Referee report on DP No. 2019-5: "Takeaways from the Special Issue on The Practice of Replication"

As its title suggests, this paper provides a synthesizing overview of the contributions to the special issue on "The Practice of Replication" from the perspective of the editor who proposed the special issue and designed the 'rules of the game' for submissions. From the outset, the emphasis was on setting out a replication plan, rather than carrying out an actual replication in detail. This encouraged attention to be focused on the core principles and essential elements of what constitutes a replication and how to interpret the results.

The paper starts by summarising the motivation for the exercise and the aspects that the replication plans were expected to address. It then classifies the eight journal articles in the special issue using the author's own typology (Reed, 2018), which distinguishes between six different categories of replications. This typology is one of several available in the literature; these typologies vary in their degree of detail and the use of terminology across different typologies is frustratingly inconsistent. However, the author's chosen framework, capturing two main dimensions (methods of measurement and/or analysis, and the source of the data) is sensible and more comprehensive than many others available. Most of the studies end up being classified, in the data source dimension, as using the same dataset as in the original studies, which is not surprising given that contributors were encouraged to select and justify the choice of a specific 'candidate' study. The paper then provides a useful summary of the main contribution of each paper, as would be expected in an overview of a special issue of this type.

The concluding section is the most interesting as it summarises the editor's 'takeaway' messages from the various contributions. These are well-judged and suggest, perhaps not surprisingly, that there is no "one-size-fits-all" characterization of how to undertake a replication. For example, it is argued that there is "no single procedure for doing replications … [and] no single measure of success". Contributors were encouraged to identify criteria to determine if a replication 'confirmed' or 'disconfirmed' the selected original study but, arguably, because of the many potential dimensions involved in a replication analysis, it is better to steer clear of any binary 'success'/'failure' classification of outcomes.

One of the editor's takeaway messages is that "researchers should formulate a 'pre-analysis' plan and make it publicly available before doing their replication". This is becoming much more common for experimental studies as a way of addressing "researcher degrees of freedom/forking paths" problems. Although empirical studies using observational data still dominate in economics, there is no reason why preregistration cannot be adopted for observational data (see, for example, Lei, Gelmand and Gitza, 2017). The main downside is the higher time cost involved, which may regrettably further discourage researchers from undertaking replications, when we really need more such work.

Minor points

Abstract, line 1: insert "the journal" or its full title after 'In July 2017,"

Additional reference

Lei, R., Gelman, A., & Ghitza Y. (2017) The 2008 Election: A Preregistered Replication Analysis, *Statistics and Public Policy*, 4:1, 1-8, DOI:10.1080/2330443X.2016.1277966