

I thank referee 2 for the positive comments on the paper and the very detailed suggestions as to how the paper should be rewritten to deal with the concerns he or she has about it. I think that referee 2 and I disagree only on a single point, which is what a replication should involve.

If I understand referee 2 correctly, he or she thinks that a replication should focus on assessing the robustness of the results reported in the original paper and not concern itself with attempting to select the best possible empirical model of the subject under consideration. In contrast, I think that a replication should do both. A replication that only shows that the original results are not robust to plausible changes in the specification of the regression model leaves the job half done. It is not enough to show that the original results are not robust. This simply tells us that no weight should be placed on them. It is also essential to say what results should have weight placed on them, and this requires the selection of the best possible empirical model. In addition to this reason for conducting a model-selection exercise, there is a second one, which is that such an exercise is needed in order to assess the robustness of the results in the original paper.

Let me expand on this second reason for model selection by considering the following statement made by referee 2 (note that (1.1) and (1.2) refer to the equations in Table 1 of my paper): “*in a replication paper* what I want is a Table that is 1.1 then 1.2 (to show exact replication) and then essentially a 1.3 which is *exactly* 1.2 *plus* some previously excluded variables (e.g. the province dummies) to show *precisely* why (in an mechanical econometric sense) the BHW results change” (emphasis in original). The difficulty I have with this proposal is that it is not possible to determine which previously excluded variables should be added to the BHW preferred regression models unless one conducts a model-selection exercise.

Here is a specific illustration of why I consider that a replication must involve a model-selection exercise in order to assess correctly the robustness of the results in the original paper. BHW’s preferred regression model for total industrialisation in 1849 is reported as equation (1.1) in my paper. In addition to the education variable, it includes 11 other variables, but it does not include the three variables which measure the distance of a Prussian county from the nearest provincial capital, from Berlin, and

from London. BHW add these three variables to their preferred 1849 specification in a robustness check which they report as equation (3) in their Table 7. The distance from the nearest provincial capital is not statistically significantly different from zero at conventional levels, but the other two distance measures are. The point estimate of the effect of education in this equation is only a little smaller than in BHW's preferred specification. Should these three distance variables be added to BHW's preferred regression model in addition to the provincial dummy variables and the date of a county's annexation which I argue must be included in any analysis of Prussian industrialisation?

It turns out that both distance to Berlin and distance to London have important effects on the estimated effect of education on total industrialisation in 1849. In the case of the latter variable, for example, omitting it from equation (3.1) in my paper changes the point estimate of the effect of education from -0.204 with standard error 0.101 to 0.022 with standard error 0.099. In other words, referee 2's statement that "as I understand it, the real action in the replication paper comes from the addition of the province dummies" is not correct. However, one could not have known that distance to London and distance to Berlin would prove to be so important before conducting the model-selection exercise. If I had not included these two variables, along with many others, in the general regression specification with which the model-selection procedure began, it is likely that neither their importance nor the negative effect of education on total industrialisation would have emerged from the analysis.

In order to avoid possible misunderstanding about the model-selection procedure, I would like to address a point that may arise from the previous paragraph. The estimated effects of education in my paper might be thought to be very sensitive to the variables that have been deleted. However, as stated on p. 27 and p. 32, the point estimates of the effect of education on total industrialisation in 1849 and 1849-1882 in Tables 3 and 5 are both very similar to those obtained from the general regression specifications with which the model-selection procedure began. The estimated effects of education are not affected very much by the process of removing variables that is involved in the general-to-specific model-selection procedure, and therefore they do not depend on selective variable deletion.

It is certainly true that the main reason I adduce in support of the view that BHW's analysis is flawed is that they neglect the importance of regional effects in Prussian industrialisation. It might, therefore, be argued that I should report the results of adding the provincial dummy variables, possibly also interacted with the date of a county's annexation by Prussia, to BHW's preferred specification. At the beginning of section 4 (p. 18), the paper does state, very briefly, what happens when this is done. However, although I understand the argument that this might be the way to proceed, I do not think it is a satisfactory approach, for the reasons that I set out on p. 19 of the paper. These are the general reasons why I do not agree with referee 2's view that a replication should seek to assess the robustness of the results in the original paper without engaging in a model-selection exercise.

Although I cannot agree with referee 2's view that a model-selection exercise should not form part of a replication, I do agree with his or her comments about clarifying the proximate, mechanical cause of the non-robustness of the BHW results. The points made by referee 2 about the difference between Tables 3 and 5 are valuable ones, and I shall incorporate them into a revised version of my paper.