J Econ Sci Assoc (2015) 1:127–131
DOI 10.1007/s40881-015-0018-y

EDITORIAL

Editors' preface: statistics, replications and null results

Nikos Nikiforakis[1] · **Robert Slonim**[2]

© Economic Science Association 2015

1 Introduction

This issue marks the first anniversary of the *Journal of the Economic Science Association* (JESA). It is therefore an appropriate time to share some summary statistics and information about replication studies and submissions with null results.

2 Summary statistics

By the end of September 2015, one year after starting to accept submissions, JESA had received a total of 95 new submissions. These numbers reflect the increasing use and importance of experiments in economic research (Nikiforakis and Slonim 2015; Fig. 3).[1] For comparison, *Experimental Economics* (EXEC) exceeded 100

[1] At the same time, the number of papers submitted to *Experimental Economics* (EXEC)—the companion journal to JESA—had almost the identical number of submissions as in each of the previous 2 years, suggesting that the submissions to JESA represent a net increase in submissions to the journals of the Economic Science Association.

✉ Nikos Nikiforakis
nikos.nikiforakis@nyu.edu

[1] New York University, Abu Dhabi, UAE

[2] University of Sydney, Sydney, Australia

https://doi.org/10.1007/s40881-015-0018-y Published online by Cambridge University Press

submissions per annum for the first time in 2008 (Nikiforakis and Slonim, 2015; Fig. 1), exactly 10 years after its inaugural issue.[2]

The average decision times and the acceptance rate were low during the first year. The average time until a *first* decision for all new submissions (i.e., accept, reject, revise and resubmit) was 34 days. The average time for all papers with a *final* decision has been 57 days. The acceptance rate for papers with final decisions was 20.3 %, with 14 of the 69 papers with final decisions accepted by the end of September. This rate is again similar to the acceptance rate at *Experimental Economics* in 2008 (Nikiforakis and Slonim 2015; Fig. 1). Considering only papers that were accepted for publication, the average time from original submission to final acceptance has been only 162 days (i.e., just over 5 months).

Of course, the most important metric for judging the quality of a journal is the quality of the papers it publishes. We have been fortunate to receive many excellent submissions in the first year. We have also been extremely fortunate to have the generous support of many anonymous reviewers in evaluating and improving the submissions received. If our experience has convinced us of one thing, it is that experimental economists do not free ride.

3 Replications and null results

One of the Economic Science Association's aims for JESA is to provide an outlet for publishing useful yet underrepresented papers for experimental economic research, including papers addressing experimental methodology issues, reporting replications of previous findings and null results. In the past year, we noticed a good deal of variation in replication and null result submissions to JESA, reflecting some uncertainty and variation in beliefs across authors regarding how to report these results. This feeling is shared by the editors of JESA and EXEC. We therefore jointly concluded (i.e., the editors of the JESA, EXEC and also the new President of the ESA—Yan Chen) that some comments on reporting replication and null results would be helpful. We have no intention to offer a comprehensive coverage on these vast topics; instead, our goal is to provide some practical guidance on how to best prepare these submissions.

Replications have implicitly always been part of experimental economics. Typically, when a study is published using a novel design and presenting an interesting result, follow-up studies would start their investigation by replicating the original setup in a control treatment before introducing novel treatments to address new questions. For instance, the experimental economics literature includes many studies replicating the efficiency of the double auction mechanism, cooperation unravelling in finitely repeated voluntary contribution mechanisms, and the importance of fairness concerns in ultimatum, trust, dictator and gift exchange games. It should be apparent that we commonly observe *successful* replications in

[2] We greatly appreciate the wonderful collaboration with the current editors of EXEC—David Cooper and Charles Noussair—who have repeatedly helped identify papers that may be better fits for JESA than for EXEC.

https://doi.org/10.1007/s40881-015-0018-y Published online by Cambridge University Press

control treatments of follow-up investigations. What has been far less common is the publication of (1) *failed* replications and (2) *stand-alone* replications.

There are several reasons for the dearth of *failed* replication publications. The one of interest for our purposes is that it can be difficult to identify the reason(s) a certain finding failed to replicate. To understand why, assume a new experimental study is published finding result X in environment Y. The environment here is broadly defined to include at least the experimental treatments, instructions, laboratory and recruitment procedures, and subject pool. Failure to replicate finding X could be because the original study committed a Type-I error (rejecting a true null hypothesis), the replicating study altered something in the environment which was crucial for result X,[3] or that the replicating study committed a Type-II error (failing to reject a false null hypothesis).

A failed replication is most useful when the reader can understand the likely reasons for the different results. The likelihood of a Type-I and Type-II errors depends primarily on the size of the effect, the amount of noise in the data and the sample size. Statistical power increases as the sample size and the treatment effect increase, and as the noise in the data generating process decreases. For example, if the sample size is small relative to the noise in the data generating process, then replication tests will be underpowered making a Type-II error more likely. The conventional power for an acceptable test is 80 % meaning that if the null is false, then the researchers would reject it 80 % of the time. In order for a failed replication to be informative, it is therefore necessary that the study has sufficient power to reject the null hypothesis when it is indeed false. That is, failure to replicate with an underpowered study is clearly not very informative.

There may be several reasons for the scarcity of *stand-alone* replication studies being published (either failed or successful), although they are more difficult to pin down.[4] Nonetheless, we hope that the Aims and Scope as well as this preface make it clear that JESA encourages stand-alone replication submissions and that the ESA encourages authors to cite these replication studies. Of course, authors need to motivate why their paper is interesting enough to warrant publication.

There are also several reasons for the relatively small number of papers reporting *null results* getting published. The one of interest for our purposes is again, like with failed replications, that it can be difficult to identify the reason(s) for the null effect. Failure to find a significant effect could be because the hypothesis being tested is false (and therefore does not hold in the environment studied), or that the study committed a Type-II error (failing to reject a false null hypothesis) because of insufficient power.

[3] If result X does not hold in the replicating environment because some things in the replicating and original environments are different, then we may interpret the outcome of the "replicating" study as providing information about the boundary conditions for the generalizability of X rather than necessarily concluding it was a failed replication.

[4] For instance, as mentioned, the convention to replicate a past result in the baseline condition of a study means many replications occur already. Further, incentives to conduct stand-alone replications may be insufficient if researchers believe journals are reluctant to publish them or they would not be cited well.

https://doi.org/10.1007/s40881-015-0018-y Published online by Cambridge University Press

4 Reporting power calculations for replication and null-effect studies

The editors of JESA and EXEC strongly encourage authors of replication studies (failed or successful) and studies reporting null results to present power calculations following the guidelines below.[5]

4.1 Replications

For replication studies, successful or failed, the information regarding the size of the treatment effect and (if applicable) standard deviations *must come from the original study*, whereas the number of observations must come from the replication study. For an example of how to report the power for a successful replication see Kingsley (2015) in this issue. In addition, all failed replication studies should ideally have a sample size that provides *at least* 80 % power to detect the original study's effect size at the 5 % significance level; with less than 80 % power, a failed replication may be regarded as being insufficiently powered to clearly identify the reason for the failed replication.

4.2 Null result

For null results, authors should specify the treatment effect size that they could have detected with 80 % power at a 5 % level of significance *using their sample size and data*. Examples of studies reporting power calculations for null results are included in the current issue (Drichoutis et al. 2015; Kingsley 2015) and the first issue (van de Ven and Villeval 2015). Note that appropriate adjustments need to be made when the data include observations which are not independent; for example and references see the appendix of Drichoutis et al. (2015).

5 Concluding remarks

By starting JESA, the ESA wishes to encourage the timely replication of important new findings and the publication of null results. The decision whether to publish such studies will depend on the appropriateness of the experimental design, the power to detect significant effects and its added value; i.e., simply finding a null result or replicating a study does not automatically guarantee publication at JESA. Authors need to motivate why their paper is interesting enough to warrant publication. In any case, the decision to publish a replication will not depend on the

[5] Software packages such as STATA include power calculation tools that can facilitate these computations. If appropriate, authors simply have to input the required information to obtain the power of the test. In some cases, however, where the appropriate statistical test necessitates more involved econometric issues (e.g., estimating Tobit or discrete-choice models, inclusion of control variables, panel data and non-independent observations), we are unaware of any pre-programmed software to provide power calculations. In these cases, authors may need to use simulation techniques to appropriately determine power.

https://doi.org/10.1007/s40881-015-0018-y Published online by Cambridge University Press

findings of the study. Our belief as editors is that economic experiments need replication to better inform theory and policy.

References

Drichoutis, A. C., Lusk, J. L., Nayga, R. M., Jr. (2015). The veil of experimental currency units in second price auctions. *Journal of the Economic Science Association*. doi:10.1007/s40881-015-0014-2.

Kingsley, D. C. (2015). Peer punishment across payoff equivalent public good and common pool resource experiments. *Journal of the Economic Science Association*. doi:10.1007/s40881-015-0017-z.

Nikiforakis, N., & Slonim, R. (2015). Editor's preface: introducing JESA. *Journal of the Economic Science Association, 1*(1), 1–7.

van de Ven, J., & Villeval, M. C. (2015). Dishonesty under scrutiny. *Journal of the Economic Science Association, 1*(1), 86–99.

https://doi.org/10.1007/s40881-015-0018-y Published online by Cambridge University Press