Referee Report on
“Replication to Assess Statistical Adequacy”

The author argues that LSE-style testing for statistical adequacy can play a role in replication. He discusses the elements of testing for statistical adequacy and cites appropriate references. The author satisfies the four points required of the paper:
(i) a general discussion of principles about how one should do a replication
(ii) an explanation of why the “candidate” paper was selected for replication
(iii) a replication plan that applies these principles to the “candidate” article, and
(iv) a discussion of how to interpret the results of the replication (e.g., how does one know when the replication study “replicates” the original study).

I have some minor suggestions.

(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.

(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. 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.

(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.

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

(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.