RESPONSES TO REVIEWER 1'S COMMENTS

We appreciate the reviewer's thoughtful review of our paper. Responses to individual comments are given below. The reviewer's comments have resulted in substantial improvements in the manuscript.

1) <u>COMMENT</u>: ***"In my view, the idea of this paper is very interesting and the motivation of this work is clearly explained and the outline of whole paper is stated in the introduction section to make readers understand the process of analysis."***

<u>Response</u>: We thank the reviewer for this encouraging comment.

2) <u>COMMENT</u>: ***"On page 6 and 7, the paper shows values of α from Cronbach's α method to check reliability of SC variables but authors just explain that the values obtained from the test are consistent with previous research on social capital … They also give a reason "there are no formal standards for acceptable values of α". Mention the exact values without providing "cutpoints" information makes readers unclear about it; authors raise the point but leave the ending of the paragraph unclear."***

<u>Response</u>: The revised version now adds that a value of 0.70 or higher is commonly considered an acceptable value of internal consistency. Values less than 0.50 are considered unacceptable (George & Mallery, 2003). See the last full paragraph on page 6.

3) <u>COMMENT</u>: ***"After checking for heteroscedasticity, the existing of the problem is confirmed so authors decide to use Heteroscedastic Probit Models. However there is a caution of using it; the heteroskedastic probit does not allow researcher to distinguish between non-constant variance and a misspecified mean function. And misspecified mean function leads researcher to sum up that the variance is non-constant. If we made conclusions about the non-constant variance from the statistical significance of coefficients in the model, we would be in a wrong way."***

<u>Response</u>: The reviewer is correct in noting that a misspecified mean function can cause one to incorrectly identify heteroskedasticity. This is no different than the OLS case, except that heteroskedasticity in nonlinear models like the Heteroskedastic Probit can result in inconsistent coefficient estimates. We take the reviewer's comment on board by noting this in a footnote (see Footnote #12). However, it is worth noting that subsequent results in the paper make clear that our conclusions are unaffected by how we handle heteroskedasticity.

4) <u>COMMENT</u>: ***"Authors explain too much detail in endogeneity topic (7 pages). Additionally, they only emphasis endogeneity from causality but in fact it is not from causality only. Other causes of the problem are Omitted variable and Measurement error which should be considered as well.***

<u>Response</u>: The original version of this paper did not address endogeneity and the responses from journals was overwhelmingly that this was a defect of the paper that needed to be addressed. As a result, the version of the paper submittted to *Economics E-Journal* expends substantial effort to address this concern. Given that other readers are concerned by endogeneity, we believe it is appropriate to provide detailed information. However, we agree with the reviewer that simultaneity is not the only cause of endogeneity. The revised version of the manuscript also points out that endogeneity bias can be generated by omitted variables. See the last full paragraph on page 15.

5) <u>COMMENT</u>: ***"Instrument variables are used to solve "unmeasurable confounders" problem, which is generally found in Administrative data. However, the factors that qualify as "instrument" are difficult to find and they are often skeptical. And it produces less precision outcomes (broader Confidence Interval band), which is not suitable for small sample. I understood that authors look for IVs that are correlated with the SC variables and uncorrelated with health outcomes. But some IVs that they introduced are not likely to be correlated."***

<u>Response</u>: We don't disagree with the reviewer that finding good instruments is a challenge. Nor that variables that satisfy the tests for a good instrument will necessarily produce "better" results, as indicated by the reviewer's note about precision. However, as reported in TABLE 5, all the variables that we use as IVs in our study are highly correlated with the endogenous variables.

6) <u>COMMENT</u>: ***For example, the authors attempt to find out the possible instrumental variables in order to correct endogeneity from the model. On page 17, they found Phone_CountyMean (average number of phones per household in the respondent's county) to be used as one of instrumental variables for SC variables. They explained that the availability of telephones where a respondent lives reduces interacting cost with others and it allows more frequent contact which is a better opportunity to develop "trust relationship". This reason might be true but in the real situation, respondent cannot accept 2 calls at the same time. In this case, having more telephones in the house doesn't make more chance to contact others and develop trust relationships. In addition, number of phone that equals to number of people in the place where respondents live sounds plausible to develop trust in this sense."***

<u>Response</u>: The average-number-of-phones-variable includes both landline and mobile phones. It combines questions about number of landline phone (*Dianhua* in Chinese) and number of mobile phones (*Shouji* in Chinese). Thus, the issue is not so much whether a respondent can accept 2 calls at the same time. Rather it is a measure of

the ease in being able to contact other members in the community. Accordingly, we believe it is an appropriate instrument for social trust. See the revised discussion in the second full paragraph on page 17.

7) <u>COMMENT</u>: ***Authors have explained about diagnostic tests to confirm Instrumental Variables' validity. But the tests for weak instruments and underidentification test, they do not explain about IID (Independent and Identically Distributed) assumption test to confirm whether the IID assumption fails. They just assume that the error terms are IID. This is being concerned because there are no concrete results in the literature to test for weak IVs when the IID assumption fails. So researchers just use asymptotic justification in a test of underidentification. In other words, the Kleibergen-Paap test can be used to test for underidentification without the IID assumption but the justification is different from that underlying the Stock-Yogo critical values. So it is somehow hand-wavy."***

<u>Response</u>: Here is what the user-written help documentation for the Stata command *ivreg210* says about the tests for underidentification when the IID assumption is not invoked: "*When the i.i.d. assumption is dropped and ivreg210 is invoked with the robust, bw or cluster options, the Cragg-Donald-based weak instruments test is no longer valid. ivreg210 instead reports a correspondingly-robust Kleibergen-Paap Wald rk F statistic. The degrees of freedom adjustment for the rk statistic is (N-L)/L1, as with the Cragg-Donald F statistic, except in the cluster-robust case, when the adjustment is N/(N-1) * (N_clust-1)/N_clust, following the standard Stata small-sample adjustment for cluster-robust. In the case of two-way clustering, N_clust is the minimum of N_clust1 and N_clust2. The critical values reported by ivreg210 for the Kleibergen-Paap statistic are the Stock-Yogo critical values for the Cragg-Donald i.i.d. case.*"

Our interpretation of this is that the sample statistics are corrected for nonspherical errors, but the critical values are not. We agree with the reviewer that the underlying statistical basis for the tests is "hand-wavy." However, we are unaware of an alternative approach, and it is our understanding that this represents current best practice when testing for underidentification.

The original version of the manuscript acknowledged this point by stating: "Note that critical values do not exist for all estimation scenarios, and that those that do exist assume that the error terms are iid" (see the bottom of page 19 in the original manuscript). In addition, footnotes below TABLE 5 state: "Note that the critical values assume iid error terms. Critical values for nonspherical errors have not been calculated." We do not know how better to communicate this point to the reader.

8) <u>COMMENT</u>: ***"Some methodology information and empirical findings in page 2 shouldn't be put in the introduction section."***

<u>Response</u>: Following the reviewer's comments, we have relocated the discussion of methodology to the bottom of page 9f.

9) **<u>COMMENT</u>: *"I couldn't find Table 1 (the pairwise result) mentioned in page 12."***

Response: The "pairwise results" statement mentioned on page 12 refers back to the discussion of TABLE 1 in the last full paragraph on page 9, which states: "Of course, these pairwise associations do not control for the influence of the other variables." To make the connection clearer, we changed the wording on page 12 (now the top of page 13) from "pairwise results" to "pairwise associations".

10) **<u>COMMENT</u>: *"Specific abbreviations should be identified at the first time that they have been introduced in the paper, e.g. IID. It is useful for readers who are not specializing in statistics or econometrics. However, it is not necessary to explain every single abbreviation, just for the confusing ones."***

Response: Following the reviewer's instructions, we have identified IID as "independently and identically distributed" when it is introduced in the paper.

11) **<u>COMMENT #1</u>: *"There are too many tables and some of them are not sequenced by the content. Also, there are unnecessary tables. If authors could sum up with less tables would be easy for readers to look at and understand."***

<u>COMMENT #2</u>: *"From the content of the paper, it seems like the paper emphasizes more on endogeneity with too much details for this issue rather than the relationship between the interested SC and health, which is a main research question. Also, there is a lot of unnecessary content. It would be enough if authors just show 2-3 best relevant calculations instead of showing all."*

<u>COMMENT #3</u>: *"Gender issue, which was introduced in the latter section, is very interesting but it would be better if authors exclude it from this paper and do it in another paper with more details and explanations."*

Response: All of the comments above address the length and coverage of the paper, so we respond to them together. Regarding the first comment, please see our response to Comment #4 above. Regarding the second comment, we do not address endogeneity until page 16 in the manuscript, so we do not think it distracts from the main research question of the paper. And, finally, we follow the reviewer's suggestion and delete Section IIIB analyzing the role of gender as a mediator of the SC/health relationship.