By: Yoke Hong Si

1005815806

Eco 375: Applied Econometrics

University of Toronto

Department of Economics

Assignment 2

April 2021

Abstract:

This paper explores the effect of income per capita on the democracy index in different countries using simple and multi linear regression models as well as difference-in-differences model in our analysis. From our results, we were able to determine a positive correlation between income and democracy in countries. However, there still exists omitted variables that would lead to bias in our results thus, we are unable to conclude that there is purely a causal effect of income per capita on the democracy index.

1. Introduction

In this paper, we will be examining the effects on democracy and political freedom as income, namely GDP per capita of a country grow. We will be utilising econometric methods such as simple linear regressions, multiple linear regressions and difference-in-differences all whilst accounting for confounders and fixed effects in order to conduct statistical tests. In our results, we determined strong positive correlations between the lagged log real GDP per capita and the democracy index. We will explore if these correlations create causal links between income and democracy.

2. The Context and Data

It is known that the relationship between income per capita and democracy yields a positive correlation as showed in Figure 1. However, we are curious as to whether if the correlation is explained by a causal connection, whereby a higher income per capita indicates that a country becomes more democratic. Thus, we will analyse this relationship through the panel dataset from the paper, "Income and Democracy" written by Acemoglu, et al (2008). To introduce a brief discussion of the summary statistic of our variables, we can look at the size of estimated effects in context. The main democracy index is measured using the Augmented Freedom House Political Rights Index which ranges between 0 and 1, whereby a number closer to 1 would indicate a more democratic country. In other words, a country would have political rights come closest to the ideals suggested by a checklist of questions i.e., whether there are free and fair elections etc. From Table 1, we determined the average of the democracy index (Augmented Freedom House Political Rights Index) between countries to be 0.56830. This indicates that on average, most countries are leaning towards democracy since the average index happened to be greater than 0.5. As shown in figure 2, summary statistics such as education tend to increase from low income to high income countries. On the other hand, colonial countries tend to have lower education levels compared to non-colonial countries as shown in figure 3.

3. Regression analysis and Base Specifications

3.1 Base Specification 1.

The estimated simple linear regression of the Augmented Freedom House Political Rights Index (AFHPRI democracy index) regressed on the one period lag of Log real GDP per capita (PWT) with robust standard errors is as follows:

$$\widehat{fhpolrigaug} = -1.257608 + 0.2254815\ L1.\widehat{lrgdpch}$$
$$(0.0624196) \qquad (0.0072955)$$
$$n = 1023,\ R^2 = 0.4034$$

Firstly, if L1.lrgdpch (lagged log real GDP per capita) = 0, the predicted fhpolrigaug (democracy index) will be the intercept which is 7.323544. If we write the predicted change in fhpolrigaug as a function of change in L1.lrgdpch: $\widehat{fhpolrigaug} = 0.2254815\ (\Delta L1.lrgdpch)$, it means that if the L1.lrgdpch increases by one unit, $\Delta L1.lrgdpch = 1$, then fhpolrigaug is predicted to change by about 0.2254815. In practical terms, this implies that the democracy index increases by 0.2254815 for every additional unit increase in one period lag of log real GDP per capita.

The t-statistic gives a value of 30.91 which falls outside the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy is statistically significant. The 95% confidence interval for this estimate is computed to be [0.2111657, 0.2397974]. In this case, zero is not part of this interval so we reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between 0.2111657 and 0.2397974.

3.2 Base Specification 2.

In this specification, we estimate an SLR of the AFHPRI regressed on the one period lag of Log real GDP per capita with standard errors clustered at the country level. In this regression, we obtained identical point estimates for both the coefficient of the log real GDP per capita, and the intercept as Base Specification 1. However, the 95% confidence level is determined to be [0.2008087, 0.2501544] which differs since the robust standard errors are different. Similarly, zero is not part of this interval so we reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between 0.2111657 and 0.2397974 for clustered standard errors. Clustered standard errors allow for unrestricted forms of serial correlation and heteroskedasticity in panel data. Since we expect the error term of the regression to be correlated across countries or over time, it is important to control for this factor using clustered standard errors. Therefore, the regression would suffer from omitted variable bias since time-invariant confounders could lurk in the error term that is not accounted for. An example of a time invariant factor that varies with democracy and income systematically which may bias the results is as follows. Consider a comparison between the United States and North Korea where the US is both richer and more democratic thus, a simple cross-country comparison that does not control for fixed country effects would suggest that higher per capita GDP causes democracy. Therefore, we should account for the fixed country effects in order to investigate “within-country variation” instead of, for instance, if the US is more likely to be relatively democratic as it becomes relatively richer.

3.3 Base Specification 3.

Here, we investigate a regression of the AFHPRI on the one period lag of log real GDP per capita with the inclusion of country fixed effects (code_numeric) in the regression. The estimated coefficient on the lag of lrgdpch is determined to be 0.0550704 which is relatively small in terms of magnitude and a lot smaller than in specification (2). Thus, the point estimate has changed from specification (2). Additionally, it means that if the L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about 0.0550704. In practical terms, this implies that the democracy index increases by 0.0550704 for every additional unit increase in one period lag of log real GDP per capita, whilst controlling for country fixed effects. The t-statistic gives a value of 1.67 which falls within the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy with country fixed effects is small and not statistically significant. The 95% confidence interval for this estimate is computed to be [-0.0100479, 0.1201887]. In this case, zero is part of this interval so we do not reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between -0.0100479 and 0.1201887. However, the regression would still suffer from omitted variable bias since time varying confounders could be present, affecting the point estimate. An example of a time varying factor that varies by country which may also vary with democracy and income systematically is as follows. Consider a comparison between the United Kingdom and Vietnam during the Vietnam War. Thus, we would expect the UK to be both richer and more democratic during this war period which would invalidate any causal interpretation from the regression.

3.4 Base Specification 4

In the fourth base specification, we regress the AFHPRI on the one period lag of log real GDP per capita with the inclusion of both country fixed effects (code_numeric) and year fixed effects (year_numeric) in the regression. The estimated coefficient on the lag of lrgdpch is determined to be 0.0587276 which has changed from specification (3) since is slightly larger than 0.0550704 in specification (3) but still relatively small in terms of overall magnitude. Thus, the coefficient has changed from (3). The coefficient implies that if the L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about 0.0587276. In practical terms, this implies that the democracy index increases by 0.0587276 for every additional unit increase in one period lag of log real GDP per capita, whilst controlling for country and year fixed effects. The t-statistic gives a value of 1.29 which falls within the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy with country and year fixed effects is small and not statistically significant. The 95% confidence interval for this estimate is computed to be [-0.0310453, 0.1485006]. In this case, zero is part of this interval so

we do not reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between -0.0310453 and 0.1485006. By testing the exclusion of the year fixed effects, we obtain an F-statistic of 106.39854 and p-value of 1.313e-23.

3.5 Base Specification 5

In this base specification, we first regressed the democracy index on the one period lag income per capita, including country and year fixed effects as well as lagged demographic controls (age, education and log population). We then defined a subsample using this regression. Finally, we ran specification (4) on the subsample generated and obtain the following regression model.

$$\widehat{fhpolrigaug} = 0.8798518 - 0.0228115\ L1.\widehat{lrgdpch}$$

$$(0.4895312) \qquad (0.053769)$$

$$n = 685,\ R^2 = 0.7259$$

3.6 Base Specification 6

For our last base specification, we first regressed the democracy index on the one period lag income per capita, including country and year fixed effects as well as lagged demographic controls (age, education and log population). The estimated coefficient on the lag of lrgdpch is determined to be 0.0159635 which is smaller than 0.0587276 in specification (4) and still relatively small in terms of overall magnitude. Thus, the coefficient has changed from (4). The coefficient implies that if L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about 0.0159635. In practical terms, this implies that the democracy index increases by 0.0159635 for every additional unit increase in one period lag of log real GDP per capita, whilst controlling for country and year fixed effects as well as demographics. This could be due to the change of sample between specifications (6) and (4) where specification (6) has 685 observations and includes demographic controls whereas (4) have 1023 observations and does not include demographic controls. The t-statistic gives a value of 0.30 which falls within the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy with the controls in specification (6) is not statistically significant.

By testing the exclusion of the year fixed effects, we obtain an F-statistic of 50.928385 and p-value of 2.889e-12. Testing the exclusion of the age variables, we obtain an F-statistic of 4.3295226 and p-value of 0.00185436. Lastly, testing the exclusion of the year fixed effects, we obtain an F-statistic of 42.127508 and p-value of 1.199e-42.

The estimated coefficients of the lagged age group variables ranging from 15-30, 30-45, 45-60 and 60-years old are as follows. In order from the youngest age group to the oldest, we have -0.5094067,

-2.61807, 0.4518443 and 0.8482768. The coefficient implies that if L1.age_young, L1.age_midage, L1.age_old, and L1.age_very old increases by one percent, fhpolrigaug is predicted to change by their respective values mentioned, divided by 100. Since only the 30-45 age group is statistically significant where the t-statistic, -2.76, falls outside the interval of [-1.96, 1.96], it is the only coefficient to have an effect at the population level. In practical terms, this implies that the democracy index decreases by -2.61807/100 = -0.0262 for every one percent increase increase in one period lag of the 30-45 age group, whilst controlling for country and year fixed effects as well as education and log population. The estimated coefficient on the lag of education is determined to be -0.0035109. The coefficient implies that if L1.education increases by one unit, then fhpolrigaug is predicted to change by about -0.0035109. In practical terms, this implies that the democracy index decreases by 0.0035109 for every additional unit increase in one period lag of education, whilst controlling for country and year fixed effects as well as demographics. Lastly, the estimated coefficient on the lag of log population is determined to be -0.1044869. The coefficient implies that if L1.lpop increases by one unit, then fhpolrigaug is predicted to change by about -0.0035109. In practical terms, this implies that the democracy index decreases by 0.0035109 for every additional unit increase in one period lag of log population, whilst controlling for country and year fixed effects as well as demographics.

The F-statistics obtained for year fixed effects, age controls and demographic controls are 50.928385, 4.3295226 and 42.127508, respectively. Since these values are above the 5% critical value of 1.96, we reject the null hypothesis. In other words, year fixed effects, age controls and demographic are individually jointly significant in the democracy index regression. Therefore, these demographic controls contributes to the slope parameter which affects the democracy index.

4. Extension

In this part of the paper, I will be extending my previous analysis by exploring specifications that include interaction terms. The regression models I will be exploring are as follows. Note: We control for lagged demographic variables including age groups, education and log population.

4.1 Extension Specification 1

$$\widehat{fhpolrigaug} = 2.718267 - 0.0282674\ L1.\widehat{lrgdpch} - 0.3482413\ L1.\widehat{nsave}$$

$$(1.634611) \qquad (0.060824) \qquad (1.119956)$$

$$+0.072851\ L1.\widehat{(lrgdpch \times nsave)} - 0.0023416\ L1.\widehat{education} - 0.1309011\ \widehat{L1.lpop}$$

$$(0.1419401) \qquad (0.0233862) \qquad (0.133988)$$

$$n = 685,\ R^2 = 0.7366$$

In specification 1, we explore the interaction effect between the income per capita (lrgdpch) and the nominal savings rate (nsave). The coefficient of the interaction term is determined to be 0.072851. Since we are interested in the effects of income on democracy in a country, we determine this effect to be $\frac{\Delta fhpolrigaug}{\Delta L1.lrgdpch} = -0.0282674 + 0.072851(L1.nsave)$. In other words, in order to estimate the partial effect of L1.lrgdpch on fhpolrigaug, we must plug in a value of L1.nsave to obtain this partial effect. The mean value of nsave in the sample is 0.16519 as determined in table 1 thus, at the mean nsave, the effect of L1.lrgdpch on fhpolrigaug is $-0.0282674 + 0.072851(0.16519) = -0.01623$. Therefore, if L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about -0.01623 when L1.nsave is 0.16519. In practical terms, this implies that the democracy index decreases by 0.01623 for every additional unit increase in one period lag of log real GDP per capita, when the nominal savings rate is equal to 0.16519, whilst controlling for country and year fixed effects as well as demographics. In order to determine if the estimate -0.01623 is statistically different from zero, we rerun the regression where we replace the interaction term, $L1.(lrgdpch \times nsave)$, with $L1.(lrgdpch \times (nsave - (-0.01623)))$. Thus, this yields a t-statistic of 0.51. Hence, it is statistically insignificant at the 5% significance level since 0.51 lies between [-1.96, 1.96]. Therefore, in terms of practical significance, at the average nsave value of 0.16519, we conclude that income has no effect on income per capita at the population level.

4.2 Extension Specification 2

$$\widehat{fhpolrigaug} = 1.686952 + 0.0100972\ L1.\widehat{lrgdpch} - 9.061232\ L1.\widehat{socialist}$$

(1.483372) (0.0531308) (1.898741)

$$-1.03265\ L1.(\widehat{lrgdpch \times socialist}) - 0.0016044\ L1.\widehat{education} - 0.0692228\ \widehat{L1.lpop}$$

(0.2144691) (0.0233138) (0.1224767)

$n = 685$, $R^2 = 0.6892$

In specification 2, we explore the interaction effect between the income per capita (lrgdpch) and the socialist status (socialist). The coefficient of the interaction term is determined to be -1.03265. Since we are interested in the effects of income on democracy in a country, we determine this effect to be $\frac{\Delta fhpolrigaug}{\Delta L1.lrgdpch} = 0.0100972 - 1.03265(L1.socialist)$. Similarly, we estimate the partial effect of L1.lrgdpch on fhpolrigaug by plugging the mean value of L1.socialist, 0.14966 to obtain this partial effect. Thus, the effect of L1.lrgdpch on fhpolrigaug is

$0.01009720 - 1.03265(0.14966) = -0.14445$. Therefore, if L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about -0.14445 when L1.socialist is 0.14966. In practical terms,

this implies that the democracy index decreases by 0.14445 for every additional unit increase in one period lag of log real GDP per capita, when the lagged socialist dummy is equal to 0.14966, whilst controlling for country and year fixed effects as well as demographics. In order to determine if the estimate 0.14445 is statistically different from zero, we rerun the regression where we replace the interaction term, $L1.(lrgdpch \times socialist)$, with $L1.(lrgdpch \times (socialist - (-0.14445)))$. Thus, this yields a t-statistic of 4.81. Hence, it is statistically significant at the 5% significance level since 4.81 does not lies between [-1.96, 1.96]. Therefore, in terms of practical significance, at the average socialist value of 0.14966, we conclude that income has a statistically significant negative effect on income per capita at the population level.

5. Limitations of results

Throughout our specifications, we encounter issues of omitted variable bias. For instance, base specifications (1) through (4) omitted key elements such as year fixed effects, country fixed effects and controls for demographics. Furthermore, the use of clustered standard errors is also important when accounting for serial correlation and heteroskedasticity. We solve most of these issues in specification (6) however, there are still remaining threats to the validity of my regression results. Throughout my analysis, we argue that although income and democracy are positively correlated, there is insufficient evidence to conclude a causal effect at the population level. This is due to omitted variables that could exist which violates the zero conditional mean assumption since these variables lurk in the error term and are correlated with the income and democracy index variables. These variables could be part of the demographics which are not part of age, education or the log population. For instance, the innate ability of a person (IQ) would affect the person's income which in turn, affects the democracy index. Without the use of quasi-experimental methods such as randomized controlled trials or IV estimates, we are unable to satisfy the zero conditional mean assumption. Therefore, the bias in our results would prevent us to determine a causal interpretation of income on democracy.

6. Conclusion

In my research paper, we determined a positive correlation between income and democracy. On the other hand, we determined the effect of being part of a socialist block to have a negative impact on democracy through income by using a difference-in-differences regression model. However, correlation does not imply causation since there are omitted variables present that would cause bias in our models. Therefore, there is no evidence of a causal effect of income per capita on the Augmented Freedom House Political Rights Index.

References:

Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2008. "Income and Democracy." American Economic Review, 98 (3): 808-42.DOI: 10.1257/aer.98.3.808

Wooldridge Introductory Econometrics Textbook

Table 1: Table of descriptive statistic

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
Year of observation	11	1975	15.8	1950	2000
Augmented Freedom House Political Rights Index	1,194	0.56830	0.35775	0	1
Log real GDP per capita (PWT)	1,199	8.1650	1.0372	5.77393	10.692
Nominal savings rate	1,206	0.16519	0.13644	-0.76600	0.73962
% labor share of gross value added	504	0.37040	0.13089	.07628	0.73844
Percent population age 0-15	1,551	0.36340	0.09475	0.14292	0.51610
Percent population age 15-30	1,551	0.25460	0.02565	0.16855	0.34658
Percent population age 30-45	1,551	0.17534	0.02983	0.08966	0.29793
Percent population age 45-60	1,551	0.11838	0.03744	0.05601	0.22149
Percent population age 60-	1,551	0.08829	0.04930	0.03165	0.24063
Median age in population	1,529	23.121	6.6637	14.4	41.3
Average schooling years	722	4.4584	2.8739	0.042	12.179
Log (total population in thousands)	1,308	8.6819	1.9073	3.7136	14.049
Polity IV index	1,092	0.55627	0.37931	0	1
Dummy for Soviet Block, including iron curtain	1,617	0.14966	0.35685	0	1
Dummy former colony vs non-colony	1,617	0.70068	0.45810	0	1

Table 2: Regression Analysis of democracy index on log real GDP per capita

	(1)	(2)	(3)	(4)	(5)	(6)
Lag of log real GDP pc	0.225***	0.225***	0.0551*	0.0587	-0.0228	0.0160
	(0.0073)	(0.0125)	(0.0329)	(0.0454)	(0.0538)	(0.0532)
	[0.2112, 0.2398]	[0.2008, 0.2502]	[-0.0100, 0.1202]	[-0.0310, 0.1485]	[-0.1296, 0.0840]	[-0.0896, 0.1216]
Age:						
Age group: 15-30						-0.509
						(0.678)
Age group: 30-45						-2.618
						(0.949)
Age group: 45-60						0.452
						(1.413)
Age group: 60-						0.848
						(1.390)
Education					-0.00508	-0.00351
					(0.0244)	(0.0237)
Log population					-0.170	-0.104
					(0.121)	(0.133)
F-statistics & P-values:						
Year fixed Effects				106.4***	60.99***	50.93***
				(1.3e-23)	(2.7e-14)	(2.9e-12)
Age Controls						4.33***
						(.00185)
Demographic Controls						42.13***
						(1.2e-42)
Year Fixed Effects	No	No	No	Yes	Yes	Yes
Country Fixed Effects	No	No	Yes	Yes	Yes	Yes
Clustered Standard Err.	No	Yes	Yes	Yes	Yes	Yes
Demographic Sample	No	No	No	No	Yes	Yes
Number of Obs.	1023	1023	1023	1023	685	685
Number of Countries	147	147	147	147	94	94
R-Squared	0.4034	0.4034	0.7272	0.7570	0.6973	0.7354

The values in Table 2 are computed using Stata. The dataset used to compute the values is obtained from the Acemoglu, et al. (2008) panel dataset. The time period of this dataset is from years 1950 to 2000. The dependent variable is the Augmented Freedom House Political Rights Index which refers to the democracy index. All variables with estimates in this table are based on a five-year lag period. The quantities in parentheses below the estimates are the standard errors. Confidence Intervals are below the standard errors and are in square brackets. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 3: Regression Analysis of democracy index on log real GDP per capita with interaction terms

	(1)	(2)
Lag of log real GDP pc	-0.0283	0.0101
	(0.0608)	(0.0531)
Savings Rate	-0.348	
	(1.120)	
Socialist Status		-9.061***
		(1.899)
Education	-0.00234	-0.00160
	(0.0234)	(0.0233)
Log Population	-0.131	-0.0692
	(0.134)	(0.122)
Interaction Terms (Lagged):		
Income × Savings Rate	0.0729	
	(0.142)	
Income × Socialist		1.033***
		(0.214)
Age:		
Age group: 15-30	-0.567	-0.368
	(0.634)	(0.629)
Age group: 30-45	-2.777**	-2.536**
	(0.984)	(0.951)
Age group: 45-60	0.384	0.417
	(1.407)	(1.406)
Age group: 60-	0.991	1.167
	(1.384)	(1.318)
Year Fixed Effects	Yes	Yes
Country Fixed Effects	Yes	Yes
Clustered Standard Err.	Yes	Yes
Demographic Sample	Yes	Yes
Number of Obs.	685	685
Number of Countries	94	94
R-Squared	0.7366	0.7383

The values in Table 3 are computed using Stata. The dataset used to compute the values is obtained from the Acemoglu, et al. (2008) panel dataset. The time period of this dataset is from years 1950 to 2000. The dependent variable is the Augmented Freedom House Political Rights Index which refers to the democracy index. All variables with estimates in this table are based on a five-year lag period. The quantities in parentheses below the estimates are the standard errors.
***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Figure 1

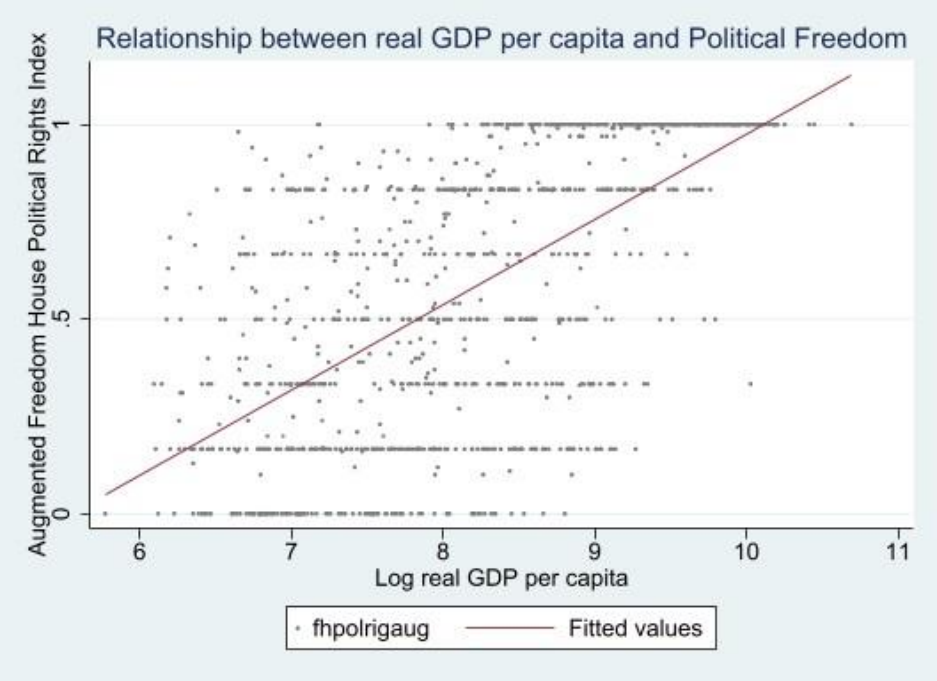


Figure 2

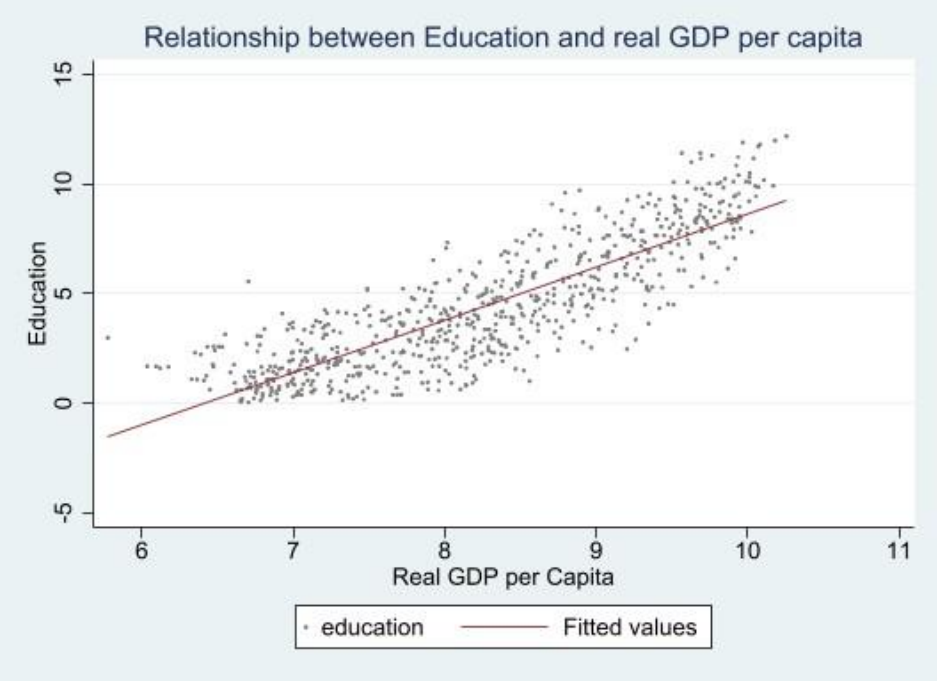
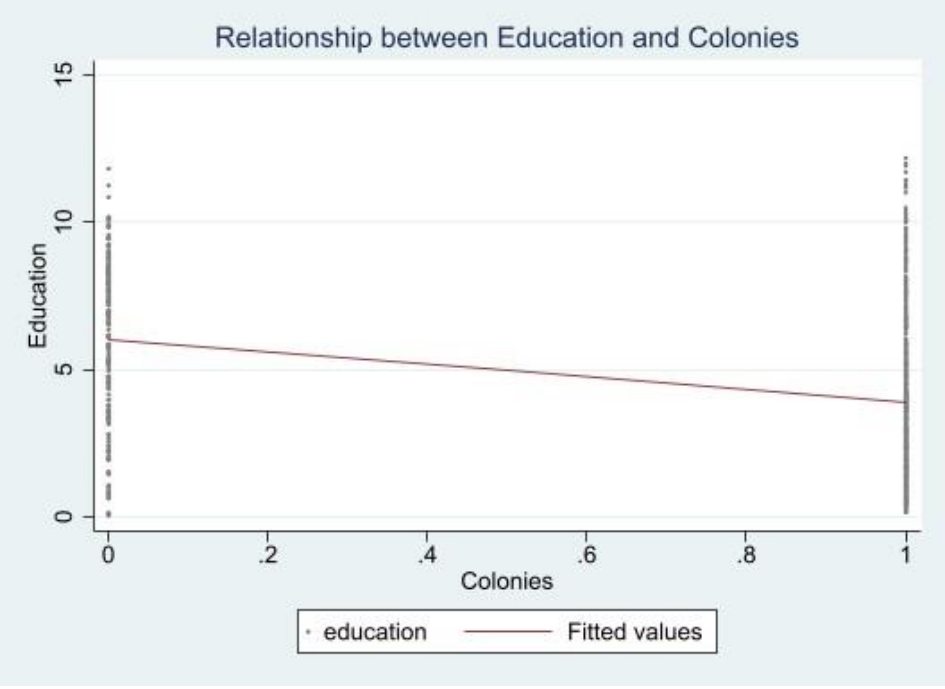


Figure 3


Relationship between Education and Colonies
Education
0
5
10
15
0
.2
.4
.6
.8
1
Colonies
education
Fitted values