Assessing the impact of increasing education provision on EU productivity

# Introduction

Over the past 50 years, there has been a significant expansion in provision of education across economically developed countries. In the early 1960s, very few students were able to access higher education, and even secondary education was unavailable to most young people in many countries.[[1]](#footnote-1)

This trend gives rise to a core policy question: “has the increase in education provision in economically developed countries positively impacted nations’ productivity?”. This question is of particular importance to policymakers and politicians, to inform decision-making around the long-term impact of changes to provision of education. As outlined below, due to data limitations, this analysis focuses solely on EU nations, as well as the United Kingdom.

# Methodology

This paper considers if, when accounting 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.

The originally intended scope of this analysis included 28 European nations (the 27 EU member states plus the United Kingdom), to produce generalised estimates of the impact of an additional year of education on productivity. However, due to data limitations, this analysis only covers 22 of those nations. These nations were chosen due for their generally rich provision of historical economic and educational data, although future analysis might attempt to build on this by broadening the scope of the data used.

Further details of those nations excluded in this analysis can be found in the technical annex.

To compare the impact of education on productivity across multiple countries and over time, a panel data format was constructed. There is a reporting lag for some of the variables used, and the average number of years of education was only reported in five-year intervals. Therefore, the analysis shown covers the period 1970-2015 in five-year intervals.

Model estimates for this analysis were produced using a Pooled OLS approach. This method was selected as it allows us to consider the broad impact of an additional year of education in a straightforward way and does not consider country-specific effects. It is therefore important to caveat that these results represent generalised results for the impact of education, and coefficients might vary in scale by country.

Do OLS assumptions hold?

# Table 1: Model Descriptions and Formulae

|  |  |
| --- | --- |
| **Model Description** | **Model Formula** |
| Model 1 - Pooled OLS with core variables | rgdpo.pop = B0 + B1·yearorig + B2·yrssch + B3·voc + |
| Model 2 - Pooled OLS with core variables and ctfp | rgdpo.pop = B0 + B1·yearorig + B2·yrssch + B3·voc + B4·ctfp + |
| Model 3 - Pooled OLS with log core variables, but share of vocational students squared | ln(rgdpo.pop.roll) = B0 + B1·ln(yearorig) + B2·ln(yrssch) + B3·vocpc2 + B4·ln(ctfp) + |
| Model 4 - Pooled OLS with all variables in log form | ln(rgdpo.pop.roll) = B0 + B1·ln(yearorig) + B2·ln(yrssch) + B3·ln(vocpc) + B4·ln(ctfp) + |
| Model 5 - Pooled OLS regression model, with all variables in log form and a dummy variable for 2010 | ln(rgdpo.pop.roll) = B0 + B1·ln(yearorig) + B2·ln(yrssch) + B3·ln(vocpc) + B4·ln(ctfp)  + B5· D2010 + |

Going forwards, the above models will be referred to in shortened form (e.g., ‘model 5’). Out first step was to consider construction of a baseline model for this analysis. For this, we constructed model 1, which was a simple linear regression of rgdpo.pop against yearorig, yrssch, and voc. We then also constructed a second version, model 2, which also included the variable ctfp to account for changes in total factor productivity over time. Model 2 had a higher adjusted R2 value than model 1 (0.614 against 0.475), and ctfp was found to be significant at the 90% significance level. Therefore, we considered model 2’s more complete set of core variables to be ideal.

The next step was to examine the voc variable more closely. Although the value was not statistically significant at the 90% significance level, with p-values of 0.16 (model 1) and 0.19 (model 2) respectively, it was still relatively close to the threshold. Additionally, alongside policy implications of spending more time in education, it might be useful to have some idea of the impact of a more vocational style of education, even if tentative. We therefore kept the vocational variable in, switching to vocpc in the hope of better capturing more of the variance in productivity than was achieved using voc.

We then tested for autocorrelation using the Durbin-Watson (DW) test. For both models the DW statistic was in the zone of indecision, meaning that we are not able to not reject or reject the presence of autocorrelation.

Additionally, we also tested for heteroscedasticity using White’s test, which concluded that there was heteroscedasticity present in the data. We attempted to reduce these issues using log-transformed variables, which had limited positive impact.

After implementing log-transformed variables, we also implemented heteroscedasticity-robust (HAC) standard errors, and two more changes. The first was to switch from rgdpo.pop to rgdpo.pop.roll, which represents the five-year rolling average of GDP output per capita. This is useful given our five-year restriction in variable sampling, as it reduces the risk of recessionary declines in GDP producing biased coefficient estimates. The second was to switch from voc to vocpc for the reasons outlined above.

Models 3 and 4 were then produced. Given the log-log form of these models, this also gave the advantage of making it much simpler to interpret model coefficients. For instance, from model 4 we can see that a 1% increase in yrssch will lead to a 0.69% increase in GDP output per capita, when accounting for factors such as changes in total factor productivity and vocational offering. [Add more about model 3 vs model 4?].

We then also produced model 5, which included a dummy variable for the year 2010 to control for the well-documented negative impacts of the 2008 financial crisis on GDP output per capita.[[2]](#footnote-2) However, this produced a counter-intuitive result, suggesting that GDP output per capita was above-trend in 2010. Multiple empirical hypotheses were considered for this result, such as the theory that the per capita GDP data used might only include employed persons, however none appeared to explain this coefficient. Because of this, model 4 is considered the preferred model of those produced.

# 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…

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.

My data are panel data of economic and educational variables for 22 European countries between 1970 and 2015.

# 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

Do not remove

1. Gurría, A. (2011) Editorial: Fifty years of change in education [Online], p. 1. Available from <https://www.oecd.org/education/skills-beyond-school/48642586.pdf> [Accessed 16th November 2023] [↑](#footnote-ref-1)
2. Oulton, N. and Sebastiá-Barriel, M. (2013) Long and short-term effects of the financial crisis on labour productivity, capital and output [Online], p. 2. Available from <https://www.bankofengland.co.uk/working-paper/2013/long-and-short-term-effects-of-the-financial-crisis-on-labour-productivity-capital-and-output> [Accessed 16th November 2023] [↑](#footnote-ref-2)