Review of: "Digitalization of research: do ICT improve scientific production in developing countries?"

Hermann Ndoya

Potential competing interests: No potential competing interests to declare.

Digitalization of research: do ICT improve scientific production in developing countries?

Overall, while the topic is interesting, the paper needs to address a couple of limitations and shortcomings in order to be publishable. Below I outline my specific concerns.

1. The introduction should be reorganized. In the penultimate of the introduction, extant studies are not substantiated in the positioning of the study within the remit of extant literature. The study is poorly motivated, and the following questions are not clearly answered. (i) What is the policy syndrome or problem statement? (ii) What has been done by the extant contemporary literature to address the problem statement? (iii) What are gaps in the engaged literature? (iv) How does this study address the identified gaps? (v) How is addressing these gaps relevant to policy? Without clear answers to these questions, the study is not clearly situated within the context of extant knowledge economy literature.
2. You need to go in-depth with the discussion of theoretical and empirical literature. To do this, it's best to move this discussion to a separate literature review section. Moreover, There is an absence of empirical literature in the manuscript.
3. Review your way of citing. In the same manuscript, sometimes you cite as “Acemoglu, Laibson and List (2014)” and the other times “Acemoglu et al. (2014)”. The citations should be uniform.
4. In equation (2) the parameters are not numbered correctly. You move from β2 to β5 and then, from β7 to β12. What justify that?
5. The choice of variables is not justified. Indeed, no definition of the variables is presented. And the authors also don't answer the question of why these variables were chosen.
6. The discussion of the estimation strategy should be improved, and also the choice of its use. How did the authors address the issues of identification, simultaneity, and exclusion restrictions? In addition, in Appendix 2 the descriptive statistics show that the dependent variables are counting data. Why not use a counting model like the negative binomial model as your first choice? why the generalized method of moments estimator?
7. Provide the full definition of the acronym (CAR) in a footnote. refer to in Appendix 1.
8. In all of your GMM estimates there is an absence of the lagged dependent variable. You should include it in the result tables. Moreover, you should harmonized the presentation of coefficients in the tables. Sometimes coefficients are reported with 3 digits, and other times with 4 digits or 5 decimal (See Table 1 for example).
9. In addition, your results are discussable. Indeed, the validity test of Sargan's instruments is significant in all your tables. This shows that the technique used to validate the instruments is invalid. As a result, you should check your

results and include the Hansen Statistics.

10. Finally, in Table 4, the authors lump together government stability, corruption, and democracy in the same model. That variables are highly correlated and can lead to a multicollinearity issue. The authors should use them in turn, not simultaneously. I am not happy with the approach used to test the transmission channels. The authors should conduct either an interaction, or a mediation analysis.