and related fields. By addressing data availability challenges and refining journal code and data review procedures, the academic community can work collaboratively to improve reproducibility. These efforts are essential to ensuring that robust research findings continue to guide decision-making and contribute to the advancement of knowledge.

References

Ankel-Peters, J., Fiala, N. and Neubauer, F. (2023), 'Do economists replicate?', *Journal of Economic Behavior & Organization* **212**, 219–232.

Artner, R., Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F. and Vanpaemel, W. (2021), 'The reproducibility of statistical results in psychological research: An investigation using unpublished raw data', *Psychological Methods* **26**(5), 527–546.

Brodeur, A., Lé, 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. et al. (2016), 'Evaluating replicability of laboratory experiments in economics', *Science* **351**(6280), 1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M. et al. (2018), 'Evaluating the replicability of social science experiments in nature and science between 2010 and 2015', *Nature Human Behaviour* **2**(9), 637–644.

Chang, A. C. and 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.

Christensen, G. and Miguel, E. (2018), 'Transparency, reproducibility, and the credibility of economics research', *Journal of Economic Literature* **56**(3), 920–980.

Clemens, M. A. (2017), 'The meaning of failed replications: A review and proposal', *Journal of Economic Surveys* **31**(1), 326–342.

Colliard, J.-E., Hurlin, C. and Pérignon, C. (2023), 'The economics of computational reproducibility', HEC Paris Research Paper No. FIN-2019-1345.

Davis, A. M., Flicker, B., Hyndman, K. B., Katok, E., Keppler, S., Leider, S. et al. (2023), 'A replication study of operations management experiments in management science', *Management Science* **69**(9), 4973–5693.

De Long, J. B. and Lang, K. (1992), 'Are all economic hypotheses false?', *Journal of Political Economy* **100**(6), 1257–1272.

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* pp. 587–603.

Dreber, A. and Johannesson, M. (2023), A framework for evaluating reproducibility and replicability in economics. Working Paper.

Eubank, N. (2016), 'Lessons from a decade of replications at the quarterly journal of political science', *PS: Political Science & Politics* **49**(2), 273–276.

Freese, J. and Peterson, D. (2017), 'Replication in social science', *Annual Review of Sociology* **43**, 147–165.

Gertler, P., Galiani, S. and Romero, M. (2018), 'How to make replication the norm', *Nature* **554**(7693), 417–419.

Glandon, P. J. (2011), 'Appendix to the report of the editor: Report on the american economic review data availability compliance project', *American Economic Review: Papers & Proceedings* **101**(3), 695–9.

Hamermesh, D. S. (2007), 'Replication in economics', *Canadian Journal of Economics* **40**(3), 715–733.

Herbert, S., Kingi, H., Stanchi, F. and Vilhuber, L. (2023), 'The reproducibility of economics research: A case study', Working Paper, Banque de France.

Hornik, K. (2005), 'A clue for cluster ensembles', *Journal of Statistical Software* **14**, 1–25.

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

Ioannidis, J. P. (2005), 'Why most published research findings are false', *PLoS Medicine* **2**(8), e124.

Ioannidis, J. P., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C. et al. (2009), 'Repeatability of published microarray gene expression analyses', *Nature genetics* **41**(2), 149–155.

Ioannidis, J. P. and Doucouliagos, C. (2013), 'What's to know about the credibility of empirical economics?', *Journal of Economic Surveys* **27**(5), 997–1004.

John, L. K., Loewenstein, G. and Prelec, D. (2012), 'Measuring the prevalence of questionable research practices with incentives for truth telling', *Psychological Science* **23**(5), 524–532.

Kuhn, H. W. (1955), 'The Hungarian method for the assignment problem', *Naval Research Logistics Quarterly* **2**, 83–97.

List, J. A., Bailey, C. D., Euzent, P. J. and Martin, T. L. (2001), 'Academic economists behaving badly? a survey on three areas of unethical behavior', *Economic Inquiry* **39**(1), 162–170.

McCullough, B. D., McGeary, K. A. and Harrison, T. D. (2006), 'Lessons from the JMCB archive', *Journal of Money, Credit and Banking* pp. 1093–1107.

McCullough, B. D., McGeary, K. A. and Harrison, T. D. (2008), 'Do economics journal archives promote replicable research?', *Canadian Journal of Economics/Revue canadienne d'économique* **41**(4), 1406–1420.

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M. et al. (2023), 'Non-standard errors', *Journal of Finance* . Forthcoming.

Nagel, S. (2018), 'Code-sharing policy: Update, march 6, 2018', Journal of Finance (Editor's Blog).

Naudet, F., Sakarovitch, C., Janiaud, P., Cristea, I., Fanelli, D., Moher, D. and Ioannidis, J. P. (2018), 'Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in the bmj and plos medicine', *BMJ* **360**.

Nosek, B. A., Spies, J. R. and Motyl, M. (2012), 'Scientific utopia: II. restructuring incentives and practices to promote truth over publishability', *Perspectives on Psychological Science* **7**(6), 615–631.

Open Science Collaboration (2015), 'Estimating the reproducibility of psychological science', *Science* **349**(6251), aac4716.

Pérignon, C., Akmansoy, O., Hurlin, C., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Menkveld, A. J., Razen, M. et al. (2023), Computational reproducibility in finance: Evidence from 1,000 tests. Working Paper.

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

Simmons, J. P., Nelson, L. D. and Simonsohn, U. (2011), 'False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant', *Psychological Science* **22**(11), 1359–1366.

Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., McCarthy, R. J., Riegelman, A., Silberzahn, R. and Nosek, B. A. (2019), 'Scientific utopia III: Crowdsourcing science', *Perspectives on Psychological Science* **14**(5), 711–733.

Vilhuber, L. (2019), 'Report by the aea data editor', *American Economic Review: Papers and Proceedings* **109**, 718–729.

Vlaeminck, S. (2021), 'Dawning of a new age? economics journals' data policies on the test bench', *LIBER Quarterly: The Journal of the Association of European Research Libraries* **31**(1), 1–29.

Welch, I. (2019), 'Reproducing, extending, updating, replicating, reexamining, and reconciling', *Critical Finance Review* **8**(1-2), 301–304.

Xiong, X. and Cribben, I. (2023), 'The state of play of reproducibility in statistics: an empirical analysis', *The American Statistician* **77**(2), 115–126.

Online Appendix

for Fišar, Greiner, Huber, Katok, Ozkes, and the Management Science Reproducibility Collaboration, "Reproducibility in Management Science", *Management Science* 2024.

A The Management Science Reproducibility Collaboration

The following co-authors lent their time and expertise as reproducibility reviewers to the Management Science Reproducibility project and are credited as "Management Science Reproducibility Collaboration" in the author string.

Diya Abraham, University of Reading
Gabrielle S. Adams, University of Virginia
Arzi Adbi, National University of Singapore, Business School
Jawad M. Addoum, Cornell University
Maja Adena, WZB Berlin
Laxminarayana Yashaswy Akella, Indian Institute of Management Ahmedabad
Pat Akey, University of Toronto
Olivier Akmansoy, HEC Paris; CNRS
Andres Alban, Harvard University, Harvard Medical School
Vitali Alexeev, University of Technology Sydney
Azizjon Alimov, IESEG School of Management
Argun Aman, University of Mannheim
Ali Aouad, London Business School
Gil Appel, George Washington University, School of Business
Nick Arnosti, University of Minnesota
Kashish Arora, Indian School of Business
Thibaut Arpinon, Georg-August Universität Göttingen
Florian M. Artinger, Max Planck Institute for Human Development; Simply Rational - The Decision Institute; Berlin International University of Applied Sciences
Joachim Arts, University of Luxembourg
Lennart Baardman, University of Michigan, Ross School of Business
Zakaria Babutsidze, SKEMA Business School
Golnaz Bahrami, Pennsylvania State University
Somnath Banerjee, North Dakota State University
Chenzhang Bao, Oklahoma State University
Te Bao, Nanyang Technological University, School of Social Science
Opher Baron, University of Toronto, Rotman School of Management
Xabier Barriola, INSEAD
Pedro Monteiro e Silva Barroso, Universidade Católica Portuguesa
Ernest Baskin, Saint Joseph's University
Robert J. Batt, University of Wisconsin-Madison, Wisconsin School of Business
George Batta, Claremont McKenna College
Anahid Bauer, Institut Mines-Télécom Business School, LITEM, Paris Saclay
Konstantin Bauman, Temple University, Fox School of Business
William Bazley, University of Kansas
Michael Becker-Peth, Erasmus University, Rotterdam School of Management
Mehmet Begen, Western University, Ivey Business School
Nazire Begen, Gebze Technical University
Sylvain Benoit, Université Paris Dauphine - PSL
Loic Berger, University of Lille, IESEG School of Management, LEM - Lille Economie Management; CNRS; iRisk Research Center on Risk and Uncertainty
Noémi Berlin, CNRS, EconomiX, Université Paris Nanterre
Lars Peter Berling, Norwegian University of Science and Technology
Anna Bernard, Catolica Lisbon School of Business and Economics
Jeremy Bertomeu, Washington University in St. Louis
Jędrzej Białkowski, University of Canterbury
Pawel Bilinski, City University of London, Bayes Business School
Jannis Bischof, University of Mannheim
Jeffrey R. Black, University of Memphis
Hayley Blunden, American University
Dion Bongaerts, Erasmus University, Rotterdam School of Management

Felix Bönisch, WZB Berlin
Marieke Bos, Swedish House of Finance
Ciril Bosch-Rosa, Technical University of Berlin
Sylvain Bourjade, TBS Business School
Andrew Boysen, University of North Carolina at Chapel Hill, Kenan-Flagler Business School
Craig Brimhall, University of California Los Angeles, Anderson School of Management
Zuzana Brokesova, University of Economics in Bratislava
J. Paul Brooks, Virginia Commonwealth University
Stephan B. Bruns, Hasselt University
Georgia Buckle, UK Office for National Statistics
Guido Buenstorf, University of Kassel
Gordon Burtch, Boston University
Benjamin Bushong, Michigan State University
Sabrina Buti, Université Paris Dauphine - PSL
Patrick Callery, University of Vermont
Mehmet Canayaz, Pennsylvania State University
Jie Cao, Hong Kong Polytechnic University
Wei Cao, Shanghai University of Finance and Economics
Xinyu Cao, The Chinese University of Hong Kong
Martin Carree, Maastricht University, School of Business and Economics
Vincent Castellani, Pennsylvania State University
Yann Joel Cerasi, Norges Bank
Hannah H. Chang, Singapore Management University
Jin Wook Chang, Korea University Business School
Michelle Chang, Nanyang Technological University
Yanru Chang, City University of New York, Baruch College
Aadhaar Chaturvedi, University of Auckland Business School
Jasmina Chauvin, Georgetown University
Daniel E. Chavez, University of Tennessee
Christopher Chen, Indiana University
Fadong Chen, School of Management & Neuromanagement Lab, Zhejiang University
Josie I Chen, National Taiwan University
Peng-Chu Chen, University of Hong Kong
Roy Chen, RWTH Aachen University
Wei Chen, University of Connecticut
Wei James Chen, National Taiwan University, Department of Agricultural Economics
Yuanyuan Chen, University of Alabama
Zepeng Chen, Hong Kong Polytechnic University
Zhuoqiong Chen, Harbin Institute of Technology, Shenzhen
Lydia Chew, Harvard University, Harvard Business School
Param Pal Singh Chhabra, University of Alberta
Sai Chand Chintala, Cornell University
Ga-Young Choi, City University of London
Seungho Choi, Hanyang University; Queensland University of Technology
Vivek Choudhary, Nanyang Technological University, Nanyang Business School
Vincent Tsz Fai Chow, Hong Kong Polytechnic University, Faculty of Business
Katherine L. Christensen, Indiana University, Kelley School of Business
Doug J. Chung, University of Texas at Austin
Melissa Cinelli, University of Mississippi
Lubomír Cingl, Prague University of Economics and Business
Andre Augusto Cire, University of Toronto, Rotman School of Management
Jeffrey Clark, Stockholm School of Economics
Jeffrey Clement, Augsburg University
John Clithero, University of Oregon
Héloïse Cloléry, Ecole Polytechnique IP Paris, CREST
David R. Clough, University of British Columbia
Nicholas Clyde, Washington University in St. Louis
Andrea Coali, Bocconi University
Irene Comeig, University of Valencia
Nikolai Cook, Wilfrid Laurier University
Joao Correia-da-Silva, University of Porto
Elaine Costa, University of Utah
Alexander Coutts, York University
Ivor Cribben, University of Alberta, Alberta School of Business
Carina Cuculiza, Oklahoma State University
Zimeng (Simon) Cui, University of Utah
Colleen Cunningham, University of Utah, Eccles School of Business
Peter Cziraki, Texas A&M University
Étienne Dagorn, National Institute of Demographic Studies (INED)
Rui Dai, University of Pennsylvania, The Wharton School
Jason Dana, Yale University, Yale School of Management
Nicholas Patrick Danks, Trinity College Dublin, Trinity Business School
Alper Darendeli, Nanyang Technological University
Simon Dato, EBS Universität für Wirtschaft und Recht

Nebojsa Davcik, EM Normandie Business School, Metis Lab

Charles de Grazia, Léonard de Vinci Pôle Universitaire, Research Center

Jose De Sousa, Université Paris Panthéon-Assas

Jelle De Vries, Erasmus University, Rotterdam School of Management

Martijn De Vries, Vrije Universiteit Amsterdam

Oleg Deev, Masaryk University

Ryan DeFronzo, California State University, Fullerton

Lennart Dekker, De Nederlandsche Bank

Arthur Delarue, Georgia Institute of Technology, H. Milton Stewart School of Industrial & Systems Engineering

Elif E. Demiral, Austin Peay State University

Cem Demiroglu, Koc University

Aishwarrya Deore, Georgetown University

Andrew Detzel, Baylor University

Azamat Devonaev, University of Luxembourg

Archana Dhinakar Bala, National University of Singapore

Eugen Dimant, University of Pennsylvania

Drew Dimmery, University of Vienna

Stephen G. Dimmock, National University of Singapore

Cheng Ding, Emory University

Likang Ding, University of Alberta

Tingting Ding, James Madison University; Shanghai University of Finance and Economics

Yuheng Ding, University of Maryland

Lu Dong, Southern University of Science and Technology

Karen Donohue, University of Minnesota, Carlson School of Management

Andreas Drichoutis, Agricultural University of Athens

Shaoyin Du, University of North Carolina at Charlotte

Ying Duan, Simon Fraser University

Teodor Duevski, HEC Paris

Huu Nhan Duong, Monash University

Merle Ederhof, University of Zurich, Stanford University

Hussein El Hajj, Santa Clara University, Leavey School of Business

Martin Ellison, University of Oxford

Jonas Nygaard Eriksen, Aarhus University

Miguel Espinosa, Bocconi University

Francesco Fallucchi, University of Bergamo

Xiaohua Fang, Florida Atlantic University

Valeria Fanghella, Grenoble Ecole de Management

Matilde Faralli, Imperial College London

Saleh Farham, University of Alberta

Felix Fattinger, Vienna University of Economics and Business

Stephanie Feiereisen, Montpellier Business School

Yiding Feng, Microsoft Research

Elia Ferracuti, Duke University

Antonio Filippin, University of Milan

Adrien Fillon, University of Cyprus, SInnoPSis

Stefano Fiorin, Bocconi University

Geoffrey Fisher, Cornell University

Matthew Fisher, Southern Methodist University

Christoph Flath, University of Würzburg

Jens Foerderer, Technical University of Munich

Vincenz Frey, University of Groningen, Department of Sociology

Christoph Fuchs, University of Vienna

Nicolas Fugger, University of Cologne

Sebastian Gabel, Erasmus University Rotterdam, Rotterdam School of Management

Fabian Gaessler, Universitat Pompeu Fabra

Bernhard Ganglmair, University of Mannheim

Manish Gangwar, Indian School of Business

Pedro Angel Garcia Ares, Instituto Tecnologico Autonomo de Mexico

Rajiv Garg, Emory University

José Miguel Gaspar, ESSEC Business School

Chiara Gastaldi, Free University of Bozen-Bolzano

Romain Gauriot, Deakin University

Alan De Genaro, Sao Paulo School of Business Administration (FGV-EAESP)

Yuxin Geng, Tsinghua University

Konstantinos Georgalos, Lancaster University Management School

Diogo Geraldes, University College Dublin, School of Economics; Geary Institute for Public Policy

Leonie Gerhards, King's College London

William Gerken, University of Kentucky

Mike Gibson, University of Maryland, Agricultural and Resource Economics Department

Joren Gijsbrechts, Esade; Ramon Llull University

Sebastian Goerg, Technical University of Munich

Daniel Goetz, University of Toronto, Rotman School of Management

Jim Goldman, University of Warwick

Filip Gonschorek, ZEW Leibniz Centre for European Economic Research

Victor Gonzalez-Jimenez, Erasmus University Rotterdam

Jorgo T.G. Goossens, Radboud University Nijmegen, Institute for Management Research; Tilburg University, Department of Econometrics and Operations Research

Michael Gordy, Federal Reserve Board

Paul M. Gorny, Karlsruhe Institute of Technology

Indranil Goswami, University at Buffalo

Amit Goyal, University of Lausanne

Ruslan Goyenko, McGill University

Tom Grad, Copenhagen Business School

Wesley Greenblatt, Massachusetts Institute of Technology, Sloan School of Management

Martin Gregor, Charles University

Daniela Grieco, University of Milano

Manuel Grieder, UniDistance Suisse; Zurich University of Applied Sciences (ZHAW)

Max R. P. Grossmann, University of Cologne

Sven Grüner, University of Rostock

Sreyaa Guha, Universidade NOVA de Lisboa, Nova School of Business and Economics

Audrey Guo, Santa Clara University

Gang Guo, National University of Singapore

Haihao Guo, Washington University in St. Louis

Lewen Guo, University of Memphis

Dominik Gutt, Erasmus University Rotterdam

André F. Gygax, University of Melbourne

Isaac Hacamo, Indiana University

Simone Haeckl, University of Stavanger

Thomas C. Hagenberg, Northwestern University, Kellogg School of Management

David Hagmann, The Hong Kong University of Science and Technology

Jacob Haislip, Texas Tech University

Eojin Han, Southern Methodist University, Operations Research and Engineering Management

Jiatong Han, Zhejiang University; School of Management & Neuromanagement Lab

Joseph Earle Harvey, Consumer Financial Protection Bureau

Olena Havrylchyk, Université Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne

Sonali Hazarika, City University of New York, Baruch College

Leshui He, Bates College

Yuhang He, Nanyang Technological University, Nanyang Business School

William Hedgcock, University of Minnesota

Irina Heimbach, WHU Otto Beisheim School of Management

Brian Henderson, George Washington University

Jurian Hendrikse, Tilburg University

Erin Henry, University of Arkansas

Bradford Hepfer, The University of Iowa

Roberto Hernan, Burgundy School of Business

Holger Herz, University of Fribourg

Anthony Heyes, University of Birmingham

Christian Hildebrand, University of St. Gallen, Institute of Behavioral Science & Technology

Adrian Hillenbrand, Karlsruhe Institute for Technology; Leibniz Centre For European Economic Research

Alexander Hillert, Goethe University Frankfurt; Leibniz Institute for Financial Research SAFE

Michael Hilweg, University of Mannheim

Erik Hjalmarsson, University of Gothenburg

Seth Hoelscher, Missouri State University

Peter Hoffmann, European Central Bank

Brett Hollenbeck, University of California Los Angeles, Anderson School of Management

Niels Holtrop, Maastricht University

Felix Holzmeister, University of Innsbruck, Department of Economics

Swarnodeep Homroy, University of Groningen

Mallick Hossain, Federal Reserve Bank of Philadelphia

Leon Houf, Heidelberg University

Taeya Howell, Brigham Young University, Marriott School of Business

Kejia Hu, University of Oxford

Allen Huang, Hong Kong University of Science and Technology

Jing-Zhi Huang, Pennsylvania State University

Lingbo Huang, Shandong University

Sterling Huang, Singapore Management University

Stefanie J. Huber, University of Bonn

Stanton Hudja, University of Toronto

Jacquelyn Humphrey, University of Queensland

Paul Hünermund, Copenhagen Business School

William Reuben Hurst, University of Michigan, Ross School of Business

Carlos Hurtado, University of Pittsburgh

Kim P. Huynh, Bank of Canada

Kyle Hyndman, University of Texas at Dallas

Armann Ingolfsson, University of Alberta

Panos Ipeirotis, New York University

Ayelet Israeli, Harvard University, Harvard Business School
Alexey Ivashchenko, Vrije Universiteit Amsterdam
Wael Jabr, Pennsylvania State University
Pankaj K. Jain, University of Memphis
Ainhoa Jaramillo-Gutierrez, University Jaume I Castellon
Nahid Javadinarab, University of Luxembourg
Yonghua Ji, University of Alberta
Mofei Jia, Xi'an Jiaotong-Liverpool University
Hansheng Jiang, University of Toronto
Houyuan Jiang, University of Cambridge, Judge Business School
Jiashuo Jiang, Hong Kong University of Science and Technology
Jingdan Tan, Nanyang Technological University
Michal Jirásek, Masaryk University
Brandon Julio, University of Oregon
Heejung (HJ) Jung, Imperial College London, Business School
Daniel Marcel te Kaat, University of Groningen
Jonathan Kalodimos, Oregon State University
Mark Kamstra, York University, Schulich School of Business
Hyo Kang, University of Southern California
Qiang Kang, Florida International University
Salpy Kanimian, Rice University
Martin M. Kapons, University of Amsterdam
Egle Karmaziene, Vrije Universiteit Amsterdam; Swedish House of Finance; Tinbergen Institute
Asad Kausar, American University
Patrick J Kelly, University of Melbourne
Saravanan Kesavan, University of North Carolina at Chapel Hill
Menusch Khadjavi, Vrije Universiteit Amsterdam; Kiel Institute for the World Economy
Hamid Khobzi, University of Sussex
Robizon Khubulashvili, University of San Francisco
Alex G. Kim, University of Chicago
Byungyeon Kim, University of Minnesota
Chungyool Kim, University of Iowa
Dong Soo Kim, Ohio State University
Sehoon Kim, University of Florida
Seojin Kim, Drexel University
Seung Hyun Kim, Yonsei University, School of Business
Soohun Kim, Korea Institute of Advanced Science and Technology
Margarita Kirneva, Ecole Polytechnique, CREST; ENSAE Paris
Andrea Kiss, Carnegie Mellon University
Leonardo Mayer Kluppel, Ohio State University
Özgecan Koçak, Emory University
Christoph Kogler, Tilburg University
Christian König-Kersting, University of Innsbruck
Anita Kopányi-Peuker, Radboud University Nijmegen, Institute for Management Research
Lina Koppel, Linköping University
Sharon Koppman, University of California Irvine
Orestis Kopsacheilis, Technical University of Munich
Laura J. Kornish, University of Colorado Boulder, Leeds School of Business
Anne Krahn, Tufts University
Ondřej Krčál, Masaryk University
Srinivasan Krishnamurthy, North Carolina State University
Philipp Kropp, University of Munich
Santanu Kundu, University of Mannheim
Michael Kurschilgen, UniDistance Suisse
David J. Kusterer, Erasmus University Rotterdam, Rotterdam School of Management
Samet Kutuk, Vrije Universiteit Amsterdam
Olga Kuzmina, New Economic School
Ellie Kyung, Babson College
Camille Lacan, CRESEM; IAE School of Management; University of Perpignan Via Domitia
Adrian Lam, University of Pittsburgh
Thomas Lambert, Erasmus University Rotterdam
Lauren Lanahan, University of Oregon
Mike Langen, CPB Netherlands Bureau for Economic Policy Analysis
Nadzeya Laurentsyeva, Ludwig-Maximilians-Universität München
Kelvin K. F. Law, Nanyang Technological University
Quoc Thai Le, University of Trento, Department of Economics and Management
Choonsik Lee, University of Rhode Island
Daniel Lee, University of Delaware
Kyeong Hun Lee, University of Alabama, Culverhouse College of Business
Sunkee Lee, Carnegie Mellon University, Tepper School of Business
Yeonjoo Lee, University of Minnesota, Carlson School of Management
Murray Lei, Queen's University

Zhou Lei, Nanyang Technological University, Nanyang Business School
Stephan Leitner, University of Klagenfurt
Gabriele Mario Lepori, University of Southampton
David E. Levari, Harvard University, Harvard Business School
Ben William Lewis, Brigham Young University
Benjamin T. Leyden, Cornell University
Chenghuai Li, Duke University, Fuqua School of Business
Jiasun Li, George Mason University
King King Li, Shenzhen University, Shenzhen Audencia Financial Technology Institute
Linfeng Li, University of Michigan
Meng Li, University of Houston
Shukai Li, Northwestern University
Shuo Li, Singapore Management University
Ye Li, University of California Riverside
Yushen Li, Jinan University, Institute of Industrial Economics
Chuchu Liang, University of California, Irvine
Stanley Lim, Michigan State University
Mingfeng Lin, Georgia Tech
Po-Hsuan Lin, California Institute of Technology
Yunduan Lin, University of California Berkeley
Sera Linardi, University of Pittsburgh
William Lincoln, Claremont McKenna College
Michaela Lindenmayr, Technical University of Munich
Martina Linnenluecke, University of Technology Sydney
Ariel Listo, University of Maryland
Robin Litjens, Tilburg University
Chengwei Liu, European School of Management and Technology
Dingyue (Kite) Liu, University of California Santa Barbara
Fang Liu, University of the Chinese Academy of Sciences
Haibo Liu, Claremont Colleges, Keck Graduate Institute
Haiyang Liu, Nanyang Technological University
Jiaxin Liu, Morgan State University
Kaiqi Liu, Maastricht University, Department Microeconomics and Public Economics
Nan Liu, Boston College
Sheng Liu, University of Toronto
Xiaojin Liu, Virginia Commonwealth University
Neta Livneh, Tel Aviv University
Tatiana Lluent, European School of Management and Technology
Nils Loehndorf, University of Luxembourg
Matthijs Lof, Aalto University, School of Business
Youenn Loheac, Rennes School of Business
Paul Lohmann, University of Cambridge, Judge Business School
Luis Arturo Lopez, University of Illinois at Chicago
Matej Lorko, University of Economics in Bratislava; Prague University of Economics and Business
Francesca Lotti, Bank of Italy, DG Economics, Statistics and Research
Joy Lu, Carnegie Mellon University
Xinyu Lu, HEC Paris
Jonathan Luffarelli, Montpellier Business School
Wolfgang J. Luhan, University of Portsmouth
Hoang Luong, University of Queensland
Guodong Lyu, Hong Kong University of Science and Technology
Liang Ma, San Diego State University
Leonardo Madio, University of Padova
Kai Maeckle, University of Mannheim
Mahdi Mahmoudzadeh, University of Auckland Business School
Patrick Maillé, IMT Atlantique
Vincent Mak, University of Cambridge, Cambridge Judge Business School
Antoine Malézieux, Burgundy School of Business
Shawn Mankad, North Carolina State University
César Mantilla, Universidad del Rosario
Benny Mantin, University of Luxembourg
Marco Mantovani, Università degli Studi di Milano Bicocca, Dipartimento di Economia
Giacomo Marchesini, Copenhagen Business School
Juri Marcucci, Bank of Italy
Diego Marino Fages, Durham University
Aidas Masiliunas, University of Sheffield
Sébastien Massoni, Université de Lorraine; Université de Strasbourg; CNRS; BETA
Nunez Matias, Ecole Polytechnique, CREST; CNRS
Thomas Matthys, University of Technology Sydney
Martin Mattsson, National University of Singapore
Thomas Andreas Maurer, University of Hong Kong
Patrick Maus, University of Nottingham
Merve Mavuş Kütük, University of Amsterdam
Malte M. Max, Vrije Universiteit Amsterdam
Christoph Meinerding, Deutsche Bundesbank
Matt Meister, University of Colorado Boulder; University of San Francisco
Dong Meitong, University of Hong Kong

Eduardo Melero, Universidad Carlos III de Madrid
Diogo Mendes, Stockholm School of Economics
Tyler Menzer, University of Iowa
Christoph Merkle, Aarhus University
Jason Merrick, Virginia Commonwealth University
Steffen Meyer, Aarhus University; Danish Finance Institute
Tomáš Miklánek, Prague University of Economics and Business
Wladislaw Mill, University of Mannheim
Stefan Minner, Technical University of Munich
Emil Mirzayev, University College London, School of Management
Sergio Mittlaender, Fundação Getulio Vargas Law School in São Paulo; Max Planck Institute for Social Law and Social Policy
Stig Vinther Møller, Aarhus University
Andras Molnar, University of Michigan, Department of Psychology
David Moore, Loyola Marymount University
Sandra Mortal, University of Alabama
Giovanni Moscariello, Stockholm School of Economics
Yuting Mou, Southeast University
Jifeng Mu, Alabama A&M University
Clemens Mueller, University of Mannheim
Anirban Mukherjee, Cornell University; INSEAD
Sara Mustafazade, University of Montpellier
Kumar Muthuraman, University of Texas-Austin
Alper Nakkas, University of Texas at Arlington
Jim Naughton, University of Virginia
Hunter Boon Hian Ng, City University of New York, Baruch College
Lily Nguyen, University of Queensland
Mike Nguyen, University of Southern California
Ngoc Phuong Anh Nguyen, University of Technology Sydney
Thi Thuy Tien Nguyen, University of Auckland
Amy Nguyen-Chyung, University of California San Diego, Rady School of Management
Nicos Nicolaou, University of Warwick
Sven Nolte, Radboud University Nijmegen
Arjan Non, Erasmus University Rotterdam
Bernt Arne Ødegaard, University of Stavanger
Yuval Ofek-Shanny, Friedrich-Alexander-Universität Erlangen-Nürnberg
Chang Hoon Oh, University of Kansas
Christopher Yves Olivola, Carnegie Mellon University
Thomas C. Omer, University of Nebraska-Lincoln
Andreas Orland, Corvinus University of Budapest
Tizian Otto, Yale University; University of Hamburg
Manlu Ouyang, New York University, Stern School of Business
Hakan Ozyilmaz, Toulouse School of Economics
Nicholas A. Pairolero, United States Patent and Trademark Office
Stefan Palan, University of Graz
Navya Pandit, University of Cologne
Dominik Papies, University of Tuebingen, School of Business and Economics
Jiyong Park, University of North Carolina at Greensboro
Tae-Youn Park, Sungkyunkwan University
Chris Parker, American University
Vinay Patel, University of Technology Sydney
Grzegorz Pawlina, Lancaster University
Elise Payzan-Le Nestour, University of New South Wales
Graeme Pearce, Bangor University
Thomas Peeters, Erasmus University Rotterdam, Erasmus School of Economics; Tinbergen Institute; Erasmus Research Institute in Management
Jana Peliova, University of Economics in Bratislava
Zhuozhen Peng, Central University of Finance and Economics
Christophe Pérignon, HEC Paris
Noemi Peter, University of Groningen
Christian Peukert, University of Lausanne, Faculty of Business and Economics (HEC)
Hieu Phan, University of Massachusetts Lowell
Aviva Philipp-Muller, Simon Fraser University
Kenny Phua, University of Technology Sydney
Matthew Pierson, University of Pennsylvania, The Wharton School
Tomáš Plíhal, Masaryk University
Matteo Ploner, University of Trento, Department of Economics and Management
Simon Porcher, Université Paris Panthéon-Assas
Matthieu Pourieux, Rennes School of Business; Univ Rennes, CNRS, CREM-UMR6211
Susanne Preuss, University of Amsterdam
Jakub Procházka, Masaryk University, Faculty of Economics and Administration
Shaolin Pu, University of Kansas, School of Business
Žiga Puklavec, Tilburg University

Hanzhang Qin, Amazon; National University of Singapore
Tian Qiu, University of Alabama
Xincheng Qiu, University of Pennsylvania
Rima-Maria Rahal, Max Planck Institute for Research on Collective Goods
Amin Rahimian, University of Pittsburgh
Mohammadreza Rajabzadeh, York University, Schulich School of Business
Oliver Randall, University of Melbourne
Soumya Ray, National Tsing Hua University, Institute of Service Science
Oliver Rehbein, Vienna University of Economics and Business
Jurij-Andrei Reichenecker, University of Strathclyde
Nicholas Reinholtz, University of Colorado Boulder
J. Philipp Reiss, Karlsruhe Institute of Technology
Jean-Paul Renne, University of Lausanne
Sadat Reza, Nanyang Technological University
Paul Richardson, Pennsylvania State University
Steven Riddiough, University of Toronto
Marc Oliver Rieger, University of Trier; University of Economics Ho Chi Minh City
Cesare Righi, Universitat Pompeu Fabra, Department of Economics and Business; UPF Barcelona School of Management; Barcelona School of Economics
Rainer Michael Rilke, WHU Otto Beisheim School of Management
Julio Riutort, Universidad Adolfo Ibáñez
Cesare Robotti, University of Warwick
Nathalie Römer, Leibniz University Hannover
Julia Rose, Erasmus University Rotterdam, Erasmus School of Economics; Tinbergen Institute
Michael Rose, Max Planck Institute for Innovation and Competition
Paul Rosmer, Ludwig-Maximilians-Universität München
Federico Rossi, Purdue University
Borzou Rostami, University of Alberta
Kasper Roszbach, Norges Bank; University of Groningen
Kristian Rotaru, Monash University, Monash Business School
Yefim Roth, University of Haifa
Daniele Rotolo, University of Sussex; Technical University of Bari
Christina Rott, Vrije Universiteit Amsterdam; Tinbergen Institute
Bryan Routledge, Carnegie Mellon University
Brian Rubineau, McGill University
Hannes Rusch, Maastricht University
Ilya O. Ryzhov, University of Maryland
Pedro Saffi, University of Cambridge, Judge Business School
Mehmet Saglam, University of Cincinnati
Margaret Samahita, University College Dublin
Panagiotis Sarantopoulos, Athens University of Economics and Business; University of Manchester
Vahid Sarhangian, University of Toronto
Secil Savasaneril, Middle East Technical University, Industrial Engineering Department
Harald Scheule, University of Technlogy Sydney
Maximilian Schleritzko, Vienna Graduate School of Finance
Max Schnidman, University of Virginia
Daniela Stephanie Schoch, emlyon business school
Marina Schröder, Leibniz University Hannover
Erik Christian Montes Schütte, Aarhus University; Danish Finance Institute
Daniel Schwartz, University of Chile
Frederik Schwerter, Frankfurt School of Finance and Management
Robert Seamans, New York University
Matthias Seifert, IE University, IE Business School
Tom Servranckx, Ghent University, Faculty of Economics and Business Administrations
Nagarajan Sethuraman, University of Kansas
Victoria Sevcenko, INSEAD
Divyesh Rajendra Shah, University of Toronto
Rachna Shah, University of Minnesota
Kartikey Sharma, Zuse Institute Berlin
Padma Sharma, Federal Reserve Bank of Kansas City
Amy Sheneman, Ohio State University
Yunting Shi, Shanghai Jiao Tong University, Antai College of Economics and Management
Ling Shuai, Tianjin University
Simon Siegenthaler, University of Texas at Dallas
John Silberholz, University of Michigan
Rui Silva, University of East Anglia
Katherine Silz-Carson, U.S. Air Force Academy
Felipe Simon, University of Minnesota
Raghav Singal, Dartmouth College, Tuck School of Business
Nitish Ranjan Sinha, Board of Governors of the Federal Reserve System

Spyros Skouras, Athens University of Economics and Business
David Smerdon, University of Queensland
Katrin Smolka, University of Warwick, Warwick Business School
Adriaan Soetevent, University of Groningen
Elvira Sojli, University of New South Wales
Konstantin Sokolov, University of Memphis
Jeeva Somasundaram, IE Business School
Yoonseock Son, University of Notre Dame
Ju Myung Song, University of Massachusetts Lowell
Vikas Soni, University of South Florida
Doron Sonsino, University of Limassol, Cyprus
Matthew Souther, University of South Carolina
Christophe Spaenjers, University of Colorado Boulder
Martin Spann, Ludwig-Maximilians-Universität München, LMU Munich School of Management
Eirini Spiliotopoulou, Tilburg University
Jeffrey Starck, University of Cologne
Austin Starkweather, University of South Carolina
Dayton Steele, University of Minnesota, Carlson School of Management
Matthias Stefan, University of Innsbruck
Frauke Stehr, Maastricht University
Eva Steiner, Pennsylvania State University
Lucas Stich, Julius-Maximilians-Universität Würzburg
Thomas Stoeckl, MCI The Entrepreneurial School
Jan Stoop, Erasmus University Rotterdam, Erasmus School of Economics
Karoline Ströhlein, University of Regensburg
Robert Stüber, New York University Abu Dhabi
Jason Sturgess, Queen Mary University of London
Yuhan Su, Tianjin University
Yuxin Su, SKEMA Business School
Rémi Suchon, Université Catholique de Lille
Mengtian Sui, City University of New York, Baruch College
Sandra Sülz, Erasmus University Rotterdam, Erasmus School of Health Policy & Management
Elie Sung, HEC Paris
Marta Szymanowska, Erasmus University, Rotterdam School of Management
Giovanni Alberto Tabacco, Freelance researcher
David Tannenbaum, University of Utah
Necati Tereyagoglu, University of South Carolina, Darla Moore School of Business
Chloe Tergiman, Pennsylvania State University
Marco Testoni, Miami Herbert Business School, University of Miami
Richard Thakor, University of Minnesota; Massachusetts Institute of Technology, Laboratory for Financial Engineering
Wing Wah Tham, University of New South Wales
Samuel Thelaus, London School of Economics
Simon Thielen, MCI The Entrepreneurial School
Lu Tong, Southwestern University of Finance and Economics
Ozlem Tonguc, Binghamton University
Mirco Tonin, Free University of Bozen-Bolzano
Sinem Yagmur Toraman, Johns Hopkins University, Department of Economics
Marco Tortoriello, Bocconi University
J. Dustin Tracy, Augusta University
James Tremewan, IESEG School of Management
Muktak K. Tripathi, Temple University
Gunseli Tumer-Alkan, Vrije Universiteit Amsterdam
Danko Turcic, University of California Riverside
Theodore Turocy, University of East Anglia
Hanu Tyagi, University of Minnesota
Maximiliano Udenio, KU Leuven
Sezer Ulku, Georgetown University, McDonough School of Business
Michael Ungeheuer, Aalto University
Steven Utke, University of Connecticut
Cihan Uzmanoglu, SUNY, Binghamton University
Matteo Vacca, Aalto University, School of Business
Philip Valta, University of Bern
Michel Van Der Borgh, Copenhagen Business School
Jesse Van Der Geest, Tilburg University
Milan Van Steenvoort, Maastricht University
Roel Van Veldhuizen, Lund University
Prasad Vana, Dartmouth College, Tuck School of Business
Mario Vanhoucke, Ghent University; Vlerick Business School; University College London
Bart Vanneste, University College London
Joseph Vecci, Gothenburg University
Sriram Venkataraman, University of South Carolina, Darla Moore School of Business
Marcella Veronesi, Technical University of Denmark; University of Verona
Sergio Vicente, University of Luxembourg
Sebastian Villa, University of New Mexico
Marta Villamor Martin, University of Maryland

Lynne Vincent, Syracuse University

Theodor Vladasel, Universitat Pompeu Fabra, Barcelona School of Economics

Stefan Voigt, University of Copenhagen

Joachim Vosgerau, Bocconi University

Christian A. Vossler, University of Tennessee

Angela Vossmeyer, Claremont McKenna College

Hannes F. Wagner, Bocconi University

David M. Waguespack, University of Maryland

Edward Walker, University of California Los Angeles

Matthew Walker, Newcastle University

Markus Walzl, University of Innsbruck

Zhixi Wan, University of Hong Kong

Charles C.Y. Wang, Harvard University, Harvard Business School

Joseph Tao-Yi Wang, National Taiwan University, Department of Economics

Kanix Wang, University of Cincinnati

Victor Xiaoqi Wang, California State University Long Beach

Xiaohong Wang, University of Pittsburgh

Yiwei Wang, Zhejiang University

Xavier S. Warnes, Stanford University

Lilia Wasserka-Zhurakhovska, University of Duisburg-Essen

Wei Wei, University of Oklahoma

Stefan Weiergraeber, Indiana University, Department of Economics

Patrick Weiss, Reykjavik University

Jingjing Weng, Temple University

Wei-Chien Weng, National Taiwan University

James Weston, Rice University

Joshua Tyler White, Vanderbilt University

Matthias Wibral, Maastricht University

Jared Williams, University of South Florida

Ole Wilms, Hamburg University; Tilburg University

Franz Wirl, University of Vienna

Adrian Wolanski, University of California San Diego, Department of Economics

M.H. Franco Wong, University of Toronto

Daniel John Woods, University of Innsbruck

Biyu Wu, University of Nebraska-Lincoln

Yiran Wu, Vrije Universiteit Amsterdam

Ziye Wu, National University of Singapore

David Wuttke, Technical University of Munich, TUM School of Management, TUM Campus Heilbronn

Yuze Xia, Northwestern University, Kellogg School of Management

Jingui Xie, Technical University of Munich

Wen Xie, City University of New York, Baruch College

Feiyu Xu, Hong Kong University of Science and Technology

Luze Xu, University of California Davis

Sikun Xu, Washington University in St. Louis

Simon Xu, Harvard University, Harvard Business School

Yilong Xu, Utrecht University School of Economics, Utrecht University

Rui Xue, La Trobe University

Beril Yalcinkaya, University of Maryland

Ruijing Yang, Chinese University of Hong Kong

Yadi Yang, Nanjing Audit University

Huang Yao, Central South University, Business School; Hunan Agricultural University, College of Economics

Shiqing Yao, Monash University

Yaojun Ke, Nanyang Technological University

Ozge Yapar, Indiana University, Kelley School of Business

Eduard Yelagin, University of Memphis

Ira Yeung, University of British Columbia

Erdem Dogukan Yilmaz, Erasmus University Rotterdam

Levent Yilmaz, Turkish-German University

Woongsun Yoo, Central Michigan University

Simon (Seongbin) Yoon, University of California Irvine

Sora Youn, Texas A&M University

Alex Young, Hofstra University

Jin Yu, Monash University

Jungju Yu, Korea Advanced Institute of Science and Technology

Junhao Vincent Yu, Miami University, Farmer School of Business

Lizi Yu, University of Queensland

Huaiping Yuan, The Chinese University of Hong Kong-Shenzhen, SME and SFI

Yuan Yuan, Purdue University

Lei Yue, University of California Santa Barbara

Anita Zednik, Vienna University of Economics and Business

Yasser Zeinali, University of Alberta

Shenghui Zhai, University of the Chinese Academy of Sciences

Xintong Zhan, Fudan University

Aiqi Zhang, Wilfrid Laurier University, Lazaridis School of Business and Economics

Chengyu Zhang, McGill University
Huanan Zhang, University of Colorado Boulder
Huanren Zhang, University of Southern Denmark
Hulai Zhang, Tilburg University; ESCP Business School
Jack H. Zhang, Nanyang Technological University
Le (Lyla) Zhang, Macquarie University
Quan Zhang, Nanyang Technological University
Renyu Zhang, Chinese University of Hong Kong
Ruishen Zhang, Shanghai University of Finance and Economics
Shu Zhang, Shanghai University of Finance and Economics
Sili Zhang, Ludwig-Maximilians-Universität München
Walter W. Zhang, University of Chicago, Booth School of Business
Zhiqi Zhang, Washington University in St. Louis, Olin Business School
Jiayu (Kamessi) Zhao, Massachusetts Institute of Technology, Operations Research Center
Xiaofei Zhao, Georgetown University
Zhongyu Zhao, University of Hong Kong
Jiakun Zheng, Renmin University of China, School of Finance
Yaping Zheng, McGill University
Zhanzhi Zheng, University of North Carolina at Chapel Hill, Kenan–Flagler Business School
Aner Zhou, San Diego State University
Hongyi Zhu, University of Texas at San Antonio
Jason Zhu, Microsoft
Yayongrong Zhu, University of Queensland
Christian Zihlmann, University of Fribourg, Berne Business School
Marius Zoican, University of Toronto
Ro'i Zultan, Ben-Gurion University of the Negev
Zhuan Zuo, University of the Chinese Academy of Sciences

B Additional tables and figures

TABLE B.1: Software used in articles with and without report

	Has Report ($N = 459$)	No Report ($N = 30$)
Stata	60.1%	43.3%
R	19.2%	23.3%
Matlab	17.9%	26.6%
SAS	12.9%	13.3%
Python	10.7%	13.3%
Mathematica	1.7%	6.7%
SPSS	1.3%	0.0%
Other	5.7%	13.3%

TABLE B.2: Reasons for non-reproducibility for articles with replication package, by policy

	Before 2019 policy ($N = 18$)	Since 2019 policy ($N = 136$)
No access to dataset.	61.1%	88.2%
Issues with software/hardware requirements.	5.6%	2.9%
Code or parts of code/functions missing.	55.6%	12.5%
Insufficient documentation, missing information.	11.1%	7.4%
Unresolvable errors when executing code.	11.1%	5.1%
Reproduction yields (partly) different results.	11.1%	4.4%

TABLE B.3: Distribution of article types/methods for each journal department, since 2019 policy

		Lab/online experiment	Theory /Simulation /Computation	Survey study	Field experiment	Empirical data
SMS	($N = 5$)	0	100	0	0	0%
BDE	($N = 66$)	70	3	5	8	15%
ENI	($N = 10$)	10	0	0	0	90%
RMA	($N = 19$)	0	84	0	0	16%
ACC	($N = 57$)	7	0	2	0	91%
OPM	($N = 38$)	11	32	5	11	42%
OPT	($N = 6$)	0	100	0	0	0%
BDA	($N = 14$)	0	100	0	0	0%
FIN	($N = 124$)	5	15	1	1	78%
HCM	($N = 16$)	0	19	0	0	81%
INS	($N = 19$)	0	11	5	11	74%
MKG	($N = 20$)	10	5	0	15	70%
ORG	($N = 13$)	0	8	8	0	85%
BST	($N = 12$)	0	8	8	25	58%
Total	($N = 419$)	15	20	2	4	59%

Note: Department acronyms are SMS: Stochastic Models and Simulations, BDE: Behavioral Economics and Decision Analysis, ENI: Entrepreneurship and Innovation, RMA: Revenue Management and Market Analytics, ACC: Accounting, OPM: Operations Management, OPT: Optimization, BDA: Big Data Analytics/Data Science, FIN: Finance, HCM: Healthcare Management, INS: Information Systems, MKG: Marketing, ORG: Organizations, BST: Business Strategy.

C Robustness analyses

In Tables C.1 and C.2 we replicate our main results reported in Section III (Figure 1 and Table 4) based on different samples from the set of all submitted reports. In Table C.1, as a “lower” bound we report the distribution of overall assessments when using the lower assessment whenever we have obtained two reports for an article. As a randomized approach (“rand.”), we report the distribution of assessments which we obtain when simulating 10,000 replications of the dataset, in each of which one report is randomly selected when multiple reports are available. The “upper” bound is represented by the case where we select the higher assessment whenever we have two reports for an article (as reported in Figure 1).

The first three result columns in Table C.1 only consider reports for verifiable articles (i.e., where data was available if needed, and soft- and hardware requirements were met) that were subject to the 2019 disclosure policy. The second set of three columns also includes reports for non-verifiable articles, and the third set focuses on reports on articles that were accepted before the disclosure policy was introduced and voluntarily provided replication materials.

Differences between the three approaches to aggregating multiple reports (lower bound, randomized, upper bound) are in the expected direction but small in size. Compared to taking the higher overall assessment with a share of fully or largely reproduced articles of 95.3% for verifiable articles, this number is 91.4% when taking the lower assessment, and 93.8% when randomizing which of two assessments is considered. Similarly, the numbers for all assessed articles and articles from before the 2019 policy change do not vary much.

The regressions reported in Table C.2, assessing the disclosure policy effect at the report level while clustering standard errors at the article level to account for multiple reports per article, replicate our results at the article level (reported in Table 4 in the main text).

TABLE C.1: Robustness checks on overall article reproducibility assessments

	Since 2019 policy, verifiable articles ($N = 297$)			Since 2019 policy, all assessed articles ($N = 419$)			Before 2019 policy, all assessed articles ($N = 40$)		
	lower	rand.	upper	lower	rand.	upper	lower	rand.	upper
Not verifiable				29.4%	26.7%	23.9%	15.0%	12.5%	10.0%
Largely not verifiable				6.4%	6.0%	5.3%	17.5%	17.5%	17.5%
Not reproduced	4.5%	3.0%	2.0%	2.9%	2.0%	1.4%	10.0%	10.0%	10.0%
Largely not reproduced, with major issues	4.1%	3.2%	2.7%	2.6%	2.2%	1.9%	10.0%	8.8%	7.5%
Largely reproduced, with minor issues	68.4%	60.1%	52.2%	43.9%	40.5%	37.0%	37.5%	35.0%	32.5%
Fully reproduced	23.0%	33.7%	43.1%	14.8%	22.7%	30.5%	10.0%	16.2%	22.5%
Fully or largely reproduced	91.4%	93.8%	95.3%	58.7%	63.2%	67.5%	47.5%	51.2%	55.0%

Note: The percentage values in columns "lower" ("upper") are the result of only considering the more negative (positive) report in case there are two reports for the same article. The "upper" columns thus correspond to the results in Figure 1 in the main text. The values in columns "rand." are the result of 10,000 replications in each of which one report was randomly selected when there are two reports for the same article.

TABLE C.2: Regressing reproducibility on disclosure policy existence, report level

Model	(1)		(2)		(3)	
Sample of articles	All incl. no package		All with package		All verifiable	
	Coeff	StdErr	Coeff	StdErr	Coeff	StdErr
Constant	0.098***	(0.020)	0.547***	(0.077)	0.778***	(0.069)
Policy	0.526***	(0.031)	0.077	(0.081)	0.159**	(0.070)
Report observations	1,045		753		504	
R^2	0.251		0.002		0.029	

Note: The dependent variable is a binary indicator whether the article was classified as "fully reproduced" or "largely reproduced", or not. Standard errors are clustered at the article level. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

In addition to an overall assessment, we asked our reviewers to provide individual assessments for each table and figure in the article that are based on code and/or data analysis, and a summary assessment of other analyses reported in the manuscript (that is, how many of those results they could reproduce). Many reviewers did so, but not all. Some articles only included figures and/or tables that were not based on code or data analysis. As a result, the sample size in terms of articles is slightly lower for this analysis.

Table C.3 shows that, as to be expected, overall assessments and individual assessments are highly correlated. If an article was overall classified as "Fully reproduced," then more than 99% of tables and figures and more than 92% of other results could be reproduced. If an article was overall classified as "Not reproduced," the shares of reproduced tables, figures, and other results are 3%, 8%, and 25%, respectively.

TABLE C.3: SHARE OF TABLES, FIGURES, AND OTHER RESULTS ASSESSED AS AT LEAST LARGELY REPRODUCIBLE, BY OVERALL REPRODUCIBILITY ASSESSMENT, SINCE 2019 POLICY

	Tables ($N = 374$)	Figures ($N = 301$)	Other Results ($N = 145$)
Fully reproduced	99.1 %	99.7 %	92.3 %
Largely reproduced, with minor issues	86.6 %	84.9 %	63.4 %
Largely not reproduced, with major issues	12.0 %	30.5 %	0.0 %
Not reproduced	2.7 %	7.5 %	23.7 %

Figures C.1, C.2, and C.3 show the distribution of assessment outcomes for tables, figures, and other results, respectively, for different samples. The first panel of each figure displays the distributions over all tables, all figures, and all other results, respectively. To account for the fact that articles differ substantially in the number of included tables and figures, for the second panel of each figure we first calculate the distribution of assessment outcomes for each article (using the report with the higher overall assessment, as above), and then average over all articles. In the third panel, we only consider articles which have been deemed verifiable (i.e., for which the dataset was available to the reviewer and soft- and hardware requirements could be met).

We find that it makes little difference how we aggregate individual results, in particular for tables and figures. The share of at least largely reproduced tables is 58-62% (depending on the aggregation method) for all articles, and 88% when considering verifiable articles only. For figures, these shares are 68-70% for all articles and 90% for verifiable articles. For other results we only distinguish between reproducible and not reproducible and results are based on a smaller sample (not all articles report other results, and not all reviewers assessed other results). The respective numbers here are 66-83% for all articles and 75% for verifiable articles.

FIGURE C.1: Reproducibility assessments of tables, since 2019 policy

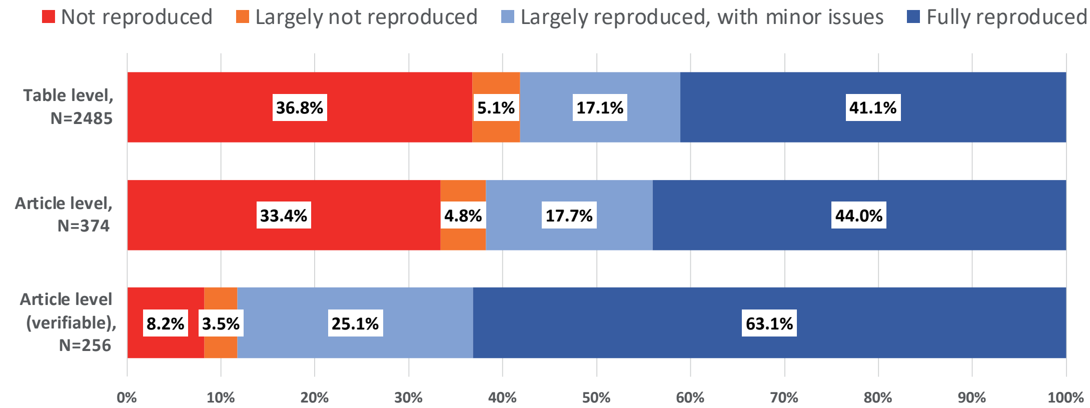


FIGURE C.2: Reproducibility assessments of figures, since 2019 policy

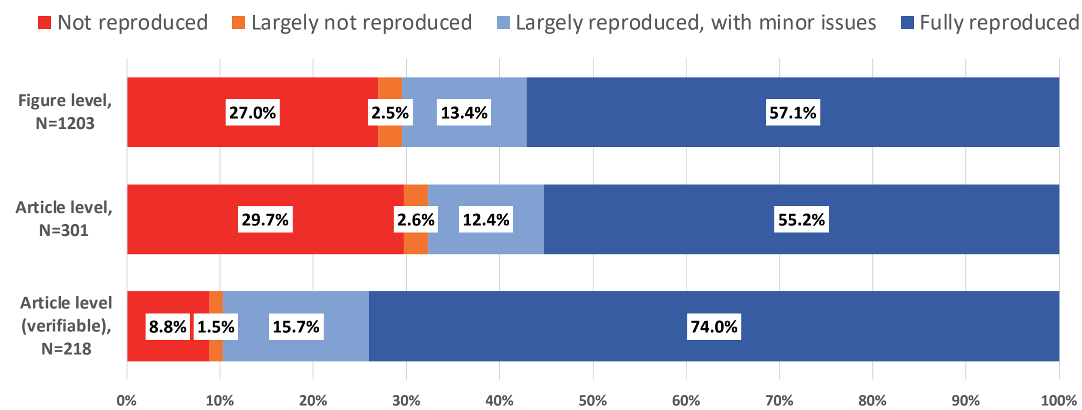
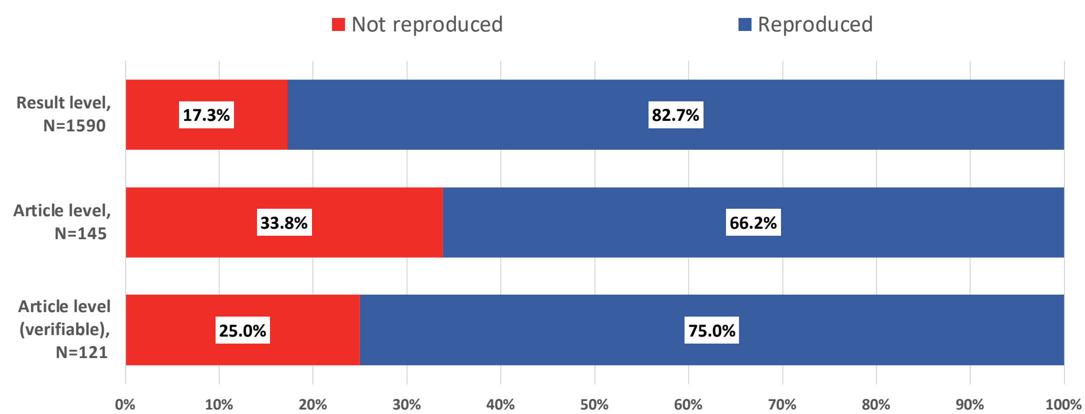


FIGURE C.3: Reproducibility assessments of other Results, since 2019 policy



D Reviewer consistency

For articles for which we were able to obtain two reviews, Table D.1 displays the assessments of the reviewer with the higher assessment and the second reviewer (with the same or lower assessment). Among the 120 reviewer pairs with different assessments, the reviewer with the lower assessment of reproducibility rated the straightforwardness of the reproduction lower (avg. of 71.7 vs. 80.9 on a scale 0-100, $p < 0.001$), was (weakly significantly) less likely to rate the readme file as sufficient ($p = 0.063$), and rated their own methodological expertise as lower (avg. of 80.9 vs. 84.8 on a scale 0-100, $p < 0.001$). No differences between reviewers with lower and higher rating were found with respect to time spent on the review (9.2 vs. 10.4 hours, $p = 0.478$), and for their self-assessed expertise in the topic of the article ($p = 0.842$).

TABLE D.1: Reviewer consistency

	Reviewer with (weakly) higher assessment			
Reviewer with (weakly) lower assessment	Fully	Largely	Largely not	Not
Fully reproduced.	31			
Largely reproduced, with minor issues.	64	65		
Largely not reproduced, with major issues.	5	20	8	
Not reproduced.	2	13	16	70

E Project documentation

E.1 Reviewer Invitation Emails

Invitation email to *Management Science* reviewers

Dear First Name,

As you may know, recently *Management Science* initiated the *Management Science Reproducibility Project (ManSciReP)*. In this project, we assess the computational reproducibility of studies published in the journal. Since 2020, the Code & Data Editor verifies that replication materials are provided but does not attempt reproduction itself. In this project, we aim to quantify the reproducibility of results published in Management Science articles before and after the new Data and Code Disclosure Policy came into effect.

I am writing to see if you would be willing to review a replication package of a paper recently accepted for publication in Management Science. You are receiving this email because you have served as a reviewer for Management Science before.

If you are willing to review, we would assign you a paper from your own field of research, and using software that you are familiar with. We would then ask you to report back within 4-6 weeks to what extent you were able to reproduce the paper's main results, and what the obstacles were.

This call for reviewers is open to any researcher in the community, including advanced Ph.D. students. Please feel free to forward this call to colleagues and students.

All participating reviewers who submit a report will become members of a "consortium co-authorship" for the final publication that reports the outcomes of the project. This consortium, the "Management Science Reproducibility Collaboration," will be listed as a co-author on the front page of the article, with all members listed by name and affiliation in the paper's appendix.

If you are willing to participate as a reviewer, we ask you to complete this short survey (before January 15, 2023), so we can match you with a paper from your field.

Begin Survey

In case of any questions, please contact the project team at ManSciReP@informs.org.

Sincerely,
David Simchi-Levi
Editor-in-Chief, *Management Science*

Invitation email to others

Dear Researcher:

We would like to draw your attention to an opportunity to join a new project on the reproducibility of studies published in Management Science as a reviewer.

In the Management Science Reproducibility Project (ManSciReP), we assess the computational reproducibility of studies published in the journal. Since 2020 the Code & Data Editor verifies that replication materials are provided but does not attempt reproduction itself. In this project, we aim to quantify the reproducibility of results published in Management Science articles before and after the new Data and Code Disclosure Policy came into effect.

If you would be willing to review, we would assign you a paper from your own field of research, and using software that you are familiar with. We would then ask you to report back within 4-6 weeks to what extent you were able to reproduce the paper's main results, and what the obstacles were.

This call for reviewers is open to any researcher in the community, including advanced PhD students. Please feel free to forward this call to colleagues and students.

All participating reviewers who submit a report will become members of a "consortium co-authorship" for the final publication that reports the outcomes of the project. This consortium, the "Management Science Reproducibility Collaboration", will be listed as a co-author on the front page of the article, with all members listed by name and affiliation in the paper's appendix.

If you are willing to participate as a reviewer, we ask you to complete this short survey, so we can match you with a paper from your field.
Survey link

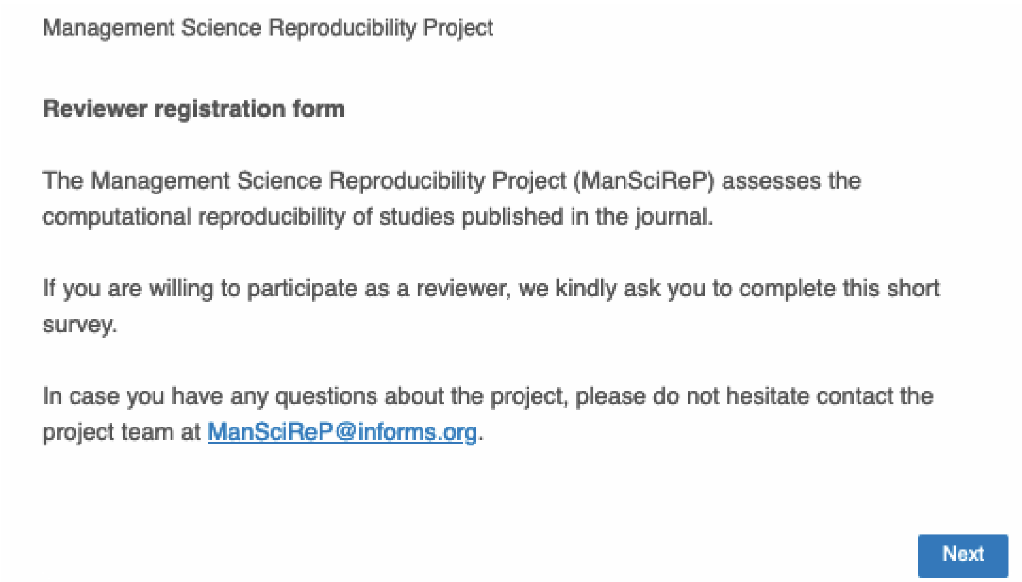
In case of any questions, please contact the project team at ManSciReP@informs.org.

Sincerely,

David Simchi-Levi
Editor-in-Chief, Management Science

Miloš Fišar, Ben Greiner, Christoph Huber, Elena Katok, and Ali Ozkes
Project coordinators

E.2 Reviewer registration survey



Management Science Reproducibility Project

Reviewer registration form

The Management Science Reproducibility Project (ManSciReP) assesses the computational reproducibility of studies published in the journal.

If you are willing to participate as a reviewer, we kindly ask you to complete this short survey.

In case you have any questions about the project, please do not hesitate contact the project team at ManSciReP@informs.org.

Next

Your full name:

Your email address:

Your affiliation:

Please do not use abbreviations. For multiple affiliations uses a semi-colon (;) to separate the affiliations.

Your current position:

- (●) Professor
- ( ) Associate Professor
- ( ) Assistant Professor
- ( ) PostDoc
- ( ) Other academic with PhD (e.g., lecturer)
- ( ) Ph.D. Candidate
- ( ) Professional with Ph.D.
- ( ) Other:

In what year did you receive your Ph.D.?

Back Next

At which departments of Management Science would you typically submit your research paper?

*Please **drag&drop** the respective departments to the box on the right, and rank them.*

Departments	My departments
Accounting	
Beh. Eco. & Decision Analysis	
Business Strategy	
Data Science	
Entrepr. and Innovation	
Finance	
Healthcare Mgmt.	
Information Systems	
Marketing	
Operations Management	
Optimization	
Organizations	
Revenue Mgmt. and Market Analytics	
Stochastic Models and Simulation	

Which programming language/analysis software/framework do you have access to and are comfortable with?

- [ ] C/C++
- [ ] Jupyter
- [ ] R
- [ ] Fortran
- [ ] Lingo
- [ ] SAS
- [ ] Gams
- [ ] Mathematica
- [ ] SPSS
- [ ] Gauss
- [ ] Matlab
- [ ] SQL
- [ ] Gurobi
- [ ] MS Office
- [ ] Stan
- [ ] Java
- [ ] Python
- [ ] Stata
- [ ] Julia

Which subscription databases do you have access to?

- [ ] Compustat
- [ ] U.S. Census Bureau
- [ ] CRSP
- [ ] WRDS
- [ ] Factset

Back Next

Management Science Reproducibility Project

Your expectations:

In your estimation, what proportion of Management Science papers **under the current Data & Code disclosure policy** (replication packages required and reviewed for completeness by Code and Data editor) can be **fully reproduced** with the available replication materials?

					%					
0	10	20	30	40	50	60	70	80	90	100

In your estimation, what proportion of Management Science papers **under the previous policy** (replication packages expected but not verified or reviewed) can be **fully reproduced** with the available replication materials?

					%					
0	10	20	30	40	50	60	70	80	90	100

Back Next

Management Science Reproducibility Project

We thank you for registering as a reviewer for the Management Science Reproducibility project.
Your registration has been recorded. We will be in touch in due course.

In case of any questions, please contact the project team at ManSciReP@informs.org.

E.3 Reproducibility report survey

Management Science Reproducibility Project

Welcome to the report survey for the **Management Science Reproducibility Project**.

Here we ask you about your attempt to reproduce the results of your assigned Management Science article.

Before you start completing this report survey, **please familiarize yourself with our guidelines for reviewers**.

Please enter your email address:

Please enter the DOI of the article (10.1287/mnsc.XXXX.XXXX) that you reviewed:

Please enter the title of the article:

If there was a second person that significantly contributed to this review and should be given credit, please list the name, email address, and affiliation.

What is your overall assessment of the reproducibility of this article's main results (tables, figures, other results in the main manuscript)?

- ◯ Fully reproduced.
- ◯ Largely reproduced, with minor issues.
- ◯ Largely not reproduced, with major issues.
- ◯ Not reproduced.
- ◯ Not reproduced but consistent with log files.
- ◯ Not based on any data analysis, simulation, or code.

Next

Management Science Reproducibility Project

The package includes a README file:

- ◉ Yes
- ◯ No

Was the README file sufficiently helpful to facilitate the reproduction?

- ◯ Yes
- ◯ No

Any comment on the README file?

Back

Next

Management Science Reproducibility Project

Does the replication package already include all the necessary DATA to reproduce the results reported in the main manuscript?

- ◯ Yes
- ◯ No, the analysis does not need data.
- ◯ No, the package includes only partial data.
- ◯ No, the package includes only sample or synthetic data.
- ◉ No, the package includes no data at all.

The missing data ...

- ◯ Can be obtained for free from publicly available sources.
- ◯ Can be obtained from a commercial provider against a one-time fee or for a subscription fee.
- ◯ Can be obtained in a different way (e.g., upon request to the data owner (not authors!), etc.).
- ◯ Cannot be obtained.

Please list the data sources used in the study. (E.g., "lab experiment", "own survey with representative panel", "Comstat, CRSP", ...)

Any other comments on data availability?

Were you able to obtain all data needed to attempt a reproduction of all results?

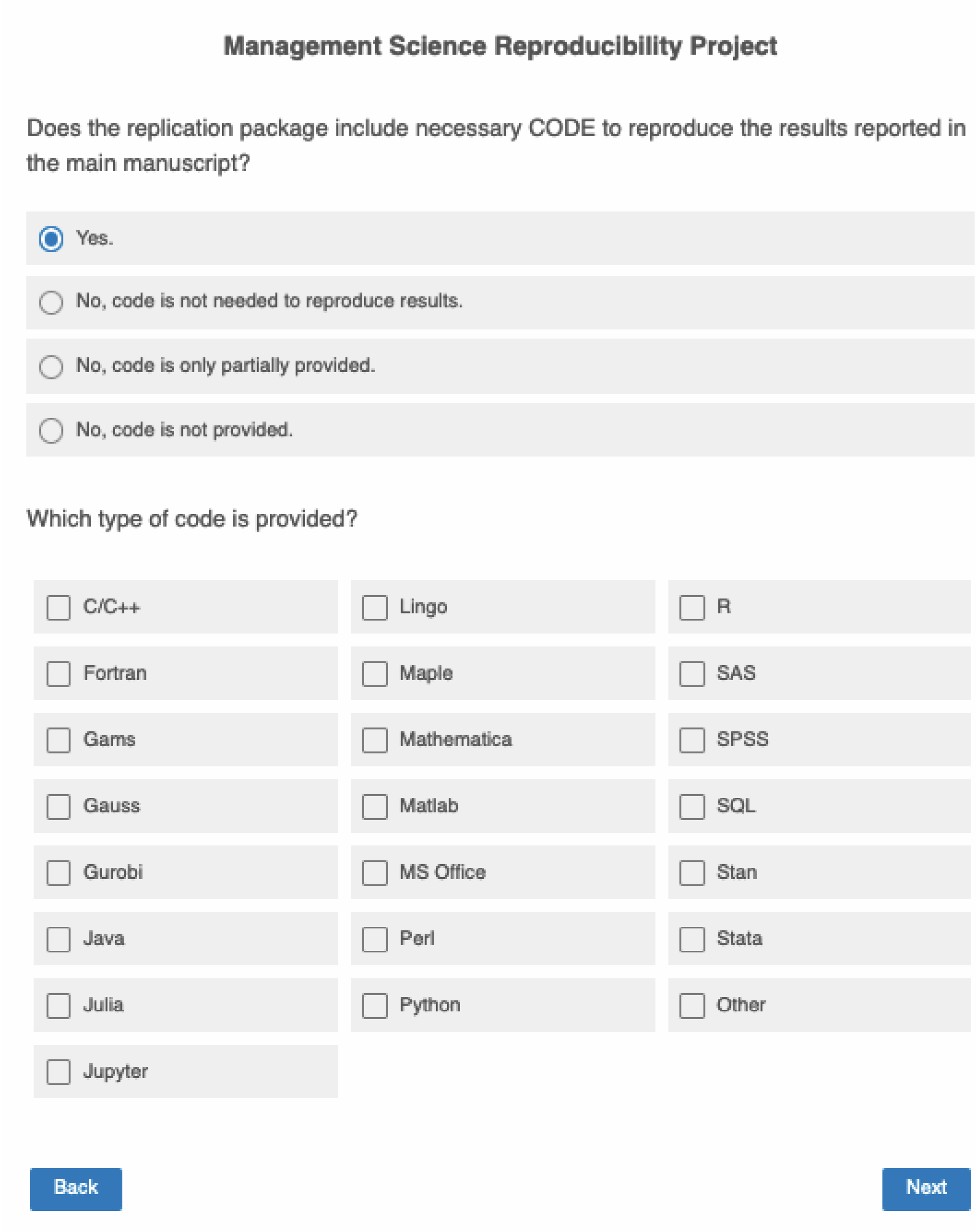
- ◯ Yes
- ◯ No

If applicable, can you please explain any obstacles you had to overcome, or obstacles you could not overcome, in obtaining a complete dataset for review?

Are log files provided from the authors' own running of the code on the original data, such that one can still compare results reported in the paper with the log file in case data cannot be obtained and/or the result cannot be reproduced?

- ◯ Yes, log files are provided for all results.
- ◯ Log files are provided for some results, but not for others.
- ◯ No, log files are not provided within the replication package.

Back | Next



Management Science Reproducibility Project

Does the replication package include necessary CODE to reproduce the results reported in the main manuscript?

- (•) Yes.
- ( ) No, code is not needed to reproduce results.
- ( ) No, code is only partially provided.
- ( ) No, code is not provided.

Which type of code is provided?

☐ C/C++	☐ Lingo	☐ R
☐ Fortran	☐ Maple	☐ SAS
☐ Gams	☐ Mathematica	☐ SPSS
☐ Gauss	☐ Matlab	☐ SQL
☐ Gurobi	☐ MS Office	☐ Stan
☐ Java	☐ Perl	☐ Stata
☐ Julia	☐ Python	☐ Other
☐ Jupyter		

Back Next

Management Science Reproducibility Project

How many tables does the main manuscript contain overall?

3

How many figures does the main manuscript contain overall?

3

Back Next

Management Science Reproducibility Project

For each TABLE in the paper, please indicate whether it is a results table (that should be reproducible), whether you were able to reproduce it, and provide any details/comments on obstacles/issues.

	Reproducible?	Can you provide any comments/details?
Table 1		
Table 2		
Table 3		
		any comments/details?

Any further comments on the reproduction of tables?

Back Next

Management Science Reproducibility Project

For each FIGURE in the paper, please indicate whether it is a results table (that should be reproducible), whether you were able to reproduce it, and provide any details/comments on obstacles/issues.

	Reproducible?	Can you provide any comments/details?
Figure 1		
Figure 2		
Figure 3		
		any comments/details?

Any further comments on the reproduction of figures?

Back Next

Management Science Reproducibility Project

How many OTHER RESULTS reported in the text of the main manuscript (e.g., p-values from statistical tests not yet reported in the tables / figures) did you identify and attempt to reproduce?

How many of these results were you able to fully reproduce?

Any comments / details on the reproduction of other results reported only in the text?

Back Next

Management Science Reproducibility Project

Please upload **one single file** (pdf, zip, etc.) that contains the log files / screenshots / outputs from your analysis that you used to check the tables and figures of the manuscript.

Drop files or click here to upload

Back

Next

Management Science Reproducibility Project

When attempting reproduction of this paper's results, did you have to change/fix any CODE (other than changing the working directory, etc.)?

- (•) Yes
- ( ) No

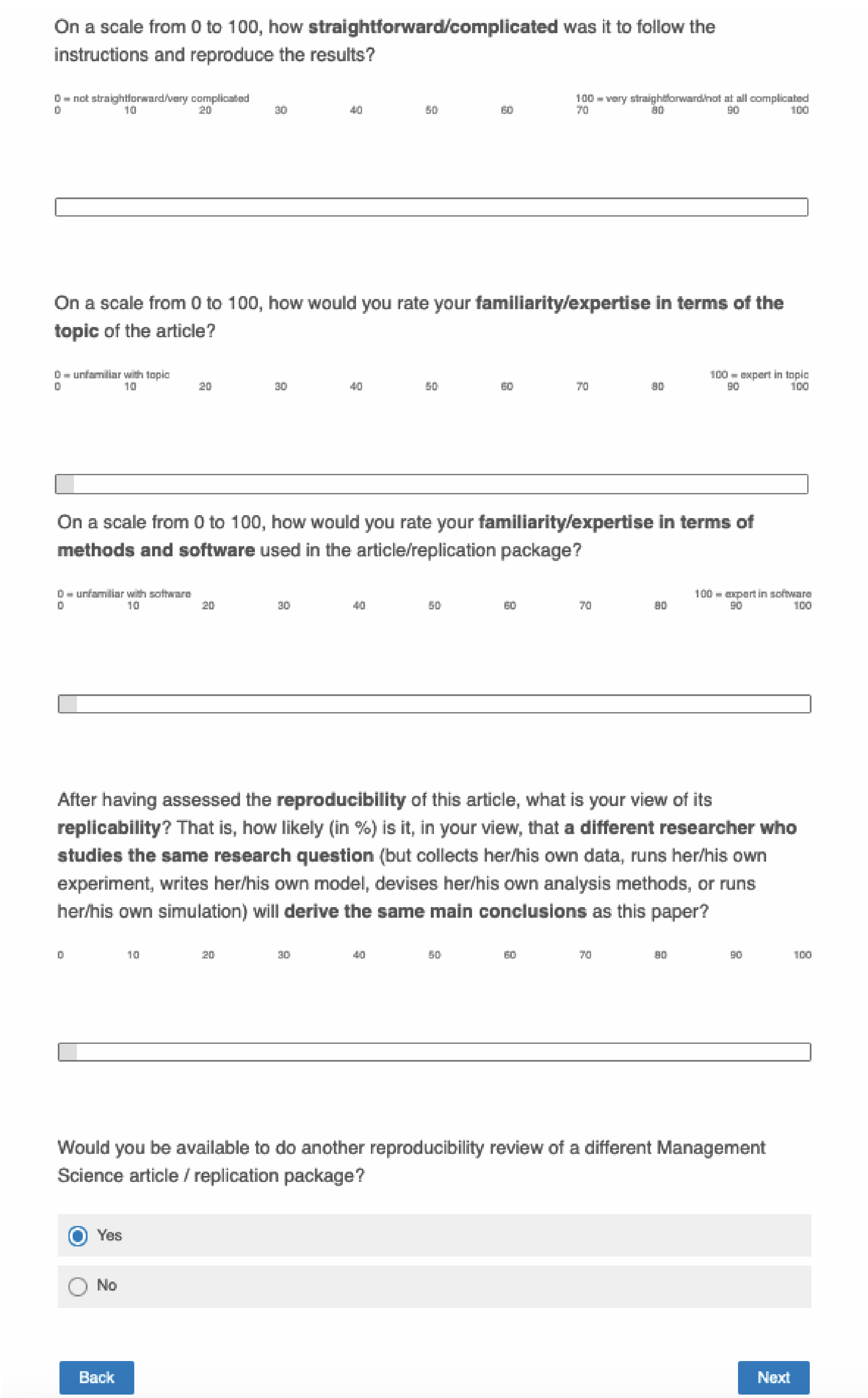
Any comments / details on type and extent of code changes?

When attempting reproduction of this paper's results, did you have to change / fix / transform any DATASETS?

- (•) Yes
- ( ) No

Any comments / details on type and extent of dataset changes?

Approximately, how much time (in hours) did you devote to the reproduction of this paper?



On a scale from 0 to 100, how **straightforward/complicated** was it to follow the instructions and reproduce the results?

0 = not straightforward/very complicated — 100 = very straightforward/not at all complicated

0 10 20 30 40 50 60 70 80 90 100

On a scale from 0 to 100, how would you rate your **familiarity/expertise in terms of the topic** of the article?

0 = unfamiliar with topic — 100 = expert in topic

0 10 20 30 40 50 60 70 80 90 100

On a scale from 0 to 100, how would you rate your **familiarity/expertise in terms of methods and software** used in the article/replication package?

0 = unfamiliar with software — 100 = expert in software

0 10 20 30 40 50 60 70 80 90 100

After having assessed the **reproducibility** of this article, what is your view of its **replicability**? That is, how likely (in %) is it, in your view, that **a different researcher who studies the same research question** (but collects her/his own data, runs her/his own experiment, writes her/his own model, devises her/his own analysis methods, or runs her/his own simulation) will **derive the same main conclusions** as this paper?

0 10 20 30 40 50 60 70 80 90 100

Would you be available to do another reproducibility review of a different Management Science article / replication package?

- Yes
- No

Back — Next

Management Science Reproducibility Project

This concludes the report survey. Thank you so much for your efforts.

When you click the "submit" button below, the report will be submitted and you will not be able to go back and make any changes.

Back Submit

E.4 Reviewer guidelines

Management Science Reproducibility Project

Reviewer Guidelines

Scope

We ask you to attempt to reproduce the results in the main manuscript of the paper. Results include tables and figures that are based on data or code, as well as results only reported verbally in the text (e.g., statistical test results not reported in tables and figures). You can ignore results reported in the appendix or in footnotes. Note that this assessment is purely about reproducibility, not about the appropriateness, soundness, or robustness of applied methods.

Some packages, in particular older ones submitted before the new code and data disclosure policy took effect, may not include data or code, or provide only limited documentation. In any case, please make an honest attempt to reproduce the results based on the information provided in the paper, appendix, and replication package. Report any barriers to reproduce the results in the final report survey.

If reproduction is not possible, some reviews may be completed very quickly. In these cases you can indicate your availability to review another article / replication package in the report survey, and we will be happy to assign you another one.

Anonymity

Please do not communicate with authors directly. We want to keep strict reviewer anonymity. The goal of this reproducibility project is to establish how many articles can be reproduced based *only* on the information provided in the paper, the appendix, and the replication package, i.e., *without* having to contact the authors in the process.

Conflicts of interest

Please apply the same ethical standards to this review as you would to a regular manuscript review at Management Science. In particular, there is a conflict of interest if one of the authors is/was your advisor or student, works at the same institution as you, is/was a co-author during the last 5 years, or if you have otherwise an interest in the outcome of the reproduction attempt. Please report any conflict of interest to us, and we will assign you to a different article/replication package.

Documentation

Please document your reproduction attempts. You can either produce log files that show your output, or make screenshots, or use any other method of documentation. In the report survey you will be asked to upload a zip file of your documentation.

The Report Survey

A full printout of the report survey is included at the end of this document. A personalized link to the survey is provided in your assignment email.

Paper/reviewer details: The first part of the survey just asks to identify yourself and the article/replication package you reviewed.

Overall assessment: We then ask for your overall assessment of the reproducibility of the whole article. Similar to the table-by-table, figure-by-figure results below, we ask you to select one of six possible assessment outcomes.

- "Fully reproduced" means that the output of your analysis shows the exact same results as reported in the paper, for all results reported in the main manuscript. You can ignore non-essential issues such as colors/line types in figures or similar.
- "Largely reproduced, with minor issues" means that there may be minor differences in your output compared to the results in the paper, but the paper's conclusions and learnings stay the same.
- "Largely not reproduced, with major issues" means that there are major differences in your output compared to the results in the paper (because you get different numbers or you are unable to reproduce the results because of missing data etc.), such that the reproduction results could not be used to support the conclusions of the paper.
- "Not reproduced" means that the results from the reproduction cannot support the conclusions drawn in the paper, either because the output is different, or because the results cannot be produced at all because of missing data or non-recoverable code.
- "Not reproduced but consistent with log files" means that you cannot reproduce the results based on running code on data, but that log files are included in the replication package, and the log files are fully consistent with the results reported in the paper.
- "Not based on any data analysis, simulation, or code" means that the paper does not include any analysis that would fall under the Code and Data Disclosure policy, i.e., analysis that is based on data, and does not use simulations or other code based-analysis. This typically only applies to pure theory papers.

Package documentation: The next part asks about the quality of documentation in the replication package, i.e., whether a README file is provided and whether it was sufficiently helpful in your reproduction attempt.

Data: The next part asks about the amount and quality of data included in the replication package, i.e., whether data, partial data, synthetic data or sample data is included or not, whether you could obtain non-included data from publicly available, private, or subscription sources, which data sources the study is based on, and whether in the end you had sufficient data to continue with the reproduction. It also asks whether log files are provided in the replication package.

Code: The next part asks whether code was included in the replication package and which type of code.

Tables/Figures: We then turn to the individual tables and figures in the main manuscript. First, we ask how many tables and figures there are overall in the manuscript, such that subsequently we can ask you for each single one of them, first for all tables, then for all figures. Please ignore tables and figures in the appendix.

You will see a table with one row per table in the manuscript. For each manuscript table, we ask via a dropdown field whether the manuscript table could be reproduced (fully, largely, largely not, not), whether there are log files consistent with the table, or whether the manuscript table was not based on data/analysis (e.g., a list of conditions, experimental design), and for details or comments.

In the dropdown field,

- "Fully reproducible" means all numbers / all output is the same in your output as reported in the paper (ignoring non-essential differences like color or line type in figures).
- "Largely reproducible, with minor issues" means that there may be small quantitative differences in reported numbers / output (e.g., due to rounding errors, different software versions, different random seeds, typos) but the qualitative conclusions and learnings from the table/figure stay the same.
- "Largely not reproducible, with major issues" means that there are significant quantitative differences in reported numbers / output such that different qualitative conclusions and learnings would be drawn, or that important parts of the table/figure cannot be produced at all. For example, while some models in a regression table can be reproduced, others yield completely different numbers.
- "Not reproducible" means that the results from the reproduction cannot support the conclusions drawn in the paper from the table/figure, either because the output is different, or because the table/figure/result cannot be produced at all because of missing data or non-recoverable code.
- "Not reproducible but consistent with provided log file" means that you cannot reproduce the results based on running code on data, but that log files are included in the replication package, and the log files are fully consistent with the results reported in the paper.
- "Table/Figure not based on data/analysis" means that this table or figure is not based on results from analyzing data or otherwise running code, such that they do not need to be documented. Examples include tables outlining experimental designs, showing a timeline of events, or listing variables, or figures providing screenshots or illustrations, or visualizing a conceptual model.

In the comments, please provide a short description of details in case you were not able to fully reproduce some results, e.g., denoting the column or cells where differences appear, or commenting which errors in the code prevent you from running a model, etc.

After tables, we ask about figures. As for manuscript tables, you will see a table with one row per manuscript figure, and for each figure, we ask via a dropdown field whether the figure could be reproduced (fully, largely, largely not, not), whether there are log files consistent with the figure, or whether the figure was not based on data/analysis (e.g., an illustration or picture). Please use the comment field to provide details on reproduction issues.

Other results: Next we ask about other results reported in the text of the main manuscript, e.g., p-values from statistical tests not yet reported in the tables/figures. For these results, we only ask for a summary report: how many results you identified, and how many you could reproduce. You can ignore results reported in the appendix or in footnotes.

Review documentation: After having reported your reproduction results, we ask you to upload log files, screenshots, or output files that you compared to the results reported in the paper. Please include all logs/screenshots in one single file (pdf, zip, etc.).

Review experience: The last part of the survey asks about your experience when reviewing the replication package. Namely, we would like to know if you needed to fix/change any code or datasets in order to be able to run the reproduction, how much time you invested, how complicated/straightforward the reproduction was, and how you assess your own expertise in terms of the article's topic and the applied methods/software. We also ask for your view on the replicability (as opposed to reproducibility) of the article.

Review availability: The final question asks whether you would be available to do another reproducibility review of a different article/replication package.