University of Nottingham
ECON 4028 UNUK: Economic Data Analysis

Do the United States' government need to develop measures that elevate the level of education quality to aid long-term economic growth?

Lecturers: Sourafel Girma & Marit Hinnosaar
Student ID: 14304274
Module Code: ECON4028
Module Title: Economic Data Analysis
Word Count: 1802

Abstract

Following news that education standards and economic growth in the United States of America have significantly decreased in the last 5 years, this paper assesses whether essential plans to develop the quality of education delivered to students in the United States could also become an integral component of plans for economic recovery, following the effects of the COVID-19 pandemic. This is done using a restricted maximum likelihood (REML) approach to estimate parameters of a linear-mixed effects model designed to investigate the significance of education on the economic growth of OECD countries. This study finds that education quality has a significant positive relationship with annual GDP growth rate per capita, therefore implying that improving education quality will significantly contribute to long-term plans for economic recovery.

Contents

1 Introduction

Worldwide responses to the COVID-19 pandemic have had disastrous economic impacts (J Emmerling et al., 2021). The IMF have forecast global economic growth in 2022 at 4.9%, a worrying depletion from the projection of 6.0% in 2021. This pattern is reflected in SP Global's predictions that, for the United States, economic recovery will also show reduced rates with GDP growth forecast to be just 2.3% in 2024, in comparison to 5.5% growth seen in 2021. Governments all over the world are now focussing efforts to maximise economic recovery and much has been published arguing that prioritising areas such employment, healthcare and climate are essential components to drive the recovery (Kroner et al., 2021; Furman J et al., 2020; O'Connor CM et al., 2020). However, there have been few publications that suggest the development of measures to improve educational systems, despite the obvious disruptions to schooling. This is likely rooted in the fact that the effects of changes to education are long term whilst costs of funding are immediate (Dickens W et al., 2006) and so the political system is often biased against making such changes, especially in the current climate.

In 2015 the Organization for Economic Cooperation and Development (OECD) published global rankings of different country's student's abilities in maths, reading and science. The table placed the United States of America at 35th, 11 points below their previous position and 20 points below the average OECD score. These worrying statistics highlight a necessity to improve the country's education system. In this paper, I investigate the statistical significance of education on economic growth. The results of which should indicate whether much needed improvements to education could also have a larger impact on the country's long-term recovery from the economic crisis created by the COVID-19 pandemic

2 Literature Review

Studies examining the main contributors to economic growth have always recognised the significant influence of education (Solow R, 1957). Such work was also shown by Denison E (1985) who estimated that, between 1929 and 1989, increased education had a significant positive economic affect. While Jorgenson D et al., (2000) attributed 8.7% of total growth between 1959 and 1998 to education. *Afzal M et al., (201*0) showed that the benefits of increased education on the economy are multifaceted. Education generates economic growth, reduces poverty, and creates an economic environment that attracts investment. There are, however, some publications which contradict the notion of a positive relationship between education and economic growth. Wolf A, (2004) argued that this idea has weak foundations and suggests that as a country generates wealth it's people seek more education for their children but there is no clear evidence that countries with increased spending on education has economic effects. This contradiction is also supported by Islam N (1995) who, after controlling for fixed effects, found human capital had insignificant effects on economic growth.

It is imperative to notice a common theme throughout these studies and those summarised in Table 1. These studies do not consider education quality, they focus solely on average years of schooling and attendance records to represent education quantity. Recent papers by M Delgado et al., (2014) and Hanushek E & Woessmann L (2007) are examples of a new phase of research investigating the links between the quality of education and economic growth. M Delgado et al., (2014) reported that quality of education has a significant positive impact on economic development and that the quantity of education's contribution was insignificant. These findings concur with those in Hanushek E & Woessmann L, (2007)'s work, who also add that such findings are a strong indicator that the quality of education delivered must be improved. Here, I follow M Delgado et al., (2014) and Hanushek E & Woessmann L, (2007), thereby providing insight into the importance for policies to develop the quality of education.

Table 1. A summary of Recent Literature Analysing the Impact of Education on Economic Growth.

Paper	Method of Analysis	Conclusion
Maneejuk & Yamaka, 2021	Time series kink regression and panel kink regression	Shows that education has significant positive affect on economic growth
Benos & Zotou, 2014	Metaregression on 57 studies	Shows that the genuine impact of education varies from depending on factors such as the measurement for education quality and model specification
Afzal et al., 2010	OLS linear regression and ARDL approach to cointegration	Finds cointegration between education and economic growth, suggests a direct correlation between education and economic growth in both the long-run and short-run
Henderson, 2010	ARDL approach to cointegration, linear regression,	Finds cointegration between school education and economic growth, also reports a direct relationship between education and economic growth in Pakistan
Hanushek & Woessmann, 2007	-	Reports a strong correlation between education and economic growth, adding that more should be done to develop the quality of education offered
Dickens et al., 2006	OLS Cross section	Finds increases in education, particularly early education, yield substantial gains in GDP and economic growth
Babatunde & Adefabi, 2005	Uses Johansen Cointegration techniques and vector error correction methodology	Establishes that a well-educated labour force significantly influences economic growth
Sala-i-Martin et al., 2004	Bayesian Averaging, Classical Estimates	Find primary school education has a positive effect on economic development, no significant relationship with higher education

3 Data and Analysis

3.1 Data Sources

In this paper, I estimate the impact of various economic and educational variables (as described in **Table 2**) on the average annual GDP growth rate per capita – which is used as the measure for economic growth.

The data for the economic variables and 'education quantity' used in this study were retrieved from the Penn World Table 10.0 (PWT 10.0). The data that represents education quality in this paper was taken from Altinock, Angrist and Patrinos' (2018) 'Global Dataset on Education Quality', which averages pooled test results from Roser, Nagdy & Ortiz-Ospina's (2013) on the 'Our World in Data' dataset.

When merging the data, using a many-to-one match merge option, I aggregated the yearly data obtained from the PWT 10.0 into 5-year periods, as the 'Global Dataset on Education Quality' averages test results for 5-year periods. For the natural log of the initial real GDP per capita and the education quantity variables, where aggregation isn't possible, I used each period's initial values. This study also filtered for data from countries that are members of the OECD. The final dataset is hierarchical and unbalanced, containing data from 37 countries from 1970 to 2015.

Table 2. Brief Description of Variables used in study's regression (See **Appendix 6.1** for a more detailed description of the variables).

Variable	Symbol	Description	Sourced From:
Average Annual GDP per capita growth rate	y	Average growth rate over 5-year period	PennWorldTable 10.0
Average Population growth rate	pop	Average growth rate over 5-year period	PennWorldTable 10.0
Log of Initial Real GDP per capita	$log(GDP_{pc})$	The logarithm of each period's initial real GDP per capita	PennWorldTable 10.0
Average Share of Gross Capital Information	$share$	Average capital formation at current PPPs over 5-year period	PennWorldTable 10.0
Education Quantity	$edquan$	Proxied by the average Human Capital Index (average years in education) per person over 5-year period	PennWorldTable 10.0
Education Quality	$edquan$	Average pooled initial test score for each 5-year period	Global Dataset on Education Quality (2018)

3.2 Exploratory Data Analysis

Variation Inflation Factor (VIF) tests were executed, as per Long J et al., (2018), to estimate the extent to which multicollinearity inflates the variance of the coefficients in the regression as the nature of the dataset used implies there may be some multicollinearity. Note that we should only be concerned where VIF is greater than 5 (Farag A, 2016; Glenn S, 2015). **Figure 2** shows that the VIF for each variable was less than 5, with a mean VIF of 1.69**,** indicating that on average the variance of the variables is 69% greater than as would be expected without any multicollinearity.

The partial correlations between the variables were investigated to control for possible cofounders, see **Table 3.** The results show that *log(*GDP_{pc}*)*, *edquan* and *share* are all significantly positively correlated. The significantly negative correlation between education quality and population growth supports common notion that countries with developed education systems have lower fertility rates (Di- Marcantonio et al., 2014).

Table 3. Summary of the partial correlations between the independent variables of the study. Values from output in **Figure 3.**

	$log(GDP_{pc})$	pop	$share$	$edquan$	$edqual$
$log(GDP_{pc})$	1.00	**0.1756	***0.2358	***0.4561	**0.1542
pop	**0.1756	1.00	***0.2458	0.0334	***-0.3341
$share$	***0.2358	***0.2458	1.00	***-0.2695	***0.3752
$eduquan$	***0.4561	0.0334	***-0.2695	1.00	***0.4295
$edqual$	*0.1542	***-0.3341	***0.3752	***0.4295	1.00
Significance Codes:	*0.05 **0.01 ***0.001				

3.3 Data Diagnostics

In this study, as per (Wooldridge J, 2013), I assume the data collected across countries is independent but the observations within the same country are not independently distributed across time. This could feasibly result in the presence of heteroscedasticity and serial correlation within panels.

As per (Di- Marcantonio et al., 2014), I executed tested for heteroscedasticity using a modified Wald test and to test for autocorrelation I performed a Wooldridge test (see **Table 4** for summary of all diagnostic tests). The test statistics and small p values yielded strongly imply we can reject the null hypothesis that there is homoscedasticity and no first order correlation within the data. As a result, I proceed to use sandwich estimators of variance – clustered at country level, which facilitate intragroup correlation (Hardin JW, 2003). The presence of heteroscedasticity and autocorrelation would yield biased OLS estimations as they clearly violate OLS assumptions.

Table 4 also shows the results of the test for random effects. The small p-value obtained from the Breusch-Pagan LM test for random effects allow the rejection of the null hypothesis that country-specific variance components is equal to zero. The test results imply that there is significant random effect in the data. Consequently, as per (Park HM, 2011), these results imply that I should use a random effects model, rather than pooled OLS.

Table 4. Results from the diagnostic tests performed on the dataset.

Diagnostic	H_0	Test-statistic	P-value
Modified Wald test for heteroscedasticity	$H_0: \sigma_i^2 = \sigma_{i,j}^2 \ \forall i,j$	4482.94	0.0000
Wooldridge Test for autocorrelation	$H_0: No\ First - Order\ Correlation$	35.735	0.0000
Breusch Pagan LM test for random effects	$H_0: Var(\omega_i) = 0$	12.92	0.0002
Cluster robust Hausman test	$H_0: No\ Systematic\ Diff. in\ coeff.$	11.24	0.0468

4 Empirical Strategy

4. 1 Model 1

The initial model used in this study, estimated with a restricted maximum likelihood (REML) approach, is a linear-mixed effects model:

$$y_{i,j} = \beta_0 + \beta_1 pop_{i,j} + \beta_2 log(GDP_{pc})_{i,j} + \beta_3 share_{i,j} + \beta_4 edquan_{i,j} + \beta_5 edqual_{i,j} + \omega_{i,j} + \varepsilon_{i,j}$$

i = 1, … ,37; j = 1970, 1975, …2015

- $\omega_{i,j}$ – denotes random effect for country
- Assume $\varepsilon_{i,j}$ and $\omega_{i,j}$ are identically and independently distributed and follow normal distribution
- As per Park HM, 2011, country effect is assumed to be uncorrelated
- i denotes the country, j denotes the observed period

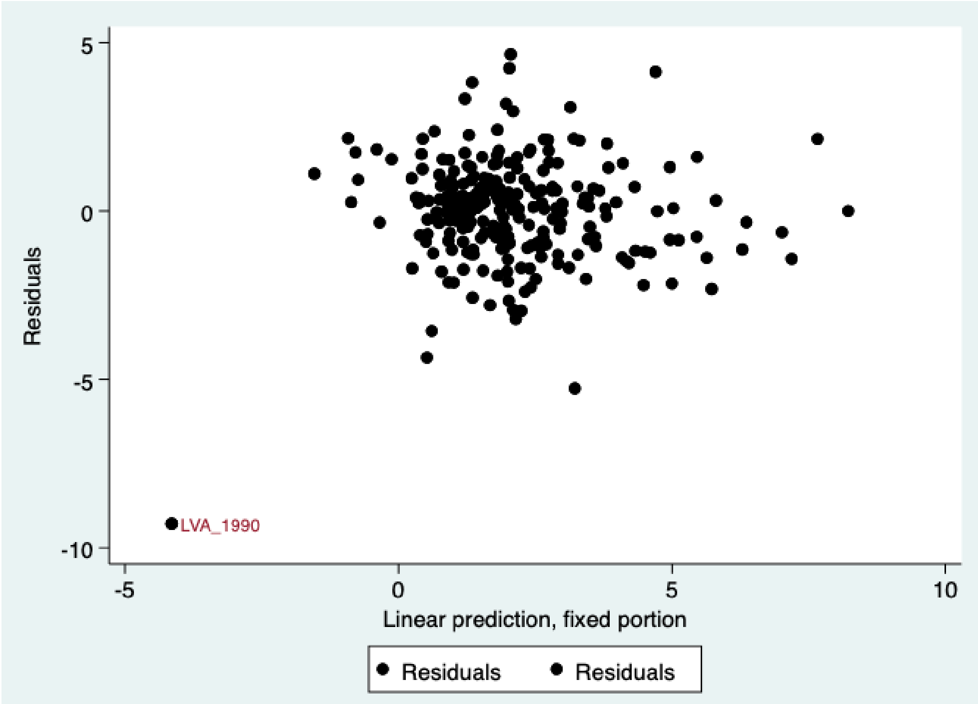
4.2 Model 1 – Results

Figure 9 shows that the estimates obtained from model 1 for both education variables have a positive relationship with the dependent variable, economic growth. The only variable estimated to have a significant impact on the dependent variable, at the 0.1 level, is education quality. The results portray that the dependent variable with increase by 0.019294 units for each point the measures for education quality increase by.
Our results also support convergence theory, that economic growth in poorer countries is quicker than in richer ones, by the estimated negative relationship between the dependent variable and the log of GDP per capita.

4.3 Model 1 – Diagnostics

As per Gelman A & Hill J (2006)**,** checking for residual normality by generating a visual of the residuals plotted against fitted values is sufficient, as residual normality does not affect our parameter estimates (see **Figure 1**).

Figure 1. Scatter plot of residuals plotted against fitted values, as per Gelman & Hill (2006)



Here we see that the residuals show no obvious trends, and there is one extreme outlier: LVA_1990. Model 1 is therefore re-estimated after the elimination of data for Latvia 1990, see **Figure 10**. Such modification reduced the goodness of fit (measured by the interclass correlation coefficient) and showed no significant improvement to the model that could justify the permanent elimination of the outlier.

4.4 Model 2 – Random Slope

To further test model 1, a random slop was added to the variable for education quality, creating Model 2:

$$y_{i,j} = \beta_0 + \beta_1 pop_{i,j} + \beta_2 log(GDP_{pc})_{i,j} + \beta_3 share_{i,j} + \beta_4 edquan_{i,j} + (\beta_5 + \beta_{5,i}) edqual_{i,j} + \omega_{i,j} + \varepsilon_{i,j}$$

i = 1, … ,37; j = 1970, 1975, …2015

A comparison using Akaike's information criterion (AIC), **Figure 12**, show that there is not significant AIC difference between model 2 and model 1 (945.176 for model 1 and 973.3477 model 2). Therefore, the results from model 1 will be used for inference.

5 Conclusion

Whilst the study is potentially limited by being an unpaired panel dataset, the results of this study concur with those found by other recent studies (Maneejuk & Yamaka, 2021) and (Delgado M et al., 2014). I found that education quality and quantity are positively correlated with economic growth, but education quality is the only significant variable at the 0.1 level.

More work should be done to investigate the importance of education quality to overall economic development of countries to build support for the argument, to an extent that governments must focus on developing adequate educational policy. Maneejuk & Yamaka (2021) also report the importance of education for economic development, however they add that higher education yields the greatest economic contribution. This should create a new avenue of investigation where more should be done to compare the impacts of different levels of education, which could affect how improved educational policies are developed.

The results from this study and the concurring past publications show that the quality of education delivered to a country's students has a significant effect on that country's economic development. I therefore conclude that the need for measures to improve education quality in the United States is necessary, not just for the improvement of their children's immediate education standards but also for long term economic growth.

References

Afzal M et al., (2010). *Relationship between school education and economic growth in Pakistan: ARDL bounds testing approach to cointegration.* Pakistan Economic and Social Review. 48(1), 39-60.

Delgado M, Henderson D & Parmeter C. (2014). *Does education matter for economic growth?* Oxford Bulletin of Economics and Statistics. 76(3), 345-359.

Denison E. 1985. *Trends in American Economic Grown, 1929-1982*. Brookings.

Dickens W, Sawhill I & Tebbs J. (2006). *The effects of investing in early education on economic growth.* Washington DC: The Brookings Institution.

Di-Marcantonio F et al., (2014). *Determinants of food production in sub-Saharan Africa: the impact of policy, market access and governance.* European Association of Agricultural Economists.

Emmerling J et al. (2021). *Will the economic impact of COVID-19 persist? Prognosis from 21st century pandemics.* IMF Working Papers. International Monetary Fund.

Farag A. (2016). *Empirical study on the spillover effect of FDI I the Egyptian and polish manufacturing sector.* International journal of business and management. 11(8), 1833-8119.

Furman J et al., (2020). *Promoting economic recovery after COVID-19.* The Aspen Institute – Economic Strategy Group.

Gelman A & Hill J. (2006). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Hanushek E & Woessmann L. (2007). *The role of education quality for economic growth.* Policy Research Working Paper. No. 4122.

Hardin JW. (2003). *The sandwich estimate of variance.* Emerald Group Publishing Limited, Bingley. 45-73.

Islam N. (1995). *Growth Empirics: A panel data approach*. The Quarterly Journal of Economics. 110(4) 1127-70

Jorgenson D et al., (2000). *Raising the speed limit: US economic growth in the information age.* Brookings Papers on Economic Activity. 1: 125-235.

Kroner et al., (2021). *COVID-era Policies and economic recovery plans: are governments building back better for protected and conserved areas?* Parks (47), 135-147.

Long J et al., (2018). *Variation inflation factor-based regression modelling of anthropometric measures and temporal-spital performance: modelling approach and implications for clinical utility.* Clinical Biomechanics. 51, 51-57.

Maneejuk & Yamaka. (2021). *The impact of higher education on economic growth in ASEAN-5 countries.* Sustainability. 13, 520-548.

O'Connor CM et al., (2020). *Economic recovery after the COVID-19 pandemic: resuming elective orthopedic surgery and total joint arthroplasty.* The Journal of Arthroplasty. 35 (7), S32-S36.

Park HM. (2011). *Practical guides to data modelling: a step-by-step analysis using stata.* Tutorial Working Paper. International University of Japan.

Solow R. (1957). *Technical change and the aggregate production function.* Review of Economic and Statistics. 38: 312-320.

Wolf A. 2004. *Education and economic performance: simplistic theories and their policy consequences.* Oxford Review of Economic Policy. 20(2), 315-333.

Wooldridge J. (2013). *Introductory Econometrics*. Cengage Learning. (5).

6 Appendix

6.1 Detailed Variables Description

Education Quality

Altinock, Angrist and Patrinos (2018) produced the 'Global Dataset on Education Quality' by collecting test scores from 163 countries from 1965 to 2015. Most standardised tests, such as Programme for International Student Assessment (PISA) tests, are new, basic, and uncoordinated – which can lead to many cases of overlapping. To overcome the limitation that is an absence of universal testing, they instead collected the individual tests and applied them to a consistent scale that enabled comparison, then pooling them together.

Education Quantity

The Penn World Table 10.0 was used to collect the other variables in the model in this study. Education quantity in this study is proxied by the average human capital per person index for each 5-year period. The human capital index is mostly obtained, in the Penn World Table 10.0, from Barro and Lee's (2013) 'Educational Attainment Data'. This contains data for each country and education level that describes the percentage of the population that do not receive education, percentage of the population that have completed primary, secondary & higher education, and the average number of years that are required to complete schooling (for each level of education) for each individual country.

Other Variables:

y: Our measurement for economic growth was calculated as follows, where Xt is the annual real GDP per capita:

$$y_t = \left[\left(\frac{X_t}{X_{t-1}}\right) - 1\right] \cdot \ 100$$

Yt is then averaged for each 5-year period.

pop: 5-year average of the annual population growth rate data obtained from the Penn World Table 10.0.

$\log(GDP_{pc})$: Penn World Table 10.0's data on the initial real GDPpc was divided by *'pop'* creating the variable to be logged.

$share$: The average share of gross capital formation at current PPPs for each 5-year period. Data sourced from the Penn World Table 10.0

6.2 Figures (STATA Output)

Figure 2: VIF Results

```
. *** VIF ***

. collin y logGDP_pc share p edquan edqual
(obs=251)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
         y      1.40    1.18    0.7123      0.2877
 logGDP_pc      2.16    1.47    0.4629      0.5371
     share      1.45    1.20    0.6916      0.3084
         p      1.19    1.09    0.8371      0.1629
    edquan      1.88    1.37    0.5316      0.4684
    edqual      2.08    1.44    0.4808      0.5192
----------------------------------------------------
  Mean VIF      1.69

                           Cond
        Eigenval          Index
---------------------------------
    1     5.9463          1.0000
    2     0.5927          3.1675
    3     0.4156          3.7828
    4     0.0305         13.9561
    5     0.0093         25.3156
    6     0.0049         34.8293
    7     0.0007         91.2090
---------------------------------
 Condition Number         91.2090
 Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)
 Det(correlation matrix)    0.2125
```

Figure 3: Partial Correlations

```
. pcorr logGDP_pc share p equan equal
(obs=251)

Partial and semipartial correlations of logGDP_pc with

             Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    corr.         corr.      corr.^2       corr.^2          value
------------+-----------------------------------------------------------------
      share |   0.2358        0.1836       0.0556        0.0337         0.0002
          p |   0.1756        0.1350       0.0308        0.0182         0.0056
      equan |   0.4561        0.3878       0.2080        0.1504         0.0000
      equal |   0.1542        0.1181       0.0238        0.0140         0.0151

. pcorr share logGDP_pc p equan equal
(obs=251)

Partial and semipartial correlations of share with

             Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    corr.         corr.      corr.^2       corr.^2          value
------------+-----------------------------------------------------------------
  logGDP_pc |   0.2358        0.2086       0.0556        0.0435         0.0002
          p |   0.2458        0.2180       0.0604        0.0475         0.0001
      equan |  -0.2695       -0.2406       0.0726        0.0579         0.0000
      equal |   0.3752        0.3480       0.1408        0.1211         0.0000

. pcorr p logGDP_pc share edquan edqual
(obs=251)

Partial and semipartial correlations of p with

             Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    corr.         corr.      corr.^2       corr.^2          value
------------+-----------------------------------------------------------------
  logGDP_pc |   0.1756        0.1633       0.0308        0.0267         0.0056
      share |   0.2458        0.2321       0.0604        0.0539         0.0001
      equan |   0.0334        0.0306       0.0011        0.0009         0.6005
      equal |  -0.3341       -0.3245       0.1116        0.1053         0.0000

. pcorr equan logGDP_pc share p equal
(obs=251)

. pcorr equan logGDP_pc share p equal
(obs=251)

Partial and semipartial correlations of equan with

             Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    corr.         corr.      corr.^2       corr.^2          value
------------+-----------------------------------------------------------------
  logGDP_pc |   0.4561        0.3737       0.2080        0.1396         0.0000
      share |  -0.2695       -0.2040       0.0726        0.0416         0.0000
          p |   0.0334        0.0244       0.0011        0.0006         0.6005
      equal |   0.4295        0.3468       0.1845        0.1203         0.0000

. pcorr edqual logGDP_pc share p edquan
(obs=251)

Partial and semipartial correlations of equal with

             Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    corr.         corr.      corr.^2       corr.^2          value
------------+-----------------------------------------------------------------
  logGDP_pc |   0.1542        0.1138       0.0238        0.0130         0.0151
      share |   0.3752        0.2951       0.1408        0.0871         0.0000
          p |  -0.3341       -0.2584       0.1116        0.0668         0.0000
      equan |   0.4295        0.3467       0.1845        0.1202         0.0000
```

Figure 4: Breusch Pagan LM Test for RE

```
. *** B-P LM Test For RE ***

. xtreg y logGDP_pc share p edquan edqual

Random-effects GLS regression                   Number of obs     =        251
Group variable: countrycode2                    Number of groups  =         37

R-squared:                                      Obs per group:
     Within  = 0.3839                                         min =          3
     Between = 0.3236                                         avg =        6.8
     Overall = 0.2856                                         max =         10

                                                Wald chi2(5)      =     119.15
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   logGDP_pc |  -3.061171   .4034232    -7.59   0.000    -3.851866   -2.270476
       share |   12.44578   2.568633     4.85   0.000     7.411355    17.48021
           p |     .01707   .2145874     0.08   0.937    -.4035136    .4376535
      edquan |   .1542945    .520952     0.30   0.767    -.8667527    1.175342
      edqual |   .0189149   .0029951     6.32   0.000     .0130446    .0247852
       _cons |   20.51578   3.179243     6.45   0.000     14.28458    26.74698
-------------+----------------------------------------------------------------
     sigma_u |  .81134107
     sigma_e |  1.5584393
         rho |  .21324028   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        y[countrycode2,t] = Xb + u[countrycode2] + e[countrycode2,t]

        Estimated results:
                         |       Var     SD = sqrt(Var)
                ---------+-----------------------------
                       y |   4.631199       2.152022
                       e |   2.428733       1.558439
                       u |   .6582743       .8113411

        Test: Var(u) = 0
                             chibar2(01) =    12.29
                          Prob > chibar2 =   0.0002
```

Figure 5: Woolridge Test

```
. *** Wooldridge Test for Autocorrelation ***

. xtserial y logGDP_pc share p edquan edqual, output

Linear regression                               Number of obs     =        211
                                                F(5, 36)          =      29.50
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4858
                                                Root MSE          =     1.8116

                          (Std. err. adjusted for 37 clusters in countrycode2)
------------------------------------------------------------------------------
             |               Robust
         D.y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   logGDP_pc |
         D1. |  -13.17043   1.386758    -9.50   0.000    -15.98291   -10.35796
             |
       share |
         D1. |   20.04511    4.28122     4.68   0.000     11.36239    28.72783
             |
           p |
         D1. |  -.1427594   .3987972    -0.36   0.722    -.9515577    .6660388
             |
      edquan |
         D1. |   11.56344   2.122626     5.45   0.000     7.258553    15.86832
             |
      edqual |
         D1. |   .0167718   .0053908     3.11   0.004     .0058387     .027705
------------------------------------------------------------------------------

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,      36) =     35.754
           Prob > F =      0.0000
```

Figure 6: Wald Test

```
. *** Wald Test ***

. quietly xtreg logGDP_pc share p edquan edqual , fe vce(cluster countrycode2)
d
. xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (37)  =    4482.94
Prob>chi2 =      0.0000
```

Figure 6: Hausman Test

```
. *** Hausman Test ***

. quietly xtreg y logGDP_pc share p edquan edqual , fe vce(cluster countrycode2)

. estimates store fixed2

. quietly xtreg y logGDP_pc share p edquan edqual , re vce(cluster countrycode2)

. estimates store random2

. rhausman fixed2 random2, cluster
bootstrap in progress
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.................................................. 50
.................................................. 100
---------------------------------------------------------------------------
Cluster-Robust Hausman Test
(based on 100 bootstrap repetitions)

b1: obtained from xtreg y logGDP_pc share p edquan edqual , fe vce(cluster countrycode2)
b2: obtained from xtreg y logGDP_pc share p edquan edqual , re vce(cluster countrycode2)

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b1-b2)' * [V_bootstrapped(b1-b2)]^(-1) * (b1-b2)
                          =       11.24
                Prob>chi2 =      0.0468
```

Figure 8: ICC

```
. *** ICC ***

. quietly xtmixed  y logGDP_pc share p edquan edqual||countrycode2: , vce(cluster
countrycode2)

. estat icc

Residual intraclass correlation

------------------------------------------------------------------------------
                       Level |        ICC   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
                countrycode2 |   .2456265   .1391397      .0695361    .5865414
------------------------------------------------------------------------------
```

Figure 9: Model 1 Estimates

```
. *** Model 1 ***

. xtmixed y logGDP_pc share p edquan edqual ||countrycode2: , vce(cluster countrycode2)

Performing EM optimisation:

Performing gradient-based optimisation:

Iteration 0:   log pseudolikelihood = -497.43781
Iteration 1:   log pseudolikelihood =  -497.4378

Computing standard errors:

Mixed-effects regression                        Number of obs     =        251
Group variable: countrycode2                    Number of groups  =         37
                                                Obs per group:
                                                              min =          3
                                                              avg =        6.8
                                                              max =         10
                                                Wald chi2(5)      =      89.70
Log pseudolikelihood =  -497.4378               Prob > chi2       =     0.0000

                          (Std. err. adjusted for 37 clusters in countrycode2)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   logGDP_pc |  -3.155888   .6622399    -4.77   0.000    -4.453854   -1.857921
       share |    12.7528   3.395719     3.76   0.000     6.097313    19.40829
           p |   .0029874   .2672515     0.01   0.991    -.5208159    .5267908
      edquan |   .2067591   .6924966     0.30   0.765    -1.150509    1.564027
      edqual |    .019294   .0111077     1.74   0.082    -.0024767    .0410646
       _cons |    21.0689    3.12584     6.74   0.000     14.94237    27.19544
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects parameters  |   Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
countrycode2: Identity       |
                   sd(_cons) |   .9211572   .3030725      .4833676    1.755456
-----------------------------+------------------------------------------------
                sd(Residual) |   1.614319   .1827485      1.293092    2.015345
------------------------------------------------------------------------------
```

Figure 10: Model 1 Estimated without LVA_1990 Outlier

```
. *** Running Model 1 Without LVA_1990 Outlier ***

. drop if re1 == 1
(1 observation deleted)

. xtmixed y logGDP_pc share p edquan edqual||countrycode2: , vce(cluster countrycode2)

Performing EM optimisation:

Performing gradient-based optimisation:

Iteration 0:   log pseudolikelihood =   -464.588
Iteration 1:   log pseudolikelihood =   -464.588

Computing standard errors:

Mixed-effects regression                        Number of obs     =        250
Group variable: countrycode2                    Number of groups  =         37
                                                Obs per group:
                                                              min =          3
                                                              avg =        6.8
                                                              max =         10
                                                Wald chi2(5)      =     110.40
Log pseudolikelihood =   -464.588               Prob > chi2       =     0.0000

                           (Std. err. adjusted for 37 clusters in countrycode2)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   logGDP_pc |  -2.553458   .4209848    -6.07   0.000    -3.378573   -1.728343
       share |   13.45207   2.812111     4.78   0.000     7.940437    18.96371
           p |  -.3316205   .1703759    -1.95   0.052    -.6655512    .0023102
      edquan |   .5756465   .5475373     1.05   0.293    -.4975069      1.6488
      edqual |   .0044095   .0029558     1.49   0.136    -.0013838    .0102028
       _cons |   21.15518   3.338498     6.34   0.000     14.61184    27.69851
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |                Robust
  Random-effects parameters  |   Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
countrycode2: Identity       |
                   sd(_cons) |   .8102155   .2214056      .4742363    1.384224
-----------------------------+------------------------------------------------
                sd(Residual) |    1.42741   .1056294      1.234694    1.650207
------------------------------------------------------------------------------
```

Figure 11: Estimation of model with Random Slopes

```
. *** Random Slopes ***

. xtmixed  y logGDP_pc share p edquan edqual||countrycode2: , vce(cluster countrycode2)

Performing EM optimisation:

Performing gradient-based optimisation:

Iteration 0:   log pseudolikelihood =   -464.588
Iteration 1:   log pseudolikelihood =   -464.588

Computing standard errors:

Mixed-effects regression                        Number of obs     =        250
Group variable: countrycode2                    Number of groups  =         37
                                                Obs per group:
                                                              min =          3
                                                              avg =        6.8
                                                              max =         10
                                                Wald chi2(5)      =     110.40
Log pseudolikelihood =   -464.588               Prob > chi2       =     0.0000

                           (Std. err. adjusted for 37 clusters in countrycode2)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
   logGDP_pc |  -2.553458   .4209848    -6.07   0.000    -3.378573   -1.728343
       share |   13.45207   2.812111     4.78   0.000     7.940437    18.96371
           p |  -.3316205   .1703759    -1.95   0.052    -.6655512    .0023102
      edquan |   .5756465   .5475373     1.05   0.293    -.4975069      1.6488
      edqual |   .0044095   .0029558     1.49   0.136    -.0013838    .0102028
       _cons |   21.15518   3.338498     6.34   0.000     14.61184    27.69851
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
  Random-effects parameters  |   Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
countrycode2: Identity       |
                   sd(_cons) |   .8102155   .2214056      .4742363    1.384224
-----------------------------+------------------------------------------------
                sd(Residual) |    1.42741   .1056294      1.234694    1.650207
------------------------------------------------------------------------------
```

Figure 12: AIC

```
. *** AIC ***

. quietly xtmixed  y logGDP_pc share p edquan edqual||countrycode2: , vce(cluster
countrycode2)

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df       AIC        BIC
-------------+---------------------------------------------------------------
           . |        250          .   -464.588       8   945.176   973.3477
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.
```

6.3 STATA Code

```
. log using "/Users/josephthomas/Documents/Masters/Economic Data
Analysis/EDA/log.smcl"
------------------------------------------------------------------------
----------------------------
      name:  <EDA_log>
       log:  /Users/josephthomas/Documents/Masters/Economic Data
Analysis/EDA/log.smcl
  log type:  smcl
 opened on:  10 Jan 2022, 10:02:53


. cd "/Users/josephthomas/Documents/Masters/Economic Data Analysis/EDA"
/Users/josephthomas/Documents/Masters/Economic Data Analysis/EDA

. use "/Users/josephthomas/Documents/Masters/Economic Data
Analysis/EDA/pennworld10.dta"

. *Variables “country” “countrycode” “year” “rgdpna” “pop” “hc” “csh_i”
all kept manually through data
>editor, all other variables deleted*

. rename countrycode cc

. rename year yr

. rename rgdpna rg

. rename pop p

. rename csh_i c

. drop if p ==.
(2,411 observations deleted)

. drop if hc ==.
(1,762 observations deleted)

. *** Generating ***

. generate gdp_pc = rgdpna/p

. generate loggdp_pc = log(gdp_pc)

. generate gdp_gr = ((gdp_pc[_n]/gdp_pc[_n-1])-1)*100
(145 missing values generated)

. generate period = 5*floor(year/5)

. *** Creating Averages of Variables Per Period ***

. bysort countrycode period: egenerate avgdp_pc_gr = mean(gdp_gr)
(8 missing values generated)

. bysort countrycode period: generate gr_p = 100*((p[_n]/p[_n-1])-1)
(1,737 missing values generated)
```

```
. bysort countrycode period: egenerate avp_gr = mean(gr_p)
(8 missing values generated)

. bysort countrycode period: egenerate avloggdp_pc = mean(loggdp_pc)

. bysort countrycode period: egenerate avhc = mean(hc)

. bysort countrycode period: egenerate avcsh_i = mean(csh_i)

. save pennworld10.dta, replace
file pennworld10.dta saved

.*Using AAP EducationQual data *

.
. import delimited "/Users/josephthomas/Documents/Masters/Economic Data
Analysis/EDA/AAP.csv", clea
> r
(encoding automatically selected: ISO-8859-1)
(8 vars, 59,922 obs)

. *Kept variables “code” “year” “averageharmonisedlearningou
>tcome” manually through data editor*

. rename code cc

. rename averageharmonisedlearningoutcome equal

. drop if equal ==.
(59,295 observations deleted)

. drop if cc == ""
(75 observations deleted)

. save sorted_AAP.dta
file sorted_AAP.dta saved

. . *** Combining the Two ***

. use pennworl0.dta

. merge m:1 cc yr using sorted_AAP.dta
(variable cc was str3, now str8 to accommodate using data's values)

    Result                      Number of obs
    -----------------------------------------
    Not matched                         8,157
        from master                     8,121  (_merge==1)
        from using                         36  (_merge==2)

    Matched                               516  (_merge==3)
    -----------------------------------------

. drop if equal ==.
(8,121 observations deleted)

. drop if p ==.
(36 observations deleted)
```

```
. drop _merge

. ** Variables “avgdp_pc_gr” “loggdp_pc” “avcsh_i” “avp_gr” “avhc”
“equal” “cc” “period” “yr” kept manually in data editor**

. * All countries that are “OECD” (37 in total) kept manually in data
editor according to their cc*
.
. encode cc, generate(cc2)

. xtset cc2 period, delta(5)

Panel variable: cc2 (unbalanced)
 Time variable: period, 1970 to 2015, but with gaps
         Delta: 5 units

. *Cleaning names

. rename loggdp_pc logGDP_pc

. rename avcsh_i share

. rename avp_gr p

. rename avgdp_pc_gr y

. rename avhc equan

.save ds.dta
file ds.dta saved

. *** EDA ***

. *Use collin command for collinearity test – shows variety of
measures, including VIF*

. collin y logGDP_pc share p edquan edqual
(obs=251)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
         y      1.40    1.18    0.7123      0.2877
 logGDP_pc      2.16    1.47    0.4629      0.5371
     share      1.45    1.20    0.6916      0.3084
         p      1.19    1.09    0.8371      0.1629
     equan      1.88    1.37    0.5316      0.4684
    equal      2.08    1.44    0.4808      0.5192
----------------------------------------------------
  Mean VIF      1.69

                           Cond
        Eigenval          Index
---------------------------------
    1     5.9463          1.0000
```

```
    2     0.5927          3.1675
    3     0.4156          3.7828
    4     0.0305         13.9561
    5     0.0093         25.3156
    6     0.0049         34.8293
    7     0.0007         91.2090
---------------------------------
 Condition Number        91.2090
 Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)
 Det(correlation matrix)    0.2125


. xtreg y logGDP_pc share p equan equal

Random-effects GLS regression                   Number of obs     =        251
Group variable: cc2                             Number of groups  =         37

R-squared:                                      Obs per group:
     Within  = 0.3839                                         min =          3
     Between = 0.3236                                         avg =        6.8
     Overall = 0.2856                                         max =         10

                                                Wald chi2(5)      =     119.15
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      y | Coefficient  Std. err.   z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
logGDP  | -3.061171  .4034232    -7.59   0.000    -3.851866   -2.270476
  share | 12.44578   2.568633     4.85   0.000     7.411355    17.48021
      p |   .01707   .2145874     0.08   0.937    -.4035136    .4376535
  equan |  .1542945   .520952     0.30   0.767    -.8667527    1.175342
  equal | .0189149   .0029951     6.32   0.000     .0130446    .0247852
_cons   | 20.51578   3.179243     6.45   0.000     14.28458    26.74698
-------------+----------------------------------------------------------------
sigma_u |  .81134107
sigma_e |  1.5584393
    rho |  .21324028   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        y[countrycode2,t] = Xb + u[countrycode2] + e[countrycode2,t]

        Estimated results:
                         |       Var     SD = sqrt(Var)
                ---------+-----------------------------
                       y |   4.631199       2.152022
                       e |   2.428733       1.558439
                       u |   .6582743       .8113411

        Test: Var(u) = 0
                             chibar2(01) =    12.29
                          Prob > chibar2 =   0.0002

. *Autocorr diagnostics
```

```
. xtserial y logGDP_pc share p equan equal, output

Linear regression                               Number of obs     =        211
                                                F(5, 36)          =      29.50
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4858
                                                Root MSE          =     1.8116

                              (Std. err. adjusted for 37 clusters in cc2)
------------------------------------------------------------------------------
      |               Robust
  D.y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
logGDP|
  D1. |  -13.17043   1.386758    -9.50   0.000    -15.98291   -10.35796
      |
share |
  D1. |   20.04511    4.28122     4.68   0.000     11.36239    28.72783
      |
    p |
  D1. |  -.1427594   .3987972    -0.36   0.722    -.9515577    .6660388
      |
 equan|
  D1. |   11.56344   2.122626     5.45   0.000     7.258553    15.86832
      |
 equal|
  D1. |   .0167718   .0053908     3.11   0.004     .0058387     .027705
------------------------------------------------------------------------------

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,      37) =     35.754
           Prob > F =      0.0000

. *Hetero diagnostics

. quietly xtreg logGDP_pc share p equan equal , fe vce(cluster cc2)

. xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (37)  =    4482.94
Prob>chi2  =      0.0000

.
. *Initial Model

. xtmixed y logGDP_pc share p equan equal ||cc2: , vce(cluster cc2)

Performing EM optimisation:

Performing gradient-based optimisation:
Iteration 0:   log pseudolikelihood = -497.43781
```

```
Iteration 1:   log pseudolikelihood =  -497.4378

Computing standard errors:

Mixed-effects regression                        Number of obs     =        251
Group variable: countrycode2                    Number of groups  =         37
                                                Obs per group:
                                                              min =          3
                                                              avg =        6.8
                                                              max =         10
                                                Wald chi2(5)      =      89.70
Log pseudolikelihood =  -497.4378               Prob > chi2       =     0.0000

                  (Std. err. adjusted for 37 clusters in countrycode2)
------------------------------------------------------------------------
       |               Robust
     y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------
logGDP|  -3.155888   .6622399    -4.77   0.000    -4.453854   -1.857921
share |    12.7528   3.395719     3.76   0.000     6.097313    19.40829
     p |   .0029874   .2672515     0.01   0.991    -.5208159    .5267908
equan |   .2067591   .6924966     0.30   0.765    -1.150509    1.564027
equal |    .019294   .0111077     1.74   0.082    -.0024767    .0410646
_cons |    21.0689    3.12584     6.74   0.000     14.94237    27.19544
------------------------------------------------------------------------

------------------------------------------------------------------------
Random-effects |                Robust
Parameters     |   Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------
cc2: Identity  |
     sd(_cons) |   .9211572   .3030725      .4833676    1.755456
-----------------------------+------------------------------------------
  sd(Residual) |   1.614319   .1827485      1.293092    2.015345
------------------------------------------------------------------------


 .



 . *Outliers

 . quietly xtmixed y logGDP_pc share p equan equal ||cc2: , vce(cluster
 cc2)

 . predict double xb, xb

 . predict double re, res
 . generate res1 = res<-8 | res>8

 . generate cp ="LVA_1990" if res>8
 (251 missing values generated)

 . replace cp ="LVA_1990" if re<-8
 variable cp was str1 now str8
 (1 real change made)

 .
 . scatter res xb, mcolor(black) msize(small) ylabel( ,
 angle(horizontal) nogrid) || scatter res xb if
```

```
>  res1==1,mlabel(cp) mcolor(black) msize(small) ylabel( ,
angle(horizontal) nogrid)

.
. graph save "Graph" "/Users/josephthomas/Documents/Masters/Economic
Data Analysis/EDA/Res Graph.gp
> h"
file /Users/josephthomas/Documents/Masters/Economic Data
Analysis/EDA/Res Graph.gph saved

. *** Running Model 1 after eliminating LVA_1990 ***

. drop if res1 == 1
(1 observation deleted)

. xtmixed y logGDP_pc share p equan equal||cc2: , vce(cluster cc2)

Performing EM optimisation:

Performing gradient-based optimisation:

Iteration 0:   log pseudolikelihood =   -464.588
Iteration 1:   log pseudolikelihood =   -464.588

Computing standard errors:

Mixed-effects regression                        Number of obs     =        250
Group variable: cc2                             Number of groups  =         37
                                                Obs per group:
                                                              min =          3
                                                              avg =        6.8
                                                              max =         10
                                                Wald chi2(5)      =     110.40
Log pseudolikelihood =   -464.588               Prob > chi2       =     0.0000

                                  (Std. err. adjusted for 37 clusters in cc2)
------------------------------------------------------------------------------
             |               Robust
           y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
logGDP|  -2.553458   .4209848    -6.07   0.000    -3.378573   -1.728343
share |   13.45207   2.812111     4.78   0.000     7.940437    18.96371
    p |  -.3316205   .1703759    -1.95   0.052    -.6655512    .0023102
 equan|    .5756465   .5475373     1.05   0.293    -.4975069      1.6488
 equal|    .0044095   .0029558     1.49   0.136    -.0013838    .0102028
_cons |   21.15518   3.338498     6.34   0.000     14.61184    27.69851
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |               Robust
Random-effects parame |   Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
Cc2:           Identity |
               sd(_cons) |   .8102155   .2214056      .4742363    1.384224
-----------------------------+------------------------------------------------
           sd(Residual) |    1.42741   .1056294      1.234694    1.650207
------------------------------------------------------------------------------
 .

 .

 .
```

```
 *Potential New Model

. xtmixed  y logGDP_pc share p equan equal||cc2: , vce(cluster cc2)

Performing EM optimisation:

Performing gradient-based optimisation:

Iteration 0:   log pseudolikelihood =   -464.588
Iteration 1:   log pseudolikelihood =   -464.588

Computing standard errors:

Mixed-effects regression                    Number of obs     =        250
Group variable: cc2                         Number of groups  =         37
                                            Obs per group:
                                                          min =          3
                                                          avg =        6.8
                                                          max =         10
                                            Wald chi2(5)      =     110.40
Log pseudolikelihood =   -464.588           Prob > chi2       =     0.0000

                             (Std. err. adjusted for 37 clusters in cc2)
------------------------------------------------------------------------
      |               Robust
    y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------
logGDP|  -2.553458   .4209848    -6.07   0.000    -3.378573   -1.728343
share |   13.45207   2.812111     4.78   0.000     7.940437    18.96371
    p |  -.3316205   .1703759    -1.95   0.052    -.6655512    .0023102
 equan|   .5756465   .5475373     1.05   0.293    -.4975069      1.6488
 equal|   .0044095   .0029558     1.49   0.136    -.0013838    .0102028
_cons |   21.15518   3.338498     6.34   0.000     14.61184    27.69851
------------------------------------------------------------------------

------------------------------------------------------------------------
                          |               Robust
Random-effects parameters|Estimate   std. err.     [95% conf. interval]
-----------------------------+------------------------------------------
cc2:          Identity |
             sd(_cons) |   .8102155   .2214056      .4742363    1.384224
-----------------------------+------------------------------------------
         sd(Residual) |    1.42741   .1056294      1.234694    1.650207
------------------------------------------------------------------------

.
. *Compare models

. quietly xtmixed  y logGDP_pc share p equan equal||cc2: , vce(cluster
cc2)

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------
Model |           N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------
```

```
    . |        250          .   -464.588       8    945.176   973.3477
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. log close
      name:  <EDA_log>
       log:  /Users/josephthomas/Documents/Masters/Economic Data 
Analysis/EDA/log.smcl
  log type:  smcl
 closed on:  13 Jan 2022, 12:04:17
-----------------------------------------------------------------------------
```