Final Report

# Introduction

This paper considers if, when controlling for a variety of confounding factors, an increase in the average number of years a citizen spends in education leads to an increase in productivity, as measured by GDP output per capita.

This question was chosen to contribute to the evidence base around the economic benefits of increased education and help to support appraisal of education policy proposals.

# Methodology

The originally intended scope of this analysis was the 27 EU member states, as well as the United Kingdom. However, due to limitations surrounding data availability, the final analysis covers 22 of the 28 originally intended nations. Add reference to technical annex.

While we had planned to use World Bank data on expected (average) number of years of education, initial modelling revealed that there was very little variation in this variable between years, as it had been rounded to the nearest year. The data was also subject to significant data gaps.

We therefore pivoted to Barro and Lee educational data, which was more detailed and complete. However, this data was only available in five-year intervals, so we regress years of education against GDP output per capita in five-year intervals. Later models (model 3 onwards) use the five-year rolling average of GDP output per capita as the dependent variable, rather than the value of GDP output per capita in a particular year.

The data used for this analysis have a panel data structure, and model estimates were compiled in Gretl using a Pooled OLS approach. Add why I chose to do pooled OLS?

# Data

Initially, the modelling used log.rgdpo.pop as the dependent variable. However, after receiving feedback on this modelling approach, later models changed to the variable log.rgdpo.pop.roll, which represents the five-year rolling average of log.rgdpo.pop (two years before and two years after).

Including imputed data, there is complete coverage of all the main variables outlined below between 1970 and 2015 in five-year intervals. More details of the variable imputation approaches taken can be found in the technical annex.

The nominal value of GDP output per capita in all these models is in 2017 US Dollars (US$). These estimates can be converted into present value using US GDP Implicit Price Deflator Data (as shown in the references section below), or country-specific GDP Implicit Price Deflator Data, as applicable and available.

The panel data used includes 22 of the originally intended 28 European nations.

The R script and comma-delimited copies of the underlying data are held in a private GitHub repository, with a reproducible copy of the R approach taken included in the technical annex. Should the reader wish to examine either the reproducible script or any of the underlying dataset…

Predictor variables include the following… Summary stats and charts/plots…

# Table 1: Variable Descriptions

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Description** | **Units** | **Source** |
| rgdpo.pop | Real GDP output per capita | 2017 US$ | Penn World  Tables 10.01 |
| log.rgdpo.pop | Natural log of real GDP output per capita | 2017 US$ | Penn World  Tables 10.01 |
| rgdpo.pop.roll | Rolling five-year average of real GDP output per capita | 2017 US$ | Penn World Tables 10.01 |
| log.rgdpo.pop.roll | Natural log of rolling five-year average of real GDP output per capita | 2017 US$ | Penn World Tables 10.01 |
| year\_orig | Years since 1970 | Years | Penn World  Tables 10.01 |
| yrs\_sch | Average number of years of education | Years | Barro & Lee |
| voc | Share of all students in secondary education enrolled in vocational programmes (binary: above EU average 0/1) | Numerical Factor | World Bank |
| voc\_pc | Share of all students in secondary education enrolled in vocational programmes (%) | Numerical Factor | World Bank |
| gen | Gender ratio for average years of schooling (binary: above European average 0/1) | Numerical Factor | Barro & Lee |
| avh | Average annual hours worked by persons engaged (employed) | Hours | Penn World  Tables 10.01 |
| csh\_x | Share of merchandise exports at current PPPs | Numerical Factor | Penn World  Tables 10.01 |
| fdi | Foreign Direct Investment (FDI), net inflows as a share of GDP | % of GDP | World Bank |
| ctfp | TFP level at current PPPs (USA=1) | Numerical Factor | Penn World  Tables 10.01 |

## References

Penn World Tables 10.01 Economic Data: <https://www.rug.nl/ggdc/productivity/pwt/?lang=en>

Barro and Lee Educational Attainment Data: <http://www.barrolee.com/>

World Bank Education Statistics: <https://datatopics.worldbank.org/education/>

US GDP Implicit Price Deflator Data: <https://fred.stlouisfed.org/series/GDPDEF/>

GitHub Repository for Data Reproduction: <https://github.com/jack-n-ocallaghan/ecox-5004-analysis> (Note that this is a private repository, please contact the author for direct access).

# Reporting

Text

# Conclusion

Text

*Please note that each of your policy questions differ from each other in a number of ways. This means that there is no single correct specification for what you must and need not include in your report! With only 2000 words to report the findings of your work, you will need to use words on the parts that matter the most.*

*You will need to apply your judgement to decide which issues you give emphasis to and which to down-play based on what is most important as you attempt to answer your policy question.*

**Some further guidance on Style**

*Your report is best thought of as a scientific report. It should guide the reader from that what and why of the question through the options available to find an answer and catalogue the twists and turns of the research journey (your statistical tests and decisions you subsequently make for a next step) through to the analyses to the conclusions.*

*You can also think of the report as a story of the journey through your research project which draws your reader through the analyses, persuading them that you’ve take an objective approach at each turn. Just be sure to write it in a dispassionate Third Person style – this really helps you persuade. i.e:*

*Don’t write “next I looked at” or “I believe that x assumption is too strong”*

*Do write “our next step is to consider” or “given the …. property of the data, it is likely that x assumption could be violated”.*