Examining Predictors of Gender Parity in Education

Anepska, Veronika and Krotova, Sofiia and Mamchur, Polina and Tyshchenko, Sofiia

Kyiv School of Economics

17 May 2024

Examining Predictors of Gender Parity in Education

Veronika Anepska, Sofiia Krotova, Polina Mamchur, Sofiia Tyshchenko

Kyiv School of Economics

May 17, 2024

Examining Predictors of Gender Parity in Education

May 17, 2024

Veronika Anepska	Sofiia Krotova	Polina Mamchur	Sofiia Tyshchenko
BE26	EBD26	EBD26	EBD26
vanepska@kse.org.ua	skrotova@kse.org.ua	pmamchur@kse.org.ua	styshchenko@kse.org.ua

Abstract

Gender equality is a pivotal aspect of an inclusive and just society. Equal educational opportunities for men and women are necessary for its achievement. This study analyzes the mutual influence between various social-economic factors and the gender parity index. The purpose of this paper is to identify the relationship between the gender parity index and gross primary school enrollment, annual GDP per capita growth, government expenditure on education, life expectancy, unemployment rate, group of income of the country, and its geographical region using regression analysis. Based on the results of our research, it is possible to form an idea of the further process of increasing gender equality in certain countries through the interaction with such factors as increasing primary school enrollment, improving economic conditions, and addressing regional disparities.

Key words: Education, Gender Parity Index, Regression Analysis

Introduction

Gender equality is a phenomenon, characterized by the absence of gender-based discrimination in accessing rights and opportunities, and is considered a fundamental principle for fostering a fair and inclusive society. There is no doubt that one of the most crucial aspects of maintaining gender equality is providing equal educational opportunities for both males and females across the globe. In this study, our primary objective is to investigate the potential relationships between various socio-economic factors and the gender parity index in education. Specifically, we aim to analyze the correlations between gross primary school enrollment, annual GDP per capita growth, government expenditure on education, life expectancy, unemployment rate, income, geographical region, and the gender parity index. By delving into these interconnections, we seek to unravel the complex dynamics that underpin gender equality in education and identify the most influential factors.

The study holds particular relevance due to persisting gender disparities in the countries, despite global efforts towards equality. By bringing light on the predictors of gender parity in education, our study aims to contribute to shaping proper policies and intervention strategies for governments. Ultimately, our findings have the potential to inform initiatives aimed at promoting inclusive education systems and advancing gender equality agendas worldwide. With the help of regression analysis and, in particular, R programming language and R studio application, we are able to hold our study. Taking our assumptions and theory regarding the issue into account, we will explore this topic in-depth and describe it in detail further in this paper.

Data and methods description

Data sources

All the data was collected from the World Bank (https://data.worldbank.org/). It is the organization that is committed to the research of the developing countries to help with their economic improvements. In addition, it suggests financing and advising for those countries. The bank represents itself as an organization that fights poverty by providing assistance for the country that needs it. The World Bank took the data that is used for the further description from UNESCO Institute for Statistics. It is the United Nations body that produces analytical and statistical reports for countries at different stages of development. As a common year for the dependent variable and independent variables we selected 2020 as it had enough observations, allowing us to conduct a more comprehensive research. That data is collected for both public and private schools, and it is calculated by dividing female gross enrollment ratio by male gross enrollment ratio.

Dependent variable

Our dependent variable is gender parity index, which is an indicator that reflects how equally women and men are represented in the world, how equal their opportunities, rights, choices and material possibilities are. Gender parity can be measured by the change in its gross enrollment ratio in primary and secondary education. The gender parity index is measured in the range from 0 to 1, where 0 is complete inequality and 1 and above is complete equality between men and women. As one example, a GPI between 0.97 and 1.03 indicates parity between the genders, whereas a GPI below 0.97 indicates a disparity in favor of males.

Independent variables

The independent variables for the dataset are gross primary school enrollment, annual growth of GDP per capita, government expenditure on education, life expectancy, unemployment rate, income group, and region. How each predictor is measured is described below:

1. *School enrollment, primary (% gross)* is calculated by dividing the number of students enrolled in primary education regardless of age by the population of the age group which officially corresponds to primary education, and multiplying by 100.
2. *GDP per capita growth (annual %)* indicates the rate at which the average income per person in a country is increasing or decreasing over a year. It is calculated by dividing the annual growth rate of the country's gross domestic product (GDP) by its population size and expressing the result as a percentage.
3. *Government expenditure on education, total (% of GDP)* is calculated by dividing total government expenditure for all levels of education by the GDP, and multiplying by 100.
4. *Life expectancy at birth, total (years)* is calculated in a period life table which provides a snapshot of a population's mortality pattern at a given time.
5. *Unemployment, total (% of total labor force) (modeled ILO estimate)* is the share of the labor force that is without work but available for and seeking employment.
6. *Income Group* is a classification, based on the Gross National Income per capita of the previous calendar year.
7. *Region* refers to seven geographic regions within which the data points are categorized.

Gender parity = f(gross primary school enrollment, annual growth of GDP per capita, government expenditure on education, life expectancy, unemployment rate, income, region)

Prior expectations

Our model estimates the relationship between the gender parity index, gross primary school enrollment, annual growth of GDP per capita, government expenditure on education, life expectancy, unemployment rate, income, and region. It shows the dependence of the distribution of education opportunities between men and women in accordance with changes in the independent variables, such as gross school enrollment in primary education, annual growth of GDP per capita, government expenditure on education, life expectancy, unemployment rate, type of income based on gross national income, and region.

The model includes the gross enrollment ratio in primary education because it is one of the main factors in measuring gender parity across the world. Higher enrollment ratios will indicate greater access to education for both genders, aligning with reducing gender inequality and boosting higher gender equality. In other words, when more children, regardless of their gender, are enrolled in primary and secondary education, it creates a distribution of access to education that proposes equal opportunities for both genders and, as a result, can diminish gender disparities.

Similarly, life expectancy may reflect the overall quality of life in the country and hence impact the level of gender gap in terms of education. Furthermore, economic growth enhances a country’s economic well-being and wealthier countries care more about promoting gender equality in education than poorer countries. Public spending improves access to education and hence provides more opportunities to study for females. Additionally, continents differ in their level of economic development with North America and Europe having wealthier countries while other continents including rather economically underdeveloped or developing countries. Since gender parity in education is a function of economic prosperity, the selected countries may differ in their levels of education equality across continents. Finally, income groups capture the type of economy to which our countries belong and hence directly measure the level of development of the economies included in the analysis.

Consequently, the following hypotheses can be formulated:

Hypothesis 1: The growth of the gross enrollment ratio in primary and secondary education is expected to positively affect the gender parity index.

Hypothesis 2: The growth of life expectancy is expected to positively affect the gender parity index.

Hypothesis 3: The annual growth of GDP per capita is expected to positively affect the gender parity index.

Hypothesis 4: The growth of government spending on education is expected to positively affect the gender parity index.

Hypothesis 5: Region of the country is expected to affect the gender parity, with North America and Europe having less gender disparity in education than South America, Asia or Africa.

Hypothesis 6: The income group is expected to affect the gender parity index with higher income countries having less gender parity in education than lower income countries.

The least squares regression analysis will assess the relationship between the gender parity index and primary and secondary enrollment rates. The functions will include determining the strength and direction of the relationship between the variables. We will also find out the significance of individual predictors, test hypotheses about the coefficients, and assess the overall fit of the resulting model.

The model equation

The full model equation without interactions is supposed to be written as follows:

GenderParityIndex = $\beta_0 + \beta_1 \times$ GrossSchoolEnrollment + $\beta_2 \times$ LifeExpectancy + $\beta_3 \times$ GDPperCapitaGrowth + $\beta_4 \times$ EducationExpenditure + $\beta_5 \times$ Region + $\beta_6 \times$ IncomeGroup + ε

Where:

- β_0 is the intercept (constant term).
- β_i are the coefficients for the corresponding predictor variables.
- ε is the error term (representing the difference between the actual and predicted values of the dependent variable).

Empirical Analysis and Results

Visualizing relationships and exploring the correlations

We start the analysis by exploring the relationship in our data visually (Figure 1). While looking at the plot, we notice a rather weak connection between our dependent and independent variables. Therefore, we support our preliminary conclusions by calculating correlations between variables in the dataset. As can be seen from Figure 2 and Figure 3, all correlation values are relatively low, we can say that there is almost no correlation between our variables selected for the analysis. Since the correlation among the predictors takes very low values, we expect that the problem of multicollinearity will not arise.

Initial full model

We begin our regression analysis by including all the predictors that we collected based on the theory. In doing so, we include both continents and income groups because the continent variable goes beyond capturing variation in income levels across countries and include, to some extent, cultural differences regarding gender equality. Many predictors appear to be not statistically significant but there is no problem of collinearity in this model. According to VIFs, presented in Figure 4, none of the predictors develop a high correlation.

Reduced model

After running the full model, we determine that out of the entire set of predictors only a few appear to be statically significant. We apply a backward selection method to choose our reduced model. In the end, we include only such variables, as enrollment, life expectancy, group of income, and region since their p-values turned out to be less than the threshold of 0.05.

Interactions and quadratics

By introducing interactions, we model how the impact of our key predictors depends on the continent of the country and income group. In choosing interactions among the available predictors, our main scope was to expect as much information as possible from regression analysis. This suggests that some interactions, such as the interaction between income groups and life expectancy, might not correspond to any theory. However, its inclusion improves the explanatory power of the model. The results are presented in Model 3 (Table 2). Because introducing quadratics does not have a significant impact on the results of the model, we do not include them in the final model.

Comparing the models

After selecting only statistically significant predictors for our model, we proceed to compare the full and reduced models. The p-value is greater than the significance level of 0.05, so we fail to reject the null hypothesis, suggesting that including all predictors in the full model does not significantly improve the model fit compared to including only Life expectancy in the reduced model. Hence, we can drop the statistically insignificant predictors, such as the annual growth of GDP per capita, government spending on education, and unemployment rate. As it can be observed, Model 3 appears to have the best fit among the three models, as it has the lowest residual sum of squares and the highest F-statistic, 5.2231, with a very low p-value of 0.00001726. Therefore, Model 3 seems to be the best model for explaining the variance in the gender parity index (Figure 5).

Checking whether the model meets all the assumptions for the regression analysis

Before interpreting the coefficients in our model, we proceed to check whether the model meets all the assumptions for the regression analysis. First of all, we build fitted values against residuals plot. From Figure 6 it can be observed that the assumption of linearity is met. The variance in residuals seems to be constant. However, some outliers might be present (Figure 7). In addition, the Breusch-Pagan test suggests that the variance in residuals is non constant (Figure 8). Nonetheless, the Shapiro-Wilk test shows that the residuals are normally distributed (Figure 9). The p-value for the BP statistic is less than the typical significance level of 0.05, we reject the null hypothesis that the variance of residuals is constant. We also detected some outliers, which, however, do not affect our coefficient estimates.

Transformations in order to meet the assumptions

In order to correct for non-constancy in the variance of residuals, we use the most common variance stabilized transformation which includes log-transformation of the dependent variable. The dependent variable GPI and predictor variables are being log-transformed to potentially address issues of heteroscedasticity in the relationship with other variables in the model. By logarithmizing one or more variables we can stabilize the variance of the residuals. This can lead to a more stable relationship between residuals and fitted values (Figure 10). In this case, the adjusted R-squared increased slightly after the log transformation, indicating a slightly better fit of the transformed model.

Nonetheless, the key assumptions have not been fulfilled even after the log-transformation of variables (Figures 11, 12, and 13). This can be primarily explained by the fact that our dependent variable had a limited scale by definition and a limited dispersion of values.

Final model

As the final model we select the model 4, it includes such predictors as gross enrollment in primary education, life expectancy, region, group of income, and interactions between enrollment and region, as well as between life expectancy and group of income (Table 2).

After analyzing our model, it is possible to state that in order to achieve the greatest gender equality between men and women, it is necessary to control such a variable as gross enrollment in primary education. Positive effect of enrollment suggests that expanding education opportunities for all helps to narrow down a gender gap (United Nations, n.d.). Also it is important to note the income group and the region of the country, because belonging to the regions of East North Africa and Sub-Saharan Africa, and Low income countries also affects the change in the gender parity indicator. In other words, belonging to such groups likely results in lower gender parity index values.

Equation of the final model

The overall estimated equation of the final model is the following:

$$\log(\mathit{GPI}) = -0.115766882 + 0.002357442 \times \mathit{GrossSchoolEnrollment} - 0.001358893 \times \mathit{LifeExpectancy} + 0.064475576 \times \mathit{RegionEurope\&Central\ Asia} + 0.114124095 \times \mathit{RegionLatinAmerica\&Caribbean} + 0.431420436 \times \mathit{RegionMiddleEast\&NorthAfrica} + 1.075134089 \times \mathit{RegionNorthAmerica} + 0.025432886 \times \mathit{RegionSouthAsia} + 0.563097040 \times \mathit{RegionSub\text{-}SaharanAfrica} - 3.653501068 \times \mathit{IncomeLow} - 1.174341435 \times \mathit{IncomeLowerMiddle} - 0.312233488 \times \mathit{IncomeUpperMiddle} - 0.000696467 \times \mathit{GrossSchoolEnrollment} * \mathit{RegionEurope\&CentralAsia} - 0.001197379 \times \mathit{GrossSchoolEnrollment} * \mathit{RegionLatinAmerica\&Caribbean} - 0.004443161 \times \mathit{GrossSchoolEnrollment} * \mathit{RegionMiddleEast\&NorthAfrica} - 0.011011144 \times \mathit{GrossSchoolEnrollment} * \mathit{RegionNorthAmerica} + 0.000003267 \times \mathit{GrossSchoolEnrollment} * \mathit{RegionSouthAsia} - 0.004758585 \times \mathit{GrossSchoolEnrollment} * \mathit{RegionSub\text{-}SaharanAfrica} + 0.057756179 \times \mathit{LifeExpectancy} * \mathit{IncomeLow} + 0.016275126 \times \mathit{LifeExpectancy} * \mathit{IncomeLowerMiddle} + 0.004051164 \times \mathit{LifeExpectancy} * \mathit{IncomeUpperMiddle} + \varepsilon$$

Interpretations of the coefficients

The results indicate that there exists a positive relationship between primary school enrollment rates and the Gender Parity Index in regions such as Central and South Asia, the Pacific, Europe, North America, and Latin America. This suggests that increasing enrollment rates in primary schools can contribute to the amelioration of gender inequality in the aforementioned regions. Conversely, in North and Sub-Saharan Africa and the Middle East, higher enrollment in primary schools tends to exacerbate gender disparities, thereby worsening the situation regarding inequality.

Moreover, the impact of life expectancy on gender equality exhibits variability based on a country's economic development level. In countries categorized as upper or upper-middle income, the effect is negative and relatively weak. Contrastingly, in low-income and lower-middle-income countries, the impact is stronger and positively related. This implies that enhancing life expectancy within less economically prosperous nations could serve as a catalyst for promoting gender equality in their societies.

Standardized coefficients

To estimate the comparable impacts of each independent variable, we proceed to standardize all of the coefficients. By doing so, it is possible to distinguish the most and the least influential independent variables in the final model. The final standardized equation is the following:

$$\log(GPI) = 0.509720040 \times GrossSchoolEnrollment - 0.236891859 \times LifeExpectancy + 0.835432245 \times RegionEurope\&CentralAsia + 1.088667269 \times RegionLatinAmerica\&Caribbean + 2.520195118 \times RegionMiddleEast\&NorthAfrica + 4.037813910 \times RegionNorthAmerica + 0.116352947 \times RegionSouthAsia + 4.893731179 \times RegionSub\text{-}SaharanAfrica - 21.342372274 \times IncomeLow - 12.288678338 \times IncomeLowerMiddle - 3.724115078 \times IncomeUpperMiddle - 0.903854428 \times GrossSchoolEnrollment * RegionEurope\&CentralAsia - 1.192647081 \times GrossSchoolEnrollment * RegionLatinAmerica\&Caribbean - 2.738959206 \times GrossSchoolEnrollment * RegionMiddleEast\&NorthAfrica - 4.095298305 \times GrossSchoolEnrollment * RegionNorthAmerica + 0.001689368 \times GrossSchoolEnrollment * RegionSouthAsia - 3.930119078 \times GrossSchoolEnrollment * RegionSub\text{-}SaharanAfrica +$$

$$20.385149646 \times LifeExpectancy * IncomeLow + 11.736730899 \times LifeExpectancy * IncomeLowerMiddle + 3.547507141 \times LifeExpectancy * IncomeUpperMiddle + \varepsilon$$

In terms of continuous variables, we revealed that the least influential predictor turns out to be life expectancy, whereas the most significant one is primary school enrollment. At the same time, the influence of categorical variables varies highly. Similar variability can be observed in terms of interactions (Figure 14).

Quality of the final model

The adjusted R-square value of the final model stands at 0.3318, signifying that approximately 33.18% of the variability observed in the data can be explained by the variables included in the model, which may be considered relatively modest. However, in terms of a limited set of predictors, it is quite a good indicator.

Comparing results to expectations

Our findings indicate that several of our hypotheses have received empirical support, specifically hypotheses 1, 5, and 6. These results provide compelling evidence to reject the null hypothesis in their favor, suggesting that the relationships proposed in these hypotheses are statistically significant and align with the observed data. This implies that factors such as gross enrollment ratio in primary education, income level, and regional characteristics indeed play influential roles in shaping gender parity in the educational field. Unexpectedly, life expectancy appears to negatively impact the gender parity index in some countries. In other words, the study allowed us to reveal that if life expectancy increases, the gender parity index tends to slightly decrease. However, the topic should be explored further due to the complexity of the factors influencing gender equality. For instance, with choosing different independent variables and refinement of theoretical frameworks.

Conclusion

Our final model justifies that, indeed, achieving gender equality requires a multifaceted approach. Specifically, it highlights the importance of factors such as increasing primary school enrollment, improving economic conditions, and addressing regional disparities. By focusing on these aspects, we can make significant efforts towards narrowing the gender gap in education.

Regarding policy recommendations for increasing gender equality, countries should improve their economies. Regarding Ukraine, which is currently a lower-middle-income country, it is recommended to monitor its economic level. Also, for countries where access to education at the primary and secondary levels is limited, it is necessary to increase school primary enrollment because it has the greatest impact on the GPI indicator. By focusing on these areas, we can work not only to improve gender equality in education but also to contribute to overall social development across the globe.

With the implementation of the policy recommendations, we expect some changes in our results. A good idea would be to extend the analysis to a longer period or to study trends and dynamics over time. Thus, we will be able to understand how the relationships between the variables could change during different periods, taking into account the external changes that have occurred. This will allow us to improve and effectively adapt our strategies, which will contribute to a sustainable approach to gender equality at the global level.

References

Dalpiaz, D. *Applied Statistics with R.* https://book.stat420.org/

Gareth, J., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Learning* (2nd ed.). https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023.pdf.download.html

Krukovets, D. (2023-2024). Class seminars, Data Preparation and Visualization, Kyiv School of Economics, Kyiv. Retrieved May 17, 2024 from https://teaching.kse.org.ua/course/view.php?id=2127

Nikulina, O. (2024). Class seminars, Regression Analysis, Kyiv School of Economics, Kyiv. Retrieved May 17, 2024, from https://teaching.kse.org.ua/course/view.php?id=2213

Tamilina, L. (2024). Class lectures and seminars, Regression Analysis, Kyiv School of Economics, Kyiv. Retrieved May 17, 2024, from https://teaching.kse.org.ua/course/view.php?id=2213

United Nations. (n.d.). *Education as the Pathway towards Gender Equality | United Nations.* Retrieved May 17, 2024, from https://www.un.org/en/chronicle/article/education-pathway-towards-gender-equality

Annexes

Figure 1

The relationship between the variables

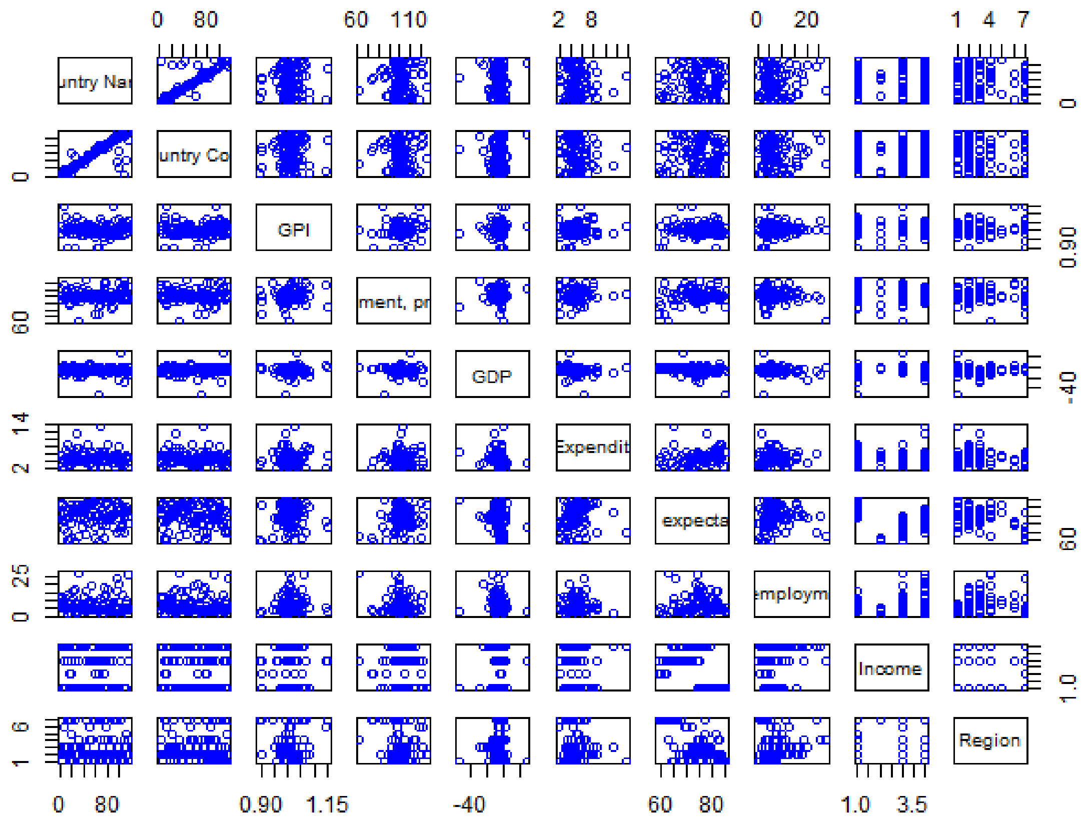

Figure 2

The correlation between the variables

```
> cor(data_plot)
                       GPI Enrollment         GDP EdExpenditures LifeExpectancy Unemployment
GPI             1.00000000 0.13941002  0.1032099   -0.034628789      0.1107179  0.094368025
Enrollment      0.13941002 1.00000000  0.1265650    0.095512824      0.1138382  0.140128589
GDP             0.10320988 0.12656499  1.0000000   -0.220483584     -0.2003972 -0.160452999
EdExpenditures -0.03462879 0.09551282 -0.2204836    1.000000000      0.3545658  0.003068576
LifeExpectancy  0.11071791 0.11383816 -0.2003972    0.354565823      1.0000000  0.102347982
Unemployment    0.09436802 0.14012859 -0.1604530    0.003068576      0.1023480  1.000000000
```

Figure 3

Correlation plot

Figure 4

Vifs for initial (full) model

```
> vif(model1)
                    Enrollment                            GDP                 EdExpenditures
                      1.438941                       1.283692                       1.452789
                LifeExpectancy                   Unemployment                IncomeLow income
                      7.470654                       1.437326                       3.586344
        IncomeLower middle income      IncomeUpper middle income    RegionEurope & Central Asia
                      4.770243                       3.304393                       2.962096
RegionLatin America & Caribbean RegionMiddle East & North Africa             RegionNorth America
                      2.365597                       1.333245                       1.264198
              RegionSouth Asia         RegionSub-Saharan Africa
                      1.436160                       4.426959
```

Figure 5

Comparison of the first three models

```
> anova(model1, model2, model3)
Analysis of Variance Table

Model 1: GPI ~ Enrollment + GDP + EdExpenditures + LifeExpectancy + Unemployment +
    Income + Region
Model 2: GPI ~ `Enrollment, primary` + `Life expectancy` + Income + Region
Model 3: GPI ~ Enrollment * Region + LifeExpectancy * Income
  Res.Df      RSS Df Sum of Sq      F    Pr(>F)
1     80 0.120802
2     83 0.123647 -3 -0.002845 0.9281    0.4316
3     74 0.075614  9  0.048033 5.2231 1.726e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 6

A fitted values versus residuals plot for Model 3

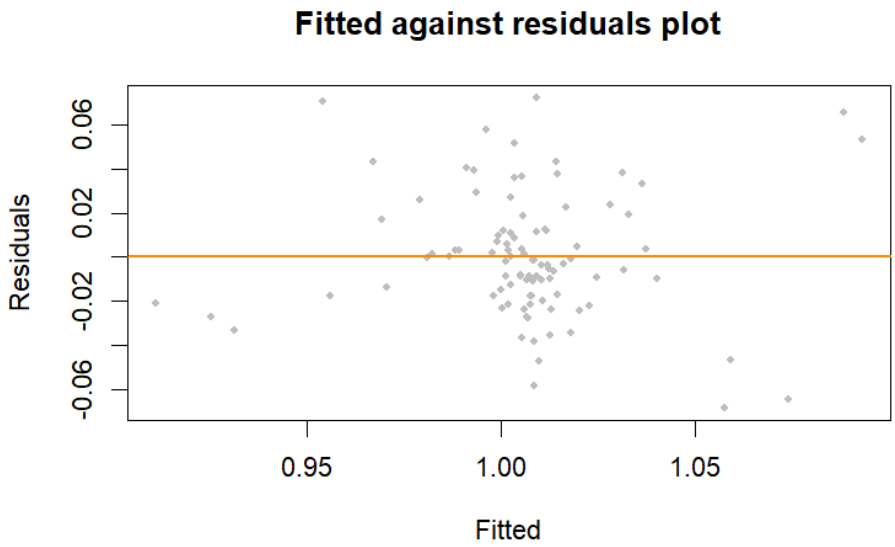

Figure 7

Tests for the Key Assumptions for Model 3. Normal Q-Q Plot

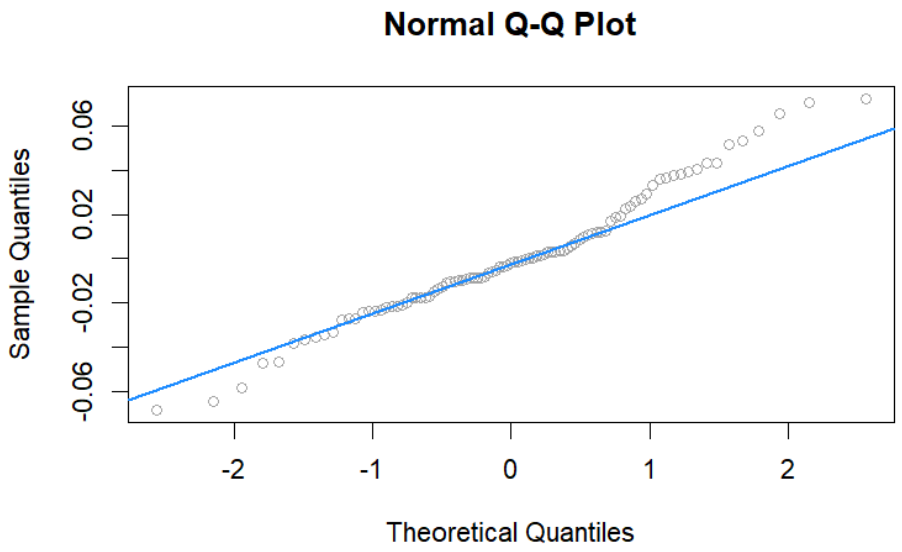

Figure 8

Tests for the Key Assumptions for Model 3. Studentized Breusch-Pagan test

```
> bptest(model3)

        studentized Breusch-Pagan test

data:  model3
BP = 39.954, df = 20, p-value = 0.005063
```

Figure 9

Tests for the Key Assumptions for Model 3. Shapiro-Wilk normality test

```
> shapiro.test(resid(model3))

        Shapiro-Wilk normality test

data:  resid(model3)
W = 0.97768, p-value = 0.1042
```

Figure 10

Checking the assumptions of regression analysis for Model 4

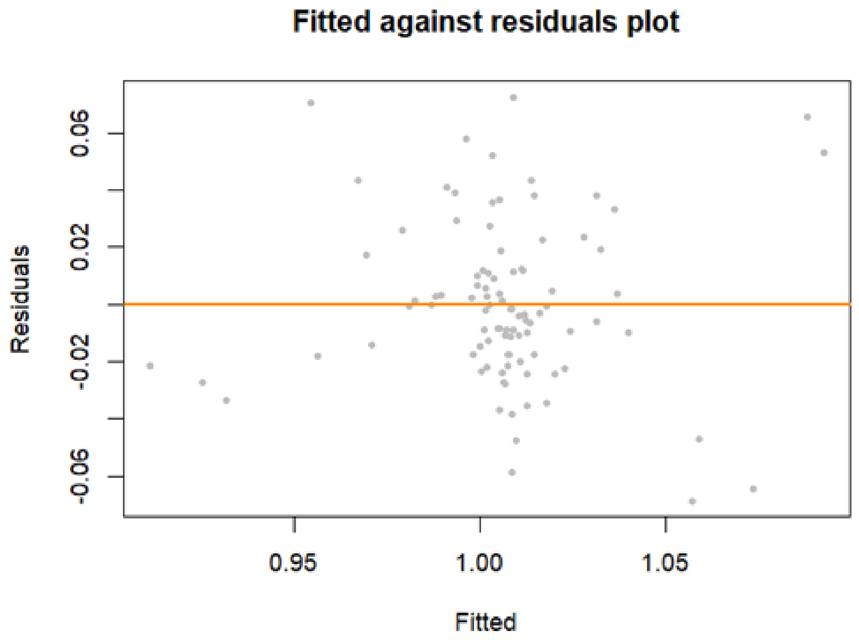

Figure 11

Tests for the Key Assumptions for Model 4. Normal Q-Q Plot

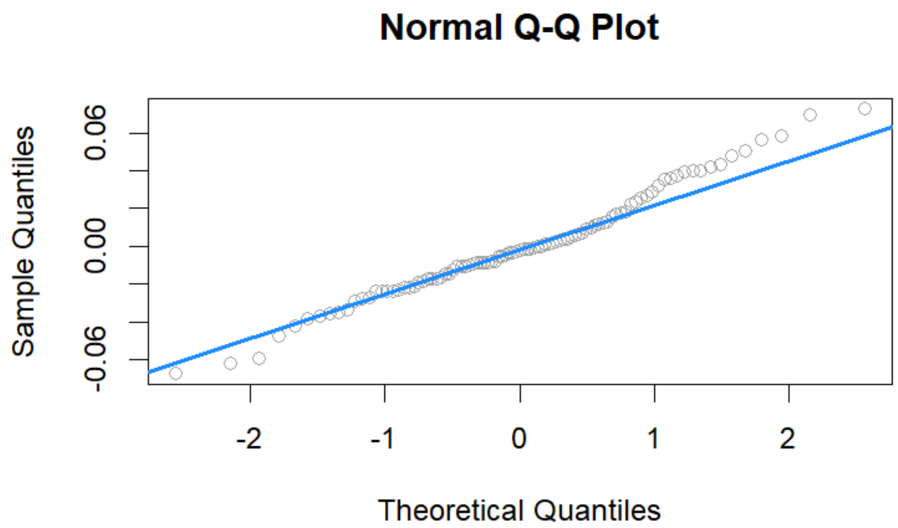

Figure 12

Tests for the Key Assumptions for Model 4. Studentized Breusch-Pagan test

```
> bptest(model4)

        studentized Breusch-Pagan test

data:  model4
BP = 38.583, df = 20, p-value = 0.007509
```

Figure 13

Tests for the Key Assumptions for Model 4. Shapiro-Wilk normality test

```
> shapiro.test(resid(model4))

        Shapiro-Wilk normality test

data:  resid(model4)
W = 0.98195, p-value = 0.2152
```

Figure 14

Standardized coefficients of Model 4

```
> lm.beta(model4)

Call:
lm(formula = log(GPI) ~ Enrollment * Region + LifeExpectancy *
    Income, data = data1)

Standardized Coefficients::
                                  (Intercept)                                    Enrollment
                                           NA                                   0.509720040
                  RegionEurope & Central Asia              RegionLatin America & Caribbean
                                  0.835432245                                   1.088667269
              RegionMiddle East & North Africa                          RegionNorth America
                                  2.520195118                                   4.037813910
                            RegionSouth Asia                      RegionSub-Saharan Africa
                                  0.116352947                                   4.893731179
                               LifeExpectancy                             IncomeLow income
                                 -0.236891859                                 -21.342372274
                      IncomeLower middle income                    IncomeUpper middle income
                                -12.288678338                                  -3.724115078
        Enrollment:RegionEurope & Central Asia   Enrollment:RegionLatin America & Caribbean
                                 -0.903854428                                  -1.192647081
   Enrollment:RegionMiddle East & North Africa              Enrollment:RegionNorth America
                                 -2.738959206                                  -4.095298305
                  Enrollment:RegionSouth Asia          Enrollment:RegionSub-Saharan Africa
                                  0.001689368                                  -3.930119078
              LifeExpectancy:IncomeLow income      LifeExpectancy:IncomeLower middle income
                                 20.385149646                                  11.736730899
     LifeExpectancy:IncomeUpper middle income
                                  3.547507141
```

Table 1

Descriptive statistics of all the variables in full model

Descriptive Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
GPI	115	1	0.039	0.89	0.99	1	1.2
Enrollment, primary	112	101	9.4	62	97	105	126
GDP	112	-5.7	7.8	-55	-8.2	-2.5	30
EdExpenditure	103	4.9	1.8	1.8	3.8	5.6	14
LifeExpectancy							
Unemployment	105	7.4	5.3	0.17	4.1	9.2	28

Table 2

Results of the model

	Dependent variable:			
	GPI			log(GPI)
	(1)	(2)	(3)	(4)
\`Enrollment, primary\`	0.0002	0.0003	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
GDP	0.001			
	(0.001)			
EdExpenditure	-0.001			
	(0.003)			
\`Life expectancy\`	0.003**	0.003**	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Unemployment	0.0003			
	(0.001)			
IncomeLow income	0.011	0.010	-3.584***	-3.654***
	(0.034)	(0.033)	(0.979)	(0.962)
IncomeLower middle income	0.025	0.027	-1.171***	-1.174***
	(0.022)	(0.021)	(0.263)	(0.258)
IncomeUpper middle income	0.024	0.024	-0.318	-0.312
	(0.016)	(0.015)	(0.217)	(0.213)
\`Enrollment, primary\`:RegionEurope & CentralAsia			-0.001	-0.001
			(0.002)	(0.002)
\`Enrollment, primary\`:RegionLatin America & Caribbean			-0.001	-0.001
			(0.002)	(0.002)
\`Enrollment, primary\`:RegionMiddle East & North Africa			-2.738	-0.004**
			(0.002)	(0.002)
\`Enrollment, primary\`:RegionNorth America			-0.011	-0.011
			(0.018)	(0.018)
\`Enrollment, primary\`:RegionSouth Asia			0.0002	0.00000
			(0.002)	(0.002)

\`Enrollment, primary\`:RegionSub-Saharan Africa			-0.005***	-0.005***
			(0.001)	(0.001)
\`Life expectancy\`:IncomeLow income			0.057***	0.058***
			(0.016)	(0.016)
\`Life expectancy\`:IncomeLower middle income			0.016***	0.016***
			(0.004)	(0.003)
\`Life expectancy\`:IncomeUpper middle income			0.004	0.004
			(0.003)	(0.003)
RegionEurope & Central Asia	-0.002	0.0001	0.063	0.064
	(0.014)	(0.012)	(0.157)	(0.154)
RegionLatin America & Caribbean	0.002	0.0003	0.114	0.114
	(0.017)	(0.015)	(0.248)	(0.243)
RegionMiddle East & North Africa	-0.008	-0.008	0.433	0.431*
	(0.021)	(0.020)	(0.220)	(0.216)
RegionNorth America	-0.010	-0.006	1.071	1.075
	(0.031)	(0.030)	(1.782)	(1.752)
RegionSouth Asia	0.051*	0.051*	0.011	0.025
	(0.027)	(0.026)	(0.231)	(0.227)
RegionSub-Saharan Africa	0.033	0.033	0.569***	0.563***
	(0.025)	(0.023)	(0.139)	(0.137)
Constant	0.720***	0.718***	0.881***	-0.116
	(0.147)	(0.139)	(0.218)	(0.215)
Observations	95	95	95	95
R^2	0.149	0.129	0.467	0.474
Adjusted R^2	-0.0002	0.013	0.323	0.332
Residual Std. Error	0.039 (df = 80)	0.039 (df = 83)	0.032 (df = 74)	0.031 (df = 74)
F Statistic	0.999 (df = 14; 80)	1.115 (df = 11; 83)	3.245*** (df = 20; 74)	3.334*** (df = 20; 74)
Note:				*p**p***p<0.01