EXPERIMENTAL ECONOMICS: DESIRABLE AGENDA

Rakesh Kumar Sudan

**Institute of Integrated and Honors Studies, Kurukshetra University, Kurukshetra, Haryana*

The experimental science community of social sciences finds itself at cross roads of 'credibility inflexion point' in matter of experimental research. Growning concerns over the replicability of published results in fields, including psychology (Nelson 2018) and economics (Camerer 2016), have inspired researchers to reconsider the way they formulate hypotheses, collect and analyse data, and interpret their findings (see, for example, Ioannidis 2005, Simmons 2011, Maniadis 2014, Brodeur 2016, and Munafó 2017). Experiments in constrained settings take away more or less the realism of the observations. Quality of existing data, data gaps is also an issue. Thus experimental necessity of social sciences becomes a buzzword in the present advanced technology of data analysis. However certain issues like replicability of the experiments hang around in the experimental domain related to economics.

However, it's not just replicability issue that lies at the core. While it is of first-order importance to embrace research practices that increase reproducibility, experimental economists also must tackle the generalizability and applicability of the evidence they produce (Banerjee 2017). For example, permanent in come is an important explanatory variable in Friedman's consumption-income relationship. Many ways can guide us to an estimate of this entity. However, one which gives the best approximation needs to be explored. Same can be said about the factors determining the shape of Phillips curve After all, we are not only interested in ensuring that the same experiment yields the same outcome when it is repeated. Ideally, we would like to be able to generalise our findings to different contexts, and produce insights that contribute to economic the oryand policy discussions.

These considerations have given rise a constructive debate that is pinned on the credibility, generalisability, and relevance of findings in experimental economics. A number of recommendations are put up here that summarise research practices that we as researchers should all domore of (Czibor2019). This is intended as an aidmemoire or a ready reck on erfor researchers in the design phase of their experiment.

While the guidelines relate to experimental economics in original, it is believed many of the suggestions are relevant in other fields of experimental science, and for scholars pursuing estimate susing observational data. For example how effective are exit polls in deducing about voters' behaviour, or how much child abuse stays as a dominant variable in explaining juvenile delinquencyata later stage

It is desirable to cover four threats to generalizability: the representativeness of population, non-random selection in to the experiment, treatment non-compliance, and characteristics of the experiment that may affect behaviour. It is important that we do not merely acknowledge these issues ex post as potential limitations of our study. They must guide decisions in the design phase that affect the type of data we generate.

What does that mean in practice?

There is good point in advocating for conducting more natural field experiments (NFEs). They are covert, and so can mitigate biases a rising from self-selection in to the experiment and experimenter demand effects, and typically involve the population of interest. NFEs offer aunique combination of control and realism. Also allied with NFEs is the ethical aspect. For example, households' economic behavior under conditions of sensitive issues like same-sex marriages, live-in relationships etc. pose challenges of such kind.

It can also be argued that lab and field experiments, as well as naturally occurring data, are complements in the production of scientific knowledge. For instance, we can begin by documenting an effect among students in the lab, then test its generalisability by repeating the experiment with different tasks and populations. Alternatively, we could first evaluate a program in a field experiment, then use additional lab experiments to test mechanisms that may explain what we observed in the field (Stoop2012).

The second challenge is the informativeness of our findings–in other words, how to design experiments that optimise learning for the students and understanding of results at common place. This requires a critical look at the practice of basing our inference solelyon p-values–an approach that leads to a high false-positive rate (Ioannidis 2005) and ignores the economic significance of findings (Ziliak and McCloskey 2004). This practice is especially dangerous in combination with specification searching (Simmons2011, Brodeur 2016) and multiple hypothesis testing (List 2016). It would be aptto discuss two ways of dealing with this issue.

lt would be worth while to advocate for more replication studies to increase the credibility of research findings. In Czibor (2019), we must apply our minds to devise ways to incentivise such studies. The incentives maybe in terms of fellowships, best research award or students' aid programmes.

Due attention ought to be given to the importance of statistical power in determining the informative value of experiments, and present ways to increase power for a given experimental budget, such as using within-subject designs (where the same participant is exposed to different treatments, with their order randomised) when appropriate, and collecting baseline characteristics to perform blocked randomisation (partitioning our sample to subgroups along relevant variables, and randomizing with in these groups).

Finally, it would be appropriate to highlight issues related to the policy relevance of experimental economics results. Even perfectly credible and reproducible findings may not inform policy discussions if they exclusively focus on short-term impacts, leave mechanism sun covered, and fail to consider scalability. Observing the longer-term outcomes ensures that promising results don’t pass away quickly, but is also important because it may take time for important general equilibrium effects to emerge due to the existence of inside and outside policy lags. For example, may be an intervention led to short-term improvements for the participants, but in the long run transformed the market in ways that were harmful for everyone.

Experiments that document a resultant effect but do not explicitly study the under lying mechanisms leave a lot of potential gains on the table. Theory can help us address the general is ability threats discussed above by explicitly model ling the participation and compliance decision. Specifying a model, then using experimental variation to identify its deep structural parameters, allows us extrapolate our results to different contexts and interventions. We can also design experiments to test the predictions of a theory, ortorunatie-breaker between competing models. If we can understand why we observe a phenomenon, we are in a better position to advise policy.

Theory also becomes important when we consider the 'science of scaling': a system atic treatment of the issues that arise when a small-scale, short-run programme is rolled out on a much larger scale. As Banerjee (2017) demonstrate, scaling up a programme is not a trivial undertaking. It requires careful planning and testing in the design phase to avoida 'voltage drop' (Al-Ubaydli2017), where by the scaled-up programme’s effect is smaller than the original, small-scale evaluation.

These considerations include general equilibrium effects (including the reaction of politicians to programs), and potential biases stemming from sample selection (the original pilot trial often includes a 'convenience' rather than are presentative sample), site selection (researchers might choose to run their experiments in places where they are 'easier' to implement), and piloting (including the fact that as a program grows larger, it needs to recruit additional workers, who may be less skilled or motivated than the ones already hired; see Davis (2017). Researchers need to 'backward induct', to address scaling-related challenges in the design phase. If they do not, they may not create programmes that work well at scale.

Grounds for optimism exist already as the researchers are becoming increasingly aware of these challenges, and are taking important steps to improve the quality of their research. The profession as a whole is raising up to new standards of evidence (an example being the success of the preregistration movement). More academic work is needed for contributing to this positive change, to prompt other experimental researchers to join the debate, to share the challenges they have identified and their proposed suggestions to overcoming them, so that step-by-step we can work together to improve the quality of scientific research.

REFERENCES

1. Nelson, L D, J Simmons, and U Simonsohn (2018), "Psychology’s renaissance", Annual Review of Psychology, 69, 511–534.
2. Al-Ubaydli, O, J A List, D LoRe, and D Suskind (2017), "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature", Journal of Economic Perspectives,31 (4):125–144.
3. Camerer, C F, A Dreber, E Forsell, T Ho, J Huber, M Kirchler, M Johannesson, M Kirchler, J Almenberg, A Altmejd, T Chan, E Heikensten, F Holzmeister, T Imai, S Isaksson, G Nave, T Pfeiffer, M Razen, H Wu (2016) "Evaluating replicability of laboratory experiments ineconomics",Science351(6280):1433–1436.
4. Ioannidis, J P A (2005), "Why most published research findings are false", P LoS Medicine 2(8):0696–0701.
5. Banerjee, A, R Banerji, J Berry, E Duflo, H Kannan, S Mukherji, M Shot land, and M Walton(2017), "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application", Journal of Economic Perspectives 31(4):73–102.
6. Simmons, JP, L D Nelson, and U Simonsohn (2011), "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant",

7. Psychological Science22 (11):1359–1366.

8. Maniadis, Z, F Tufano, and J A List (2014), "One swallow does n't make a summer: New evidence on anchoring effects", The American Economic Review 104 (1):277–290.

9. Brodeur, A, M Lé, M Sangnier, and Y Zylberberg (2016), "Star Wars: The empirics strike back", American Economic Journal: Applied Economics, 8(1):1–32.

10. Munafó, M R, BANosek, D V M Bishop, K S Button, C D Chambers, N Percie, U Simonsohn, and E-J Wagenmakers (2017), "A manifesto for reproducible science", Nature Human Behaviour1:1–9.

11. Czibor, E, D Jimenez-Gomez, and J A List (2019), "The Dozen Things Experimental Economists Should Do (Moreof)", NBER working paper 25451.

12. Stoop, J, C N Noussair, and D Van Soest (2012), "From the lab to the field: Cooperation among fishermen", Journal of Political Economy 120(6):1027–1056.

13. Stoop, J, C N Noussair, and D Van Soest (2012), "From the lab to the field: Cooperation among fishermen", Journal of Political Economy 120(6):1027–1056.

14. List, J A, AM Shaikh, and YXu (2016),"Multiple Hypothesis Testing in Experimental Economics", NBER working paper21875.