Assessing the impact of increasing education provision on EU productivity

Jack O’Callaghan

University of Kent

[jno6@kent.ac.uk](mailto:jno6@kent.ac.uk)

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)

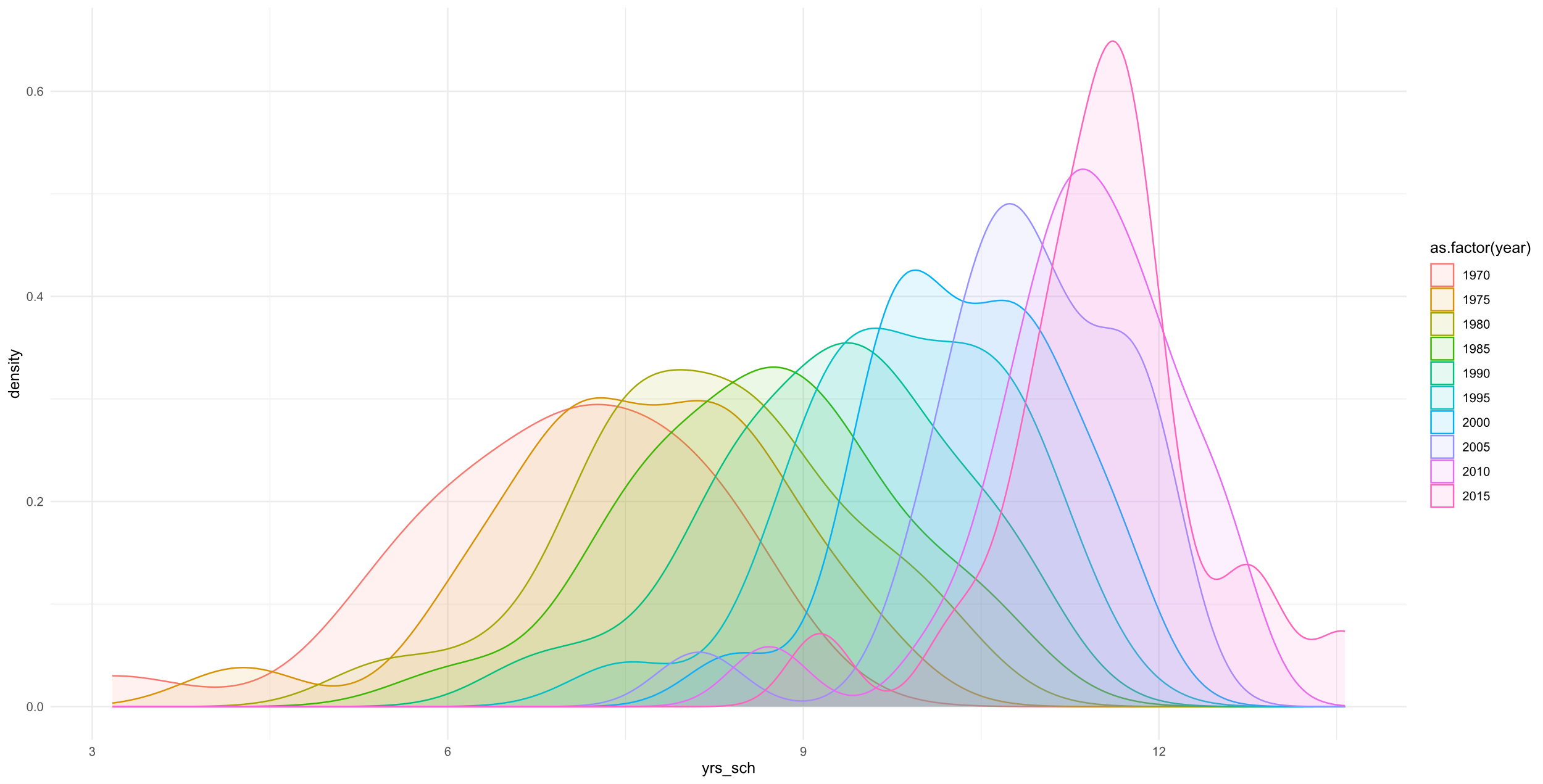
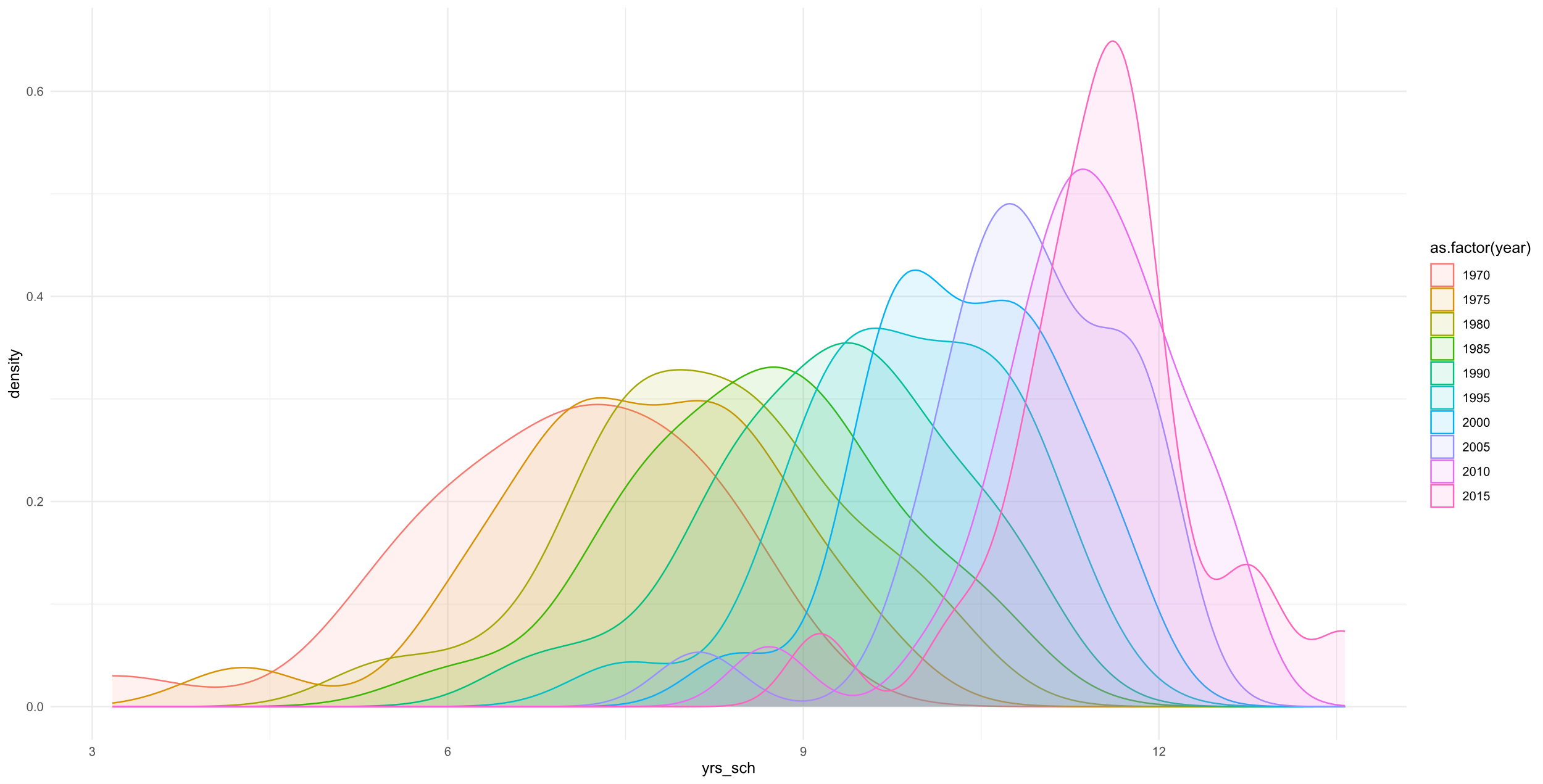
Since then, we have seen large expansions in education provision, giving rise to the question: “has the increase in education provision in economically developed countries positively impacted nations’ productivity?”. This is of particular importance to policymakers and politicians, to inform decision-making around the long-term impact of changes to provision of education. Due to data limitations, this analysis focuses solely on EU nations, as well as the United Kingdom.

# Data

The data used are panel data of economic and educational variables for 22 European countries between 1970 and 2015. The data sets used in this analysis can be obtained using either the original source link or in the GitHub repository below. The R script used to produce the combined data set can be examined using the same repository link.

Using Figure 1 below, we can identify two key trends in education over time. Firstly, we have seen a significant rise in the average number of years of education over the 45 years examined. Secondly, the data appears to have become more leptokurtic over time, as the number of years in education has become more standardised across nations. This might be a potential source of heteroscedasticity within models, as error sizes might increase over time, and may imply a need for heteroscedasticity-robust (HAC) standard errors in later models.

## Figure 1: Change in the distribution of average years of education over time



It also appears that there is a strong relationship between the number of years of education and output per capita, with 40% of the variance in output per capita explained by the average number of years in education. However, this relationship may be subject to some collinearity, where the number of years of education also increases as countries become wealthier, meaning that we must consider a wider range of factors than simply average years of education.

## Figure 2: Years of Education against GDP Output per Capita

From Table 1 below, we can see that there is a large amount of variation in output per capita between countries and over time, as represented by rgdpo.pop. Furthermore, there seems to be very little skew to the data for the number of years of education, with little difference between the mean and median values.

## Table 1: Descriptive Statistics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Minimum** | **1st Q.** | **Median** | **Mean** | **3rd Q.** | **Maximum** | **Standard Deviation** |
| rgdpo.pop | 3,002 | 15,534 | 23,531 | 25,538 | 33,447 | 82,382 | 14,032 |
| log.rgdpo.pop | 8.01 | 9.65 | 10.07 | 9.99 | 10.42 | 11.32 | 0.60 |
| yearorig | 0.00 | 10.00 | 22.50 | 22.50 | 35.00 | 45.00 | 14.40 |
| yrssch | 3.17 | 8.16 | 9.64 | 9.45 | 10.93 | 13.57 | 1.84 |
| voc | 0.00 | 0.00 | 0.50 | 0.50 | 1.00 | 1.00 | 0.50 |
| vocpc | 1.30 | 16.66 | 27.72 | 27.25 | 34.85 | 69.01 | 13.47 |
| ctfp | 0.45 | 0.76 | 0.87 | 0.88 | 0.98 | 1.44 | 0.18 |

From a visual examination of model 4’s residual plot, the model appears to be slightly heteroscedastic, which White’s Test confirms. This suggests that the log-log model form has not eliminated the heteroscedasticity issue in data. A future analysis might attempt to resolve this challenge by considering a wider range of model forms.

## 

## References

Penn World Tables 10.01: <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/>

Repository for Data Reproduction: <https://github.com/jack-n-ocallaghan/ecox-5004-analysis>

# 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 intended scope of this analysis originally included 28 European nations (the 27 EU member states plus the United Kingdom). However, due to data limitations, this analysis only covers 22 of those nations. These nations were chosen due to their historic educational and economic data availability, 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. Due to reporting lag and data restrictions, the analysis 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 impact at the broad European level and 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 2: Model Descriptions and Formulae

|  |  |
| --- | --- |
| **Model Description** | **Model Formula[[2]](#footnote-2)** |
| 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 + |

Our first step was to consider construction of a baseline model for this analysis, starting with model 1, which was a simple linear regression of rgdpo.pop against yearorig, yrssch, and voc. Model 2 expanded to include 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. This suggests that the inclusion of ctfp might improve the model.

Next, we examined the voc variable. Although not statistically significant at the 90% significance level, with p-values of 0.16 (model 1) and 0.19 (model 2), it was still relatively close to the significance threshold. Alongside the policy implications of increased education, we continue to investigate the impact of vocational education. We therefore switched to vocpc to better capture more of variance in productivity than was previously achieved.

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 cannot not reject or reject the presence of autocorrelation.

Additionally, we also repeated White’s test for heteroscedasticity on the log-transformed models, 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 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 as previously discussed.

Models 3 and 4 were then produced. Given the log-log form of these models, this made it simpler to interpret model coefficients. For instance, from model 4 we can see in Table 3 that a 1% increase in yrssch will lead to a 0.70% increase in output per capita, when accounting for factors such as changes in total factor productivity and vocational offering.

We then also produced a final model, which included a dummy variable for the year 2010 to control for the impacts of the 2008 financial crisis.[[3]](#footnote-3) However, this produced a counter-intuitive result, suggesting that output per capita was above-trend in 2010. This is likely an atypical result, and because of this, model 4 is considered the preferred model of those produced.

# Reporting

## Table 3: Summary of model coefficients

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| constant | (−) 934 | (−) 26,984\*\*\* | 7.64\*\*\* | 7.59\*\*\* | 7.71\*\*\* |
| yearorig | 508\*\*\* | 464\*\*\* | 0.33\*\*\* | 0.33\*\*\* | 0.31\*\*\* |
| yrssch | 1,488\*\* | 1,572\*\*\* | 0.70\*\*\* | 0.70\*\*\* | 0.66\*\*\* |
| voc | 1,973 | 1,606 | -- | -- | -- |
| vocpc | -- | -- | 0.00 | 0.02 | 0.02 |
| ctfp | -- | 30,131\*\*\* | 1.11\*\*\* | 1.11\*\*\* | 1.12\*\*\* |
| D2010 | -- | -- | -- | -- | 0.16\*\* |

It is important to note that due to difference in singular model structure, the scale of coefficients between these models is not directly comparable.

From model 3 onwards, results were more consistent, therefore we only consider these models going forwards.

Our estimates suggest a 10% rise in the average number of years spent in education leads to between a 6.6% and 7.0% rise in output per capita holding all else equal.

Additionally, we found no statistically significant link in this analysis between the type of education provided and output per capita. However, it is important to caveat this by emphasising that lagged variables could not be implemented due to limited sample size.

There was a strong positive link between increasing total factor productivity and output per capita.

# Conclusion

This paper examined the relationship between the number of years spent in education and productivity across EU nations, as measured by output per capita. The preferred model, model 4, took the form:

ln(rgdpo.pop.roll) = B0 + B1·ln(year\_orig) + B2·ln(yrs\_sch) + B3·ln(voc\_pc) + B4·ln(ctfp) + ε

Our modelling estimates that a 10% rise in the average number of years spent in education is associated with between a 6.6% and 7.0% rise in GDP output per capita. This result can help inform educational policymaking and appraisal but represents a generalised relationship between education and productivity across European nations, and does not account for individual member state differences.

It is important to state that the size of the effect mentioned is likely to vary significantly across different countries, depending on a variety of factors. This analysis also considers the impact of education solely in terms of the change in number of years spent in education. Other factors, such as the quality and type of educational programmes offered, are important considerations when designing education policy. Education positive spillover, not just economic impact…

Future analysis could expand on this by sourcing more detailed data for average number of years in education, thus allowing for a move from quinquennial panel data to annual. Moreover, there are some variables that this analysis did not cover due to time limitations that might improve the robustness of these findings, including adding a qualitative variable for increased female participation in the labour force over time, and the impact of other core economic variables, such as exports and foreign direct investment, and their potential relationship on GDP output.

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. rgdpo.pop = Real GDP output per capita, rgdpo.pop.roll = Rolling five year average of real GDP output per capita, B0 = Model constant, yearorig = Number of years since 1970, yrssch = Average number of years of education, voc = Dummy variable: ‘Is the share of students in vocational education above the EU average?’, vocpc = Share of students in vocational education, ctfp = Current total factor productivity, D2010 = Dummy variable for the year 2010, ε = Error term [↑](#footnote-ref-2)
3. 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-3)