Technical Annex

My initial plan was to use World Bank data on expected (average) number of years of education. However, an initial attempt to run a simple OLS regression model on this variable revealed that, there was very little variation in this variable, as it was rounded to the nearest year, and was contained missing data points in earlier years. Furthermore, when examining the data for other European countries, I found similar missing data points.

Hence, I pivoted to an alternative data source from Our World in Data, and Barro and Lee, which was much more complete. These data are updated in five-year intervals, meaning that I had to perform my regression analysis using only five-year intervals between 1970 and 2015 inclusive, but the data was much more complete.

There was still some missing data for the following countries: Croatia, Estonia, Latvia, Lithuania, Slovakia, Slovenia. This missing data covered 1970 to 2010 inclusive for all six countries, which was considered significant enough that it became necessary to exclude them from this analysis.

There were some gaps in the data for Foreign Direct Investment (FDI) inflows as a share of GDP. Where this occurred, we calculated the median FDI share for each country and imputed this in the data series.

There were some gaps in the share of all students in secondary education enrolled in vocational programmes. Where these occurred, the most recent data point was imputed.

My first two models used a dummy variable for if the share of students enrolled in vocational secondary education was above the EU average, whereas all models after this used the exact share of students enrolled in vocational education.

The first two models used a GDP per capita, whereas all models after this used a five-year rolling average of GDP per capita.

# Model 1: Pooled OLS Model

rgdpo.pop = const + B1·year\_orig + B2·yrs\_sch + B3·voc + Error

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Variable* | *Coefficient* | *Std. Error* | *t-ratio* | *p-value* |  |
| const | −934.224 | 4994.95 | −0.1870 | 0.8518 |  |
| year\_orig | 507.573 | 87.7213 | 5.786 | <0.0001 | \*\*\* |
| yrs\_sch | 1488.12 | 684.626 | 2.174 | 0.0309 | \*\* |
| voc | 1973.48 | 1409.87 | 1.400 | 0.1631 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 25537.53 |  | S.D. dependent var | 14031.57 |
| Sum squared resid. | 2.13e+10 |  | S.E. of regression | 10168.31 |
| R-squared | 0.482386 |  | Adjusted R-squared | 0.474848 |
| F (3, 206) | 63.99327 |  | P-value(F) | 2.80e-29 |
| Log-likelihood | −2233.634 |  | Akaike criterion | 4475.269 |
| Schwarz criterion | 4488.657 |  | Hannan-Quinn | 4480.681 |
| rho | 0.099718 |  | Durbin-Watson | 1.635858 |

# Model 2: Pooled OLS Model with ctfp

rgdpo.pop = const + B1·year\_orig + B2·yrs\_sch + B3·voc + B4·ctfp + Error

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Variable* | *Coefficient* | *Std. Error* | *t-ratio* | *p-value* |  |
| const | −26984.1 | 5159.49 | −5.230 | <0.0001 | \*\*\* |
| year\_orig | 463.974 | 75.6221 | 6.135 | <0.0001 | \*\*\* |
| yrs\_sch | 1572.45 | 585.182 | 2.687 | 0.0078 | \*\*\* |
| voc | 1606.11 | 1209.39 | 1.328 | 0.1857 |  |
| ctfp | 30130.7 | 3359.27 | 8.969 | <0.0001 | \*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 25828.86 |  | S.D. dependent var | 13920.36 |
| Sum squared resid. | 1.51e+10 |  | S.E. of regression | 8653.999 |
| R-squared | 0.621020 |  | Adjusted R-squared | 0.613515 |
| F(4, 202) | 82.75230 |  | P-value(F) | 1.76e-41 |
| Log-likelihood | −2167.805 |  | Akaike criterion | 4345.611 |
| Schwarz criterion | 4362.274 |  | Hannan-Quinn | 4352.349 |
| rho | 0.028359 |  | Durbin-Watson | 1.702494 |

# Model 3: Pooled OLS Model using log variables, vocational education share squared, and 5-year GDP averages.

ln(rgdpo.pop.roll) = const + B1·ln(year\_orig) + B2·ln(yrs\_sch) + B3·voc\_pc2 + Error

Robust (HAC) standard errors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Coefficient* | *Std. Error* | *t-ratio* | *p-value* |  |
| const | 7.64001 | 0.304658 | 25.08 | <0.0001 | \*\*\* |
| year\_orig | 0.325555 | 0.0683222 | 4.765 | 0.0001 | \*\*\* |
| yrs\_sch | 0.698127 | 0.167528 | 4.167 | 0.0005 | \*\*\* |
| voc\_pc | 1.18977e-05 | 2.40917e-05 | 0.4939 | 0.6268 |  |
| ctfp | 1.11083 | 0.176448 | 6.296 | <0.0001 | \*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 10.05903 |  | S.D. dependent var | 0.544711 |
| Sum squared resid. | 24.11103 |  | S.E. of regression | 0.361992 |
| R-squared | 0.567759 |  | Adjusted R-squared | 0.558363 |
| F(4, 20) | 138.0901 |  | P-value(F) | 2.89e-14 |
| Log-likelihood | −73.59656 |  | Akaike criterion | 157.1931 |
| Schwarz criterion | 173.4019 |  | Hannan-Quinn | 163.7597 |
| rho | −0.063261 |  | Durbin-Watson | 1.710862 |

# Model 4: Pooled OLS Model using all log variables and 5-year GDP averages.

ln(rgdpo.pop.roll) = const + B1·ln(year\_orig) + B2·ln(yrs\_sch) + B3·ln(voc\_pc) + Error

Robust (HAC) standard errors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Coefficient* | *Std. Error* | *t-ratio* | *p-value* |  |
| const | 7.58841 | 0.297356 | 25.52 | <0.0001 | \*\*\* |
| year\_orig | 0.325523 | 0.0667768 | 4.875 | <0.0001 | \*\*\* |
| yrs\_sch | 0.694488 | 0.158667 | 4.377 | 0.0003 | \*\*\* |
| voc\_pc | 0.0224764 | 0.0381011 | 0.5899 | 0.5619 |  |
| ctfp | 1.10803 | 0.179253 | 6.181 | <0.0001 | \*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 10.05903 |  | S.D. dependent var | 0.544711 |
| Sum squared resid. | 24.08947 |  | S.E. of regression | 0.361830 |
| R-squared | 0.568146 |  | Adjusted R-squared | 0.558758 |
| F(4, 20) | 155.8689 |  | P-value(F) | 8.99e-15 |
| Log-likelihood | −73.51202 |  | Akaike criterion | 157.0240 |
| Schwarz criterion | 173.2328 |  | Hannan-Quinn | 163.5906 |
| rho | −0.054237 |  | Durbin-Watson | 1.697535 |

# Table 4: Model Coefficient Comparison

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Model* | *R2* | *Adj. R2* | *Durbin-Watson* | *Durbin Lower* | *Durbin Upper* | *White’s Test* |
| Model 1 | 0.482 | 0.475 | 1.636 | 1.61 (k=4) | 1.74 (k=4) | 0.000013 |
| Model 2 | 0.621 | 0.614 | 1.702 | 1.59 (k=5) | 1.76 (k=5) | 0.000000 |
| Model 3 | 0.567 | 0.558 | 1.711 | 1.59 (k=5) | 1.76 (k=5) | 0.000001 |
| Model 4 | 0.568 | 0.559 | 1.698 | 1.59 (k=5) | 1.76 (k=5) | 0.000004 |