

Responses to Referee Report 1

Thank you for your comments and suggestions, all of which will be acted on to improve the clarity and flow of the arguments in the paper. Responses are given below (following the original comments and suggestions in italics).

(1) On page 2, par 2, "In principle...mechanism for highlighting unreliable results in the literature." The primary purpose of replication is to confirm or deny, not just deny.

I agree. The aim of the original wording was to link back to the first paragraph relating to reproducibility problems, but it is more important to provide an even-handed description of replication, so the wording will be changed.

(2) Page 3, first par, bottom: "From the perspective of statistical, a replication that faithfully reproduces..." Inasmuch as on page 2 you use the phrase "reproducibility crisis", you should distinguish between reproduction and replication. While you give various definitions on page 2, you should clearly define what you mean by these terms for the purpose of your paper. Having said the above, you should remark early on that reproduction is NOT replication, but it is where replication begins. If a paper cannot be reproduced, then there is no point in trying to replicate it.

A definition of 'reproduction'/'verification' will be added and a distinction drawn between reproduction and replication, making it clearer that reproduction is not replication.

... You should also emphasize the role observational data play in statistical adequacy by drawing a parallel with experimental data.

For example, for an experimental paper, I will try to reproduce it. If I can, then I will try to replicate it to confirm or deny. With observational data, if I can reproduce it, I cannot replicate it because it's observational. However, I can use statistical adequacy to deny it. If it passes statistical adequacy, while it doesn't confirm the way that a replication does for an experiment, it does lend more credence to the original study.

The status of replication with observational data depends on the definition of replication adopted. Currently, in the paper, this is interpreted in terms of Duvendack et al.'s broad definition of a replication as "any study whose main purpose is to determine the validity of one or more empirical results from a previously published study", with a specific focus on testing for statistical adequacy. From this perspective, replication is feasible for a study using observational data, with a 'successful replication' passing misspecification tests for statistical adequacy. But I agree this is not the same as what might be regarded as confirmation in the replication of an experimental study. Adding some discussion of comparisons with replications of experimental studies would therefore be useful. Indeed, this provides an opportunity to make the point that testing for statistical adequacy also has a role to play with experimental data in ensuring that the various aspects of experimental

design have been successfully applied to generate data with the expected statistical properties (Spanos and Mayo, 2015).

Spanos, A., and Mayo, D. G. (2015). Error statistical modelling and inference: Where methodology meets ontology. *Synthese*, 192, 3533–3555.

(3) page 5, line 2: "in a majority of empirical studies in economics 3." Footnote 3 does not give a citation for this claim, merely reasons for this claim. This claim needs to be documented.

This is a fair comment. The claim here is based on casual empiricism and needs to be toned down, especially the reference to "a majority of empirical studies", to make clear this is not a well-documented fact. Instead, reference will be made specifically to the studies examined in Owen (2017), for which the statement is true, and Spanos's (2017, p.15) comment that "very few applied papers in econometric journals provide sufficient evidence for the statistical adequacy of their estimated models".

(4) page 6, line 5: grammatically, "former colonies" is better than "ex colonies".
Agreed.

(5) page 12, last par of section 5: As I point out above, this needs to be reiterated at the beginning of the paper so that the reader can follow your argument. I already am quite familiar with spanometrics, but the reader who hasn't read Spanos will need this kind of help.

Agreed. This will be made clear in the Introduction and then reiterated in section 5.