Summary
This paper proposes an alternative methodology for modelling (gender differences in) self-assessed health, as well as a theoretical motivation for it. The authors motivate their approach with concerns with gender specific reporting behaviour. Unlike previous approaches, in this one SAH is viewed as a combination of current and present value of expected future health, and the model accommodates heterogeneity in the discounting of those future values. While new efforts to model SAH and to explain the gender health paradox are welcome, I have several concerns about this paper, which I explain in detail in the remainder of this report.

Comments

1. Connection of this model with reporting behaviour
Possible gender differences in reporting behaviour, and their relation to the gender health paradox are used to motivate this paper and the proposed methodology. However, the connection between the two is not made sufficiently clear, which weakens the motivation and clarity of the contribution. I started by finding the title (about reporting behaviour) and the abstract (about heterogeneity in discount factors) contradictory. And this remained unclear throughout the paper. Besides improving clarity in this respect, it would also be important to explain to what extent this methodology has better explanatory power than others.

2. Descriptives
The paper should contain descriptive statistics of all the data used, not just of SAH. A related comment, in the introduction the authors say that the gender differences are more pronounced in Turkey compared to other countries. It would be good to see some data about this too.

3. Discounting
The authors hypothesize that heterogeneity in discount factors might drive gender differences and partly explain the gender health paradox. However, it remains unclear to me what the “correction” proposed is actually doing and implies. Do the authors mean that one should compare current health rather than current combined with discounted expected future health (ie, “uncorrected” SAH)? Is this what the model is doing? All this should become much clearer.
Given the role that the discount factors play in this methodology, the respective estimation results should also be presented. There should also be an explanation and discussion of how those results relate to the differences obtained with the base model and the extended model.
I also believe that identification of the individual specific discount factors from smoking behaviour should at least be better justified and discussed. Finally, I wonder if there is/should be some rescaling of the predictions of equation (14), with explanatory variable smoking behaviour, into the discounting factor.

4. Other comments on (presentation of) methodology
I would have liked to see a better explanation of what the assumptions described under equation (4) mean, as well a discussion of their plausibility and possible implications.
Is it plausible to assume that the income growth will be the same for both gender and education groups? Again, what are the possible implications of such assumption?
The authors repeat a couple of times that SAH is categorical and that an ordered model is needed because of that (somehow presented as a limitation). I think it would work better to establish this at the beginning and avoid repeating.
Why are there no interactions in equations (12) and (13)?

5. Other comments on (presentation of) results
What are the four lambdas in Table 1? In the equations, there is only one lambda but I do not see the relation with the four estimated ones.
Why is there a constant term in Model 2 but not in Model 1? (on the other hand, it is clear that it is not identified in the base model, the standard ordered logit)
I am puzzled by the fact that the base model, supposedly more restrictive, has a slightly higher likelihood than Model 2. I also do not understand why that of Model 1 is better than Model 2. As a more general comment, it should become much clearer what Model 1 and Model 2 actually are.
Even more generally, it should become much clearer what parameters are presented in Table 1, in relation to the equations.

6. Writing
Several sentences and paragraphs should be better written. Besides some small issues with punctuation and with articles, the lack of clarity in parts of the text impairs important messages related to the general motivation and contribution of the paper, as well as the proposed methodology. These are some examples but advise the authors to revise the whole text carefully and possibly have it proofread:
- The first sentence of the second paragraph of the introduction is too long.
- I found the third paragraph quite confusing in some respects.
- In the second full paragraph of page 3, there seems to be some confusing between risk and time preferences.
- In the first paragraph of Section 3, it is unclear whether or not the authors find SAH a problematic measure.

- I also found the first paragraph of sub-section 3.2 quite confusing. It is perhaps best to start with saying what is now in footnote 7, to avoid giving the initial impression that the model estimated will be much simpler than what was presented previously.
- At the beginning of sub-section 3.3, it seems better to present clearly what is done in this paper, rather than starting with referring to the most general case in the paper of Kose and Soytas (this should perhaps be in a footnote instead).

7. Additional references
The paper would benefit from some additional references on gender health differences and gender specific health reporting such as:
Case, A.and C. Paxson. 2005. "Sex differences in morbidity and mortality." Demography 42(2):189-214.
Peracchi, F.and C. Rossetti. 2012. "Heterogeneity in health responses and anchoring vignettes." Empirical Economics 42(2):513-538.