

Assessing the impact of increasing education provision on EU productivity

Jack O'Callaghan

University of Kent

jno6@kent.ac.uk

Assessing the impact of increasing education provision on EU productivity

Introduction

There has been a significant expansion over the past 50 years 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

Additionally, growth in labour productivity post-financial crisis in the euro area has slowed.²

As a policy measure that theoretically could impact this trend, policymakers might ask:

“should education have a contributing role towards spurring slowing EU labour productivity growth?”. This paper considers the relationship between years in education and productivity, measured by GDP output per capita (output per capita), to support informed decision and policy-making around the impact of changes to provision of education on output.

Our modelling suggests that a 10% rise in years spent in education is associated with a 6.6% to 7.0% rise in output per capita. This finding is a general result across EU nations, and likely varies based on factors such as how additional years in education are distributed across citizens, as well as qualitative factors such as the type of education offered.

Data

The data used are panel data of economic and educational variables for 22 European nations between 1970 and 2015. All 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.

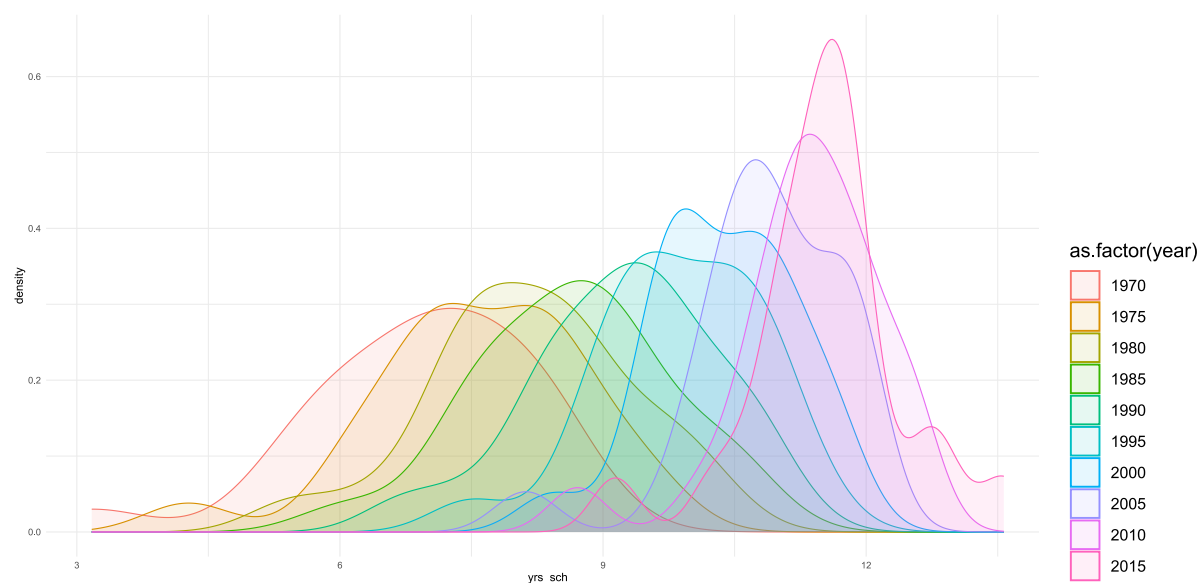
Using Figure 1 below, we identify two trends in education over time. Firstly, there has been a significant rise in the average number of years in education over the 45 years examined.

Secondly, the data appears to have become more leptokurtic over time, meaning that there is less variation in years spent in education by country. This could be a potential source of heteroscedasticity within modelling, as error sizes might increase over time, and may imply a need for heteroscedasticity-robust standard errors in this modelling.

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]

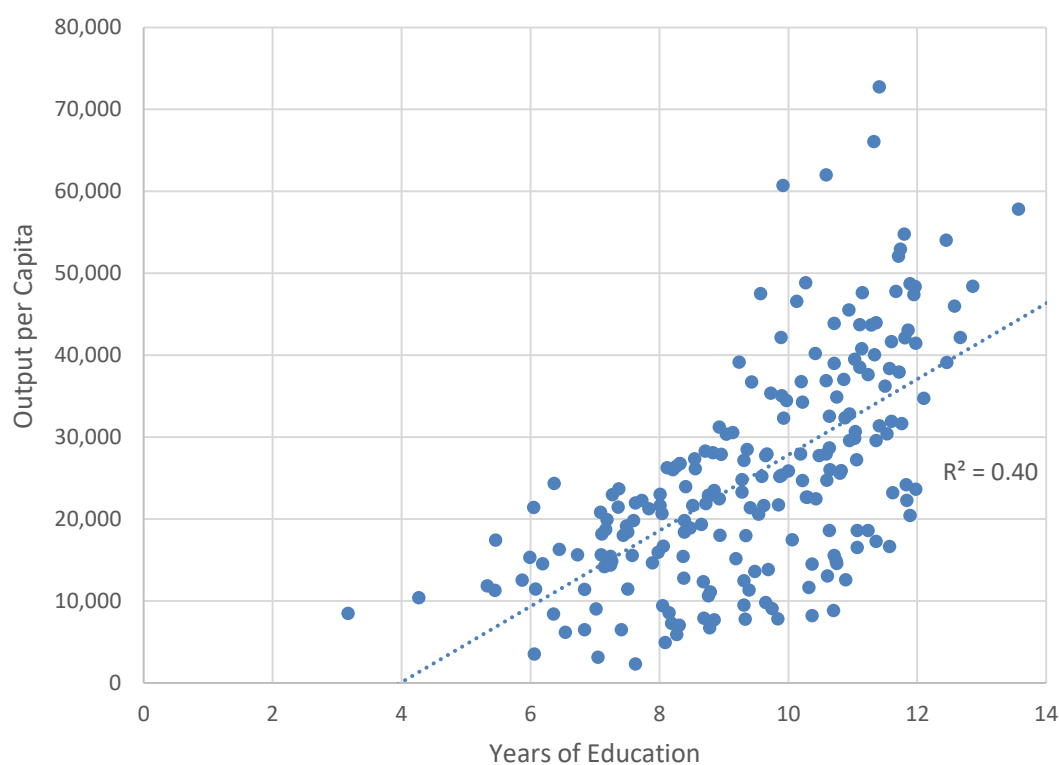
² ECB Economic Bulletin (2017) The slowdown in euro area productivity in a global context, pp. 1. Available from https://www.ecb.europa.eu/pub/pdf/other/ebart201703_01.en.pdf [Accessed 17th November 2023]

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



There is a significant relationship between years in education and output per capita, with 40% of the variance in output per capita explained by years in education. However, this relationship may be bidirectional, where the number of years in education also increases as countries become wealthier. We expand upon this univariate modelling viewpoint in our models to capture a wider range of factors, such as factor productivity improvements.

Figure 2: Years of education against output per capita



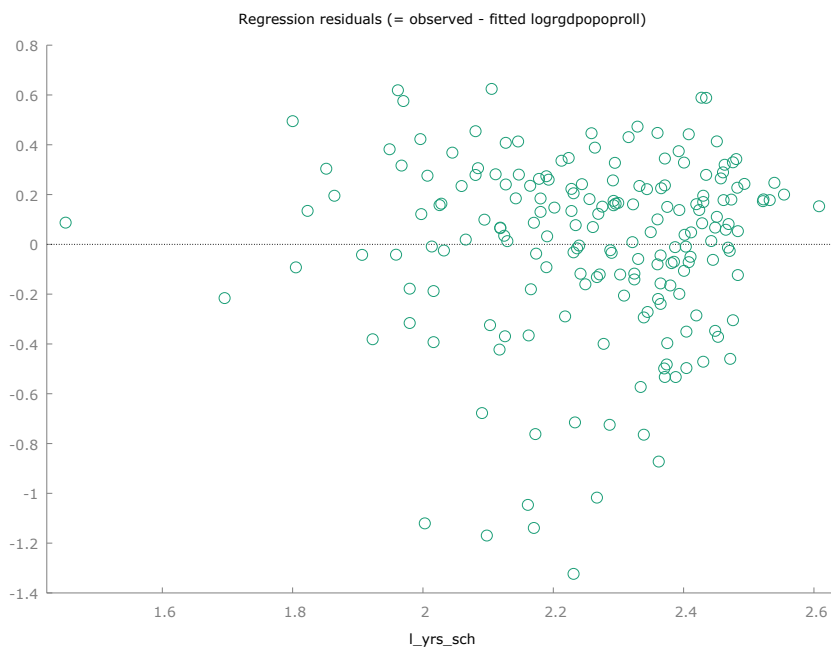
From Table 1, we observe a large amount of variation in output per capita (rgdpo.pop) between countries and over time. Furthermore, there is little difference between the mean and median values for the number of years of education, suggesting little skew to this variable.

Table 1: Descriptive statistics (Variable definitions referenced)

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
year _{orig}	0.00	10.00	22.50	22.50	35.00	45.00	14.40
yr _{Ssch}	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
voc _{pc}	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

A visual examination of Figure 3 suggests our modelling is slightly heteroscedastic, which White's Test confirms. Hence, a log-log model structure does not eliminate the heteroscedasticity issue identified in these models. A future analysis might consider a wider range of model forms.

Figure 3: Regression residuals plot for Model 4



References

Penn World Tables 10.01: <https://www.rug.nl/ggdc/productivity/pwt>
Barro and Lee Educational Attainment Data: <http://www.barrolee.com/>
World Bank Education Statistics: <https://datatopics.worldbank.org/education/>
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 years in education leads to higher productivity, as measured by 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 covers 22 of those nations. European nations were chosen due to their historic educational and economic data availability, although future analysis might build on this by including a wider range of countries. Further details of the nations excluded in this analysis are in the technical annex. To compare the impact of education on productivity across multiple countries and over time, a panel data format was used. 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, which allows us to consider variable relationships at the broad European level, not considering country-specific effects. Results represent the generalised impact of education, and coefficients might vary in scale by country.

OLS assumptions such as linearity in parameters (model specification), no autocorrelation (Durbin-Watson test), and no multicollinearity (cross-correlation matrix) have been tested and are satisfied. Assumptions such as homoscedastic error terms (White's test) have been tested and are not satisfied. Test results can be found in the repository's technical annex.

Table 2: Model descriptions and formulae

Model Description	Model Formula ³
Model 1 - Pooled OLS with core variables	$\text{rgdpo.pop} = B_0 + B_1 \cdot \text{year}_{\text{orig}} + B_2 \cdot \text{yrs}_{\text{sch}} + B_3 \cdot \text{voc} + \varepsilon$

³ rgdpo.pop = Real GDP output per capita, rgdpo.pop.roll = Rolling five-year average of real GDP output per capita, B_0 = Model constant, $\text{year}_{\text{orig}}$ = Number of years since 1970, yrs_{sch} = Average number of years of education, voc = Dummy variable: 'Is the share of students in vocational education above the EU average?', voc_{pc} = Share of students in vocational education, ctfp = Current total factor productivity, D_{2010} = Dummy variable for the year 2010, ε = Error term

Model 2 - Pooled OLS with core variables and ctfp	$\text{rgdpo.pop} = B_0 + B_1 \cdot \text{year}_{\text{orig}} + B_2 \cdot \text{yrs}_{\text{sch}} + B_3 \cdot \text{voc} + B_4 \cdot \text{ctfp} + \varepsilon$
Model 3 - Pooled OLS with log variables, but (share of vocational students) ²	$\ln(\text{rgdpo.pop.roll}) = B_0 + B_1 \cdot \ln(\text{year}_{\text{orig}}) + B_2 \cdot \ln(\text{yrs}_{\text{sch}}) + B_3 \cdot \text{voc}_{\text{pc}}^2 + B_4 \cdot \ln(\text{ctfp}) + \varepsilon$
Model 4 - Pooled OLS with all variables in log form	$\ln(\text{rgdpo.pop.roll}) = B_0 + B_1 \cdot \ln(\text{year}_{\text{orig}}) + B_2 \cdot \ln(\text{yrs}_{\text{sch}}) + B_3 \cdot \ln(\text{voc}_{\text{pc}}) + B_4 \cdot \ln(\text{ctfp}) + \varepsilon$
Model 5 - Pooled OLS regression model, with all variables in log form and a dummy variable for 2010	$\ln(\text{rgdpo.pop.roll}) = B_0 + B_1 \cdot \ln(\text{year}_{\text{orig}}) + B_2 \cdot \ln(\text{yrs}_{\text{sch}}) + B_3 \cdot \ln(\text{voc}_{\text{pc}}) + B_4 \cdot \ln(\text{ctfp}) + B_5 \cdot D_{2010} + \varepsilon$

Our first step was to consider construction of a baseline model for this analysis, starting with Model 1, which regressed rgdpo.pop against $\text{year}_{\text{orig}}$, yrs_{sch} , 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 R^2 value than Model 1 (0.614 against 0.475), and ctfp was found to be significant at the 90% significance level, suggesting that ctfp might improve modelling.

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 aim to investigate the impact of vocational education. We therefore switched from voc to voc_{pc} 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 neither reject nor fail to reject the presence of autocorrelation. Additionally, we also repeated White's test for heteroscedasticity on the log-transformed models, which had little 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 short-term recessionary declines in GDP producing biased coefficient estimates. The second was to switch from voc to voc_{pc} as previously discussed.

Models 3 and 4 were then produced. Given log-log model structure, this made it simpler to interpret model coefficients. For instance, Model 4 in Table 3 suggests that a 1% increase in yrs_{sch} 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.⁴ 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.

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***
year _{orig}	508***	464***	0.33***	0.33***	0.31***
yrSch	1,488**	1,572***	0.70***	0.70***	0.66***
voc	1,973	1,606	--	--	--
VOC _{pc}	--	--	0.00	0.02	0.02
ctfp	--	30,131***	1.11***	1.11***	1.12***
D ₂₀₁₀	--	--	--	--	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, and therefore we only consider these models in the conclusion.

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 years spent in education and productivity across EU nations, measured by output per capita. The preferred model (Model 4) was:

⁴ Oulton, N. and Sebastián-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]

$$\ln(\text{rgdpo.pop.roll}) = B_0 + B_1 \cdot \ln(\text{year_orig}) + B_2 \cdot \ln(\text{yrs_sch}) + B_3 \cdot \ln(\text{voc_pc}) + B_4 \cdot \ln(\text{ctfp}) + \varepsilon$$

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 output per capita. This result can help to inform educational policymaking and appraisal, but represents a generalised relationship, 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 change in years spent in education. Other factors, such as the quality and type of educational programmes offered, are important considerations when designing education policy. Additionally, this analysis does not consider the positive spillover effects provided by education, such as social benefits, and represents a single element of a much wider frame of reference used to appraise educational policy-making.

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 a qualitative variable for increased female participation in the labour force over time, and other core economic variables, such as exports and foreign direct investment, and their potential relationship on GDP output.