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 + B4·ln(ctfp) + 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) + B4·ln(ctfp) + 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 |

# Model 5: Pooled OLS Model using all log variables, 5-year GDP averages, and a dummy for the year 2010.

ln(rgdpo.pop.roll) = const + B1·ln(year\_orig) + B2·ln(yrs\_sch) + B3·ln(voc\_pc) + B4·ln(ctfp) + B5· D2010+ Error

Robust (HAC) standard errors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Coefficient* | *Std. Error* | *t-ratio* | *p-value* |  |
| const | 7.70510 | 0.304234 | 25.33 | <0.0001 | \*\*\* |
| year\_orig | 0.307550 | 0.0710073 | 4.331 | 0.0003 | \*\*\* |
| yrs\_sch | 0.661255 | 0.161539 | 4.093 | 0.0006 | \*\*\* |
| voc\_pc | 0.0215616 | 0.0375893 | 0.5736 | 0.5726 |  |
| ctfp | 1.11915 | 0.179258 | 6.243 | <0.0001 | \*\*\* |
| D2010 | 0.156027 | 0.0665769 | 2.344 | 0.0295 | \*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean dependent var | 10.05903 |  | S.D. dependent var | 0.544711 |
| Sum squared resid | 23.69261 |  | S.E. of regression | 0.359816 |
| R-squared | 0.575260 |  | Adjusted R-squared | 0.563655 |
| F (5, 20) | 123.4479 |  | P-value(F) | 2.52e-14 |
| Log-likelihood | −71.94221 |  | Akaike criterion | 155.8844 |
| Schwarz criterion | 175.3349 |  | Hannan-Quinn | 163.7643 |
| rho | −0.084061 |  | Durbin-Watson | 1.726653 |

# 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 |
| 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 |

# Table 2: Model Descriptions

|  |  |  |
| --- | --- | --- |
| **Model Name** | **Model Description** | **Model Formula** |
| Pooled OLS Model | Pooled OLS regression model | rgdpo.pop = const + B1·year\_orig + B2·yrs\_sch + B3·voc + Error |
| Pooled OLS Model with ctfp | Pooled OLS regression model, with a ctfp variable added | rgdpo.pop = const + B1·year\_orig + B2·yrs\_sch + B3·voc + B4·ctfp + Error |
| Pooled OLS Log-Log Model with voc\_pc2 | Pooled OLS regression model, with all variables in log form, excluding the share of vocational students | ln(rgdpo.pop.roll) = const + B1·ln(year\_orig) + B2·ln(yrs\_sch) + B3·voc\_pc2 + Error |
| Pooled OLS Log-Log Model | Pooled OLS regression model, with all variables in log form | ln(rgdpo.pop.roll) = const + B1·ln(year\_orig) + B2·ln(yrs\_sch) + B3·ln(voc\_pc) + Error |
| Pooled OLS Log-Log Model with a 2010 dummy | Pooled OLS regression model, with all variables in log form and a dummy variable for 2010 | ln(rgdpo.pop.roll) = const + B1·ln(year\_orig) + B2·ln(yrs\_sch) + B3·ln(voc\_pc) + B4·ln(ctfp) + B5· D2010+ Error |

# Table 3: Summary Statistics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Minimum** | **1st Q.** | **Median** | **Mean** | **3rd Q.** | **Maximum** | **Standard Deviation** |
| rgdpo.pop | 3002.00 | 15534.00 | 23531.00 | 25538.00 | 33447.00 | 82382.00 | 14031.57 |
| 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 |
| yrs\_sch | 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 |
| gen | 0.00 | 0.00 | 0.00 | 0.48 | 1.00 | 1.00 | 0.50 |
| avh | 1401.00 | 1627.00 | 1766.00 | 1778.00 | 1903.00 | 2334.00 | 207.67 |
| csh\_x | 0.00 | 0.16 | 0.27 | 0.34 | 0.49 | 1.22 | 0.24 |
| fdi | -15.71 | 0.58 | 1.95 | 10.36 | 5.39 | 340.26 | 31.16 |
| ctfp | 0.45 | 0.76 | 0.87 | 0.88 | 0.98 | 1.44 | 0.18 |

# 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 |
| Model 5 | 0.575 | 0.564 | 1.727 | 1.57 (k=6) | 1.78 (k=6) | 0.000032 |

# Chart 1: GDP Output per Capita Over Time by Country

One reason that the drop in 2010 does not look as severe as is might otherwise is that the 2010 figure represents the five-year average of GDP output per capita, between 2008 and 2012. It could be that, had we been able to use the annual data for this analysis, a control for the 2008 financial crisis would have been more able to capture a fall in GDP output post-2008.

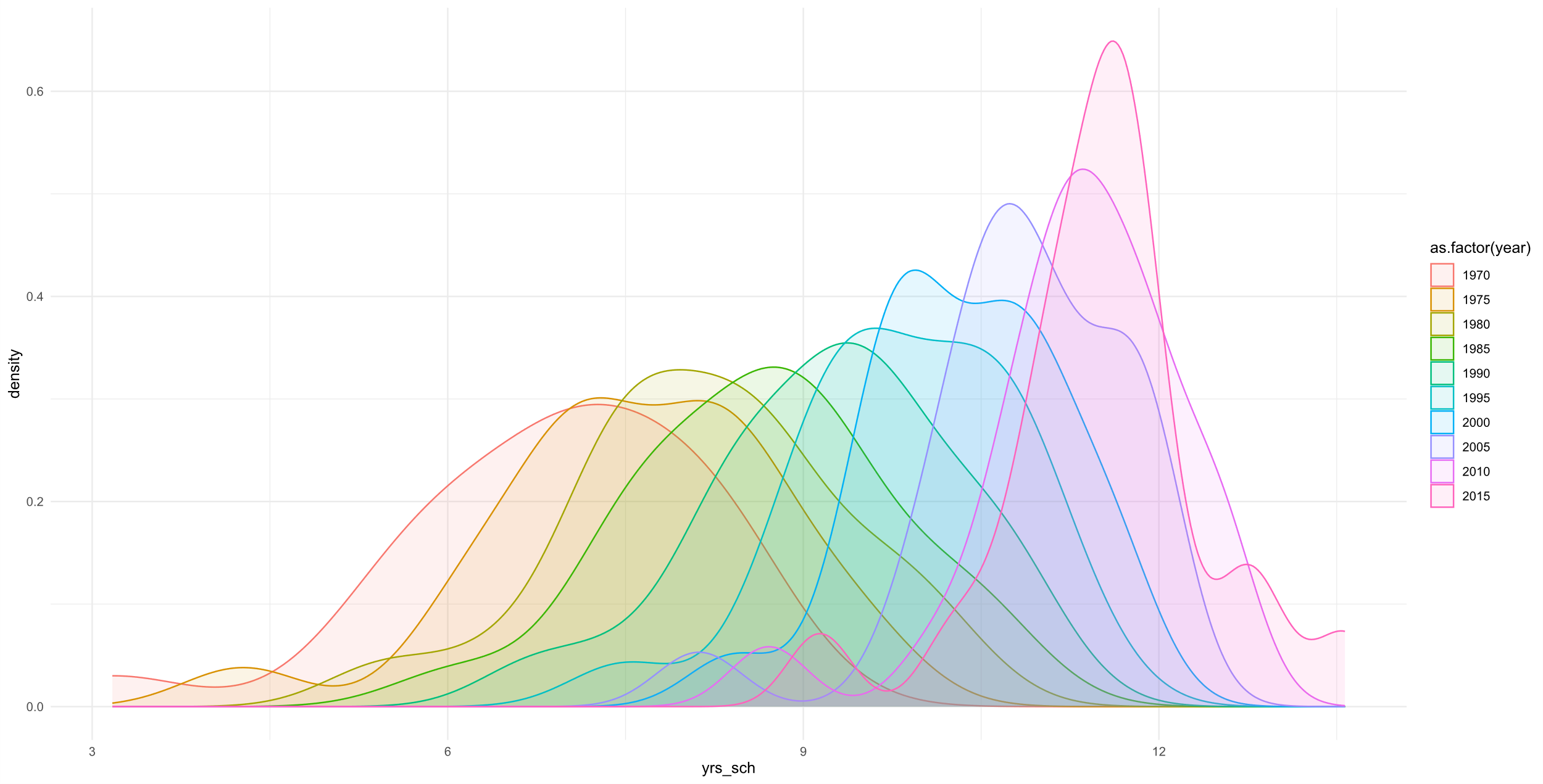
# Chart 2: Years of Education against Log of GDP Output per Capita

# Chart 3: Years of Education against GDP Output per Capita

Chart 4: Years of Education Against Log of GDP Output per Capita in 2015

Chart 5: GDP Output per Capita Over Time by Country (All?)

# Chart 6: Distribution of Years of Education Over Time



# Chart 7: Cross-Correlation Matrix for All Potential Model Variables

