

RESPONSES TO REVIEWER 2

General: We thank the reviewer for their excellent comments. In the space below, we discuss how we will address these comments if given the chance to revise and resubmit our manuscript.

1. **Comment**: “The good performance of bootstrap method is nothing new. What surprises me is that bootstrap test is not widely applied in hypothesis tests when sample size is very small.”

Response: We agree! To be fair, a major handicap in bootstrapping panel data models with both serial and cross-sectional dependence is that there has been no existing methods to do this.

The only previous paper to address this issue in the SUR space was Rilstone and Veall (1996). They worked out a bootstrapping procedure in the case of cross-sectional correlation. To be honest, this was relatively straightforward, as it only involved block bootstrapping on the cross-sectional units. The addition of serial correlation greatly complicates things. We suspect that this is why it has taken so long for somebody to come up with a bootstrapping solution for the case of both cross-sectional and serial correlation. We hope that publication of our paper will facilitate the use of bootstrapping in this area.

MINOR COMMENTS:

2. **Comment**: “The mention of parametric and nonparametric bootstrap method is not necessary, as the paper does not compare the performance of the two different methods.”

Response: We are willing to drop the comparison with the parametric bootstrap procedure if the reviewer thinks this is best. However, we believe it is useful to help the reader better understand the intuition behind the nonparametric bootstrap. It is only two lines (Step 3a and the single line of text above it).

In fact, we have estimated the parametric bootstrap model. An earlier version of this paper reported results from the parametric bootstrap. Therefore, if the reviewer is bothered by the fact that we do not include results from the parametric bootstrap, we could reference these results in a footnote. That would allow us to keep the two lines as a pedagogical aid, but also demonstrate that the parametric bootstrap is feasible and produces results very similar to the nonparametric bootstrap. We are willing to do whatever the reviewer thinks is best.

3. **Comment**: “As the paper solely focuses on the comparison of finite sample performance of test statistics, it is unfair to compare PCSE method with bootstrap method. The authors should add an additional column which reports the results of bootstrap-based PCSE method.”

Response: We prefer not to do this for two reasons. Firstly, we are interested in focusing on the PCSE estimator as it is currently employed by researchers. At the time of this writing, the associated Beck and Katz (1995) paper has been cited 2,555 times in Web of Science. We want to focus on the PCSE estimator that researchers are currently using.

Secondly, we would argue there is no reason to bootstrap the PCSE estimator. In previous research, Mantobaye, Rea and Reed (2018) demonstrate that Parks dominates the PCSE estimator with respect to efficiency. However, it suffers from poor standard error estimation. The PCSE estimator greatly improves on Parks, producing more reliable inference, albeit at the cost of efficiency.

Our paper makes clear that one can have the best of all worlds, efficiency via Parks and reliable inference via bootstrapping. In other words, whenever Parks can be estimated, there is no reason to use PCSE. A PCSE estimator with improved standard errors would still be dominated by the Parks estimator with bootstrapping. As a result, we believe there is no reason to bootstrap the PCSE estimator.

REFERENCE

Mantobaye, M., Rea, W., and Reed, W.R. Which panel data estimator should I use?: A corrigendum and extension." *Economics: The Open-Access, Open-Assessment E-Journal*, 12 (2018-4): 1–31.

Rilstone, P., & Veall, M. (1996). Using bootstrapped confidence intervals for improved inferences with seemingly unrelated regression equations. *Econometric Theory*, 12(3), 569-580.