Case B: Educational production function and policy

Assignment

Economists often think about the skill formation of students in terms of an educational production function. There are various inputs of the student, peers, teachers, schools and parents and these inputs jointly increase the skills of the student. The educational policies in a country determine the size and the quality of inputs. For example, having more teaching hours per year is more useful if there is a well-trained teacher.

In Case A the PISA data were used to evaluate the school starting age. This is only one educational institution. In this assignment you use the PISA data to empirically evaluate one or some educational policies or institutions of your choice. So, consider yourself a data driven policy advisor on educational policies with the goal to advise the different OECD countries on how they can improve the skills of 15-year-old students.

There are many educational policies that you can consider in this assignment. For example, think about class sizes, number of teaching hours per year, (early) tracking, level of skills of teachers, school closures due to Covid lockdown, composition of the peer group of classmates, etc. To investigate how these policies affect the skills of students, you can consider various variables that are available in the teacher, school principal and parents survey of PISA. For example, when considering the impact of school closures during the Covid pandemic, you may want to consider parental time investments. This might also explain possible heterogeneity in the impact between for example students with low-educated and high-educated parents.

What is important in this assignment is that you use the PISA data to make reliable policy advice. For a good policy advice, you should differentiate between causality and correlations, inform about the mechanism, consider heterogeneity in the effects and assess robustness of your results. To get additional information you may want to enrich the student database that you used in Case A with variables that are available in for example the teacher, school principal and parent survey that are provided on the official PISA website (https://www.oecd.org/pisa/data/2022database/).

APPENDIX: PISA data description

If you want to enrich the dataset from case A with the additional questionnaires for parents, teachers, schools etc. or the additional questions from the student questionnaire you can merge them as follows:

- To merge the *student data file* with the *school* or/and the *teacher data file*(s), use the country code 3-character (variable name: CNT) and the international school ID (variable name: CNTSCHID) for performing the merging process.

- To merge the *student data file* with the *parent data file*, use the country code 3-character (variable name: CNT) and the international student ID (variable name: CNTSTUID) for performing the merging process.

- To merge the *student data file* with the *cognitive or financial literacy data file(s)*, use the country code 3-character (variable name: CNT), the international school ID (variable name: CNTSCHID) and the international student ID (variable name: CNTSTUID) variables for performing the merging process.

NOTE: The Pisa dataset makes use of so-called plausible values. These plausible values are generated through multiple imputations based upon pupils' answers to the sub-set of test questions they were randomly assigned and their responses to the background questionnaires. **Analysis of such plausible values is a bit more complicated, but for the sake of brevity, you may just take the average score per student on each domain as an outcome variable for your analyses.**

NOTE 2: The Pisa dataset is based on a sample, and not on the whole population of 15-year olds. This means that you will have to use the provided sampling weight (W_FSTUWT). Technically, you would need to use another set of 80 replicate weights get correct standard errors. **However, for sake of brevity, we have omitted these and you may ignore this issue.**

If you want to dive deeper into these two issues, see here:
https://www.oecd.org/pisa/data/httpoecdorgpisadatabase-instructions.htm
At this link you can also find software packages that directly execute all steps necessary to deal with both plausible values and the sampling weights. **If you want to, you are allowed to use these packages, but whether you use them or not will not count towards the grading of your assignment.**