Women’s participation in the labour market around the world (Stefan Kuhn and Sheean Yoon).

**Comments**

Thank you for sending me through this paper that looks at women’s participation using cross sectional data from the 2016 Gallup World Poll.

I have a number of suggestions for the authors to consider.

**Methods**

Let me start with an important methodological point. The authors argue that there are lots of interactions which might affect participation and complicate the analysis. They also note that some of the variables that drive these interactions might themselves be partly determined by female participation rates, so note that the coefficients need to be interpreted with caution (specifically that they are capturing correlation, more than direct causation). All of this is OK.

My concern, however, is with the construction of the gender gap variable that is used fairly heavily in the paper – e.g. in figures 5, 6 and 7 and the discussion around those figures. This variable seems very problematic, because it is constructed using information on the female participation rate in a first stage regression. Specifically, as the paper explains on p. 9, female participation is regressed on male participation rate, GDP per capita and it’s square and the residual from this regression is then used to define high or low gap. Unless I have misunderstood something, this means that the left hand side variable (participation) also enters on the right hand side (high/low gap). This is problematic because a positive shock to female participation must reduce the gap measure and similarly a negative shock will increase the gap measure. In short the way the gap variable is constructed means that it is, by construction, endogenous.

I’m not sure that this is fixable. The authors might think about *only* using an interaction with the male participation rate? I guess they could split the samples based on high and low female participation. Splitting on the left hand side variable is also problematic, however, so I’d suggest they focus on the first solution – i.e. only interacting with male participation.

Another thing that could be improved is in the description of the data. If you need to reduce the sample because of missing data, it would be useful to know if the data that you end up with is representative of the overall data set for variables which are reported for all countries. This is probably best done in an annex. If it’s not representative, I don’t think it matters (given the emphasis on interactions) but it would be helpful to know in what way it is unrepresentative.

**Structure**

My other points are non-methodological and focus on improving the write-up of the paper.

I’d encourage the authors to think more carefully about the structure of the results section. Trying to talk about multiple dimensions for heterogeneity – life cycle versus country effects – at the same time is quite difficult to keep track of. One possibility would be to focus on prime age women first. What does the basic, un-interacted regression look like for prime-age women? They could then carefully document the factors that play a role in explaining differences in participation rate for this key group. They could then do life cycle effects without worrying about the interactions before, say, turning to the differences in life-cycle effects when splitting low/middle/high income countries. Other orderings are possible, but I’d strongly encourage them to think about some way of simplifying.

Another way to put this is that they should start with a restricted version of equation (2) and then think about ways to build up to the more generalised version with lots of interactions.

Talking of interactions, and equation (2) it would be helpful to remind the reader that the notation is given in table 1 – it took me a while to figure out that the variables had been defined. Also, what’s the difference between i (lower case) and I (upper case)? Sorry if I missed this. Finally, where are the levels terms that correspond to the interactions (I realise that the country fixed effects pick up anything that is country specific).

**Write-up**

I’d also like to see the authors work a little more on the write-up. At points, it’s rather meandering, when do things that are pretty standard – see, for example, the discussion at the bottom of p.7 and top of p.8 on the use of country fixed effects. At other points, I wasn’t sure if the authors were suggesting something that they do, or that should be done, but wasn’t (e.g. the discussion on the endogeneity of thresholds at the end of the sub-group section). Finally, despite a lot of caution in the rest of the paper, the conclusion seems rather bold: ‘this paper provides strong evidence’ has a significant positive effect. Similar statements are made on a number of other interactions. This goes against the earlier caveats around causality and the need for care in interpreting the results. I’d suggest that this caution is extended to the conclusions.

I hope these comments are helpful. This is obviously a very important question and it’s good to see new (‘biggish’) data used to help find answers.