Alkan Can Mollaibrahimoglu

CSSM 502

FINAL PROJECT

INTRODUCTION

The Covid-19 pandemic starting very beginning of 2020 has challenged every aspect of life in all the countries. Its effects have been economically, socially, politically, and psychologically threatening. Therefore, governments had to take on responsibilities to prevent the spread of the virus applying strict measures such as lockdowns, closing borders, shutting down workplaces, travel bans, mask mandates, etc. Different countries followed different ways to prevent the virus according to their social, economic, political circumstances.

The task and the responsibility to prevent the spread of the virus have not been limited to the governments, preventative measures by individuals have also been an important component fighting against the spread of coronavirus. Additionally, how the public have reacted to the measures taken by the governments had a significant role in the battle against the Covid-19 pandemic. Both of these two behaviors are expected to be affected by the education level of citizens within the country. Based on this hypothesis, this project aims to analyze the correlation between the education level of the country and the spread of the coronavirus.

METHOD AND DATA:

A cross-country OLS regression analysis will be conducted for the purpose of this project.

Dependent Variable

Infection Rate: Total number of confirmed Covid-19 cases in a country by the end of 2021 / Population of the country. Total cases are updated daily and in a cumulative way in OurWorldinData. Therefore, the total number on the date 31.12.2021 has been taken.

Data Source: OurWorldinData

Independent Variables

Explanatory Variables: Lower Secondary School Completion Rate,Total (% Relevant Age group)

Primary Completion Rate, Total (%Relevant Age group)

School Enrollment, Tertiary (%gross)

Data Source: World Bank API

For each indicator for education level, inferences will be made once at a time.

Control Variables:

a) Democracy Index(eiu): The democracy level of a country determines the ability of the government to take quick actions and impose strict measures against the spread of the virus. Democratic countries are also expected to provide their citizens with better access to education. Therefore, the democracy level is expected to affect the education level. Democracy scores of countries have been gathered from Economic Intelligence Unit (EIU)(2019). The reason why the democracy score for the year 2019 has been considered is that it is the period that corresponds to the beginning of the pandemic.
b) GDP: The economic power of a country, determines the ability of the government to sustain measures such as lockdowns, closing workplaces through compensating them for their citizens. The countries with the strong economy are expected to provide better education opportunities. GDP as an indicator for the economic level of a country, therefore, has been taken as a control variable. GDP data of countries (2019) have been gathered from World Bank API.
c) Urban Population: The data for the urban population have been associated with mobility which is a determinant factor for the spread of the virus. This data have been gathered from World Bank API.
d) Population Density: The number of people per km2 in a land area. Population density is also expected to affect the spread of the virus. The data have been gathered from OurWorldinData.
e) Gini Index: It is a measure indicating the distribution of income inequality. It has been gathered from World Bank Data and since it is not available every year, the most recent value for each country has been taken. 0 corresponds to perfect equality and 100 implies

perfect inequality. Income inequality is associated with the luxury for people to stay at home during the period of the pandemic within the scope of this analysis. It is also a factor affecting people's access to education. Therefore, Gini Index will be taken as a control variable.

MODEL

Secondary School Completion Rate (model 1)

There seems no linear relationship between infection rate and secondary school completion rate. (Figure 1)

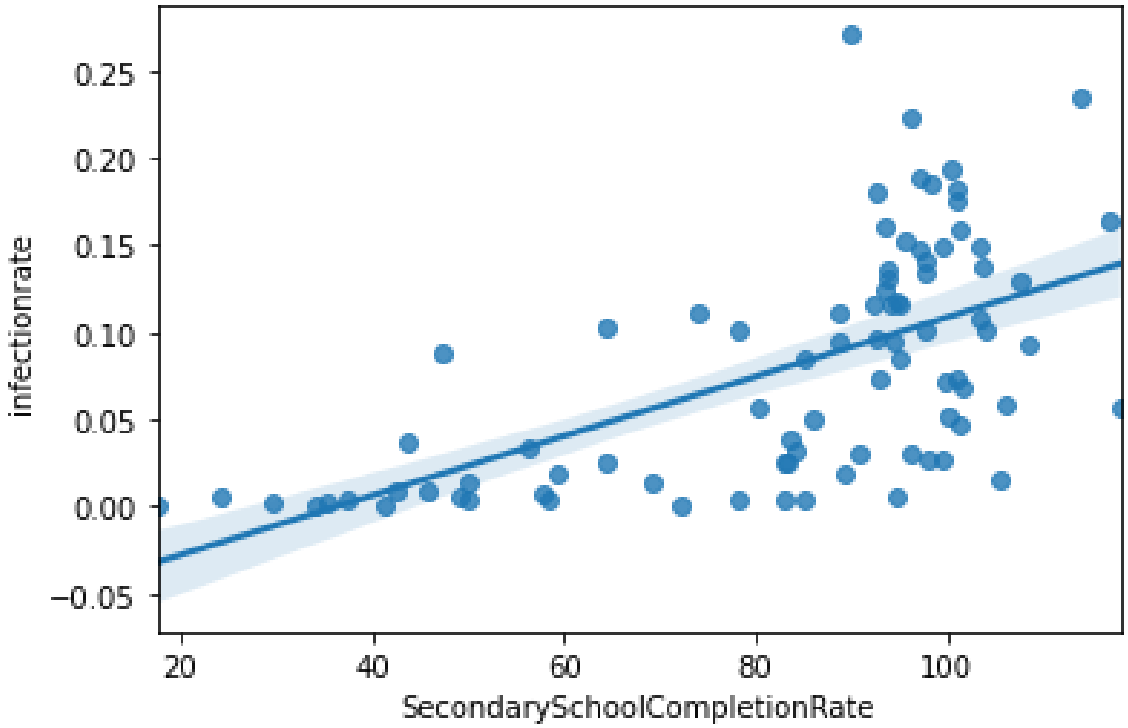


Figure 1

Transformed the data taking the natural log of both sides to provide linearity. (Figure 2)

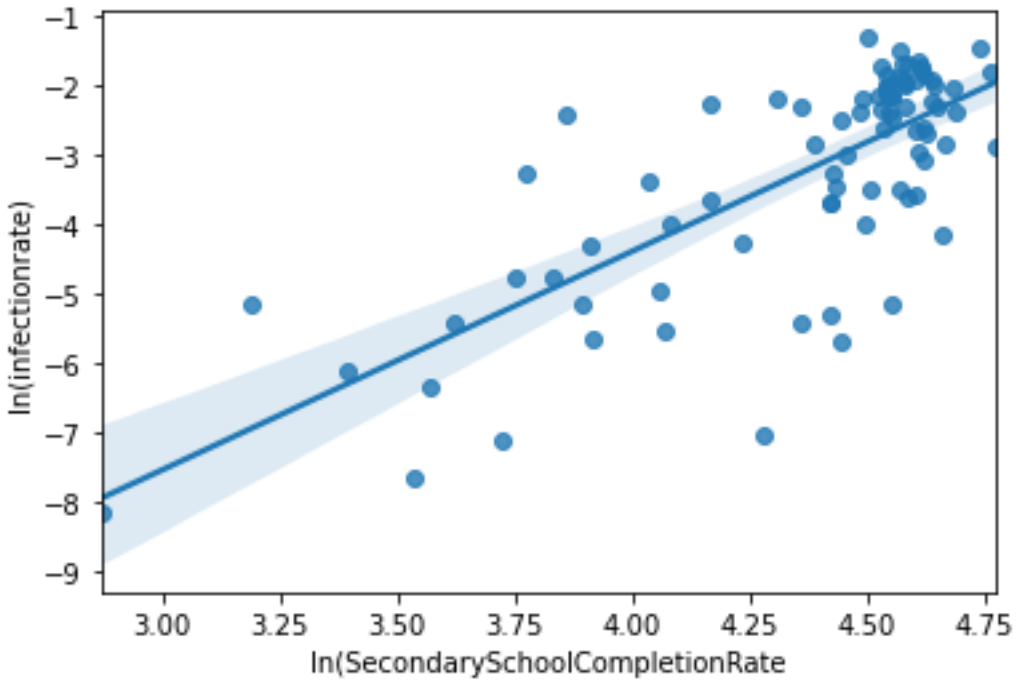


Figure 2

Regression Equation (for all the three models)

$\ln(Yi)=ß0+ß_i*\ln(X_{ij})+ \varepsilon$

Y: Infection rate of the country i

Xij: Independent Variable j for country i

OLS Regression Results

Dep. Variable:	infectionrate	R-squared:	0.751
Model:	OLS	Adj. R-squared:	0.730
Method:	Least Squares	F-statistic:	35.26
Date:	Fri, 28 Jan 2022	Prob (F-statistic):	2.67e-19
Time:	00:47:41	Log-Likelihood:	-92.243
No. Observations:	77	AIC:	198.5
Df Residuals:	70	BIC:	214.9
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-22.2532	2.793	-7.966	0.000	-27.825	-16.682
SecondarySchoolCompletionRate	1.1287	0.426	2.649	0.010	0.279	1.978
GDP	1.0375	0.185	5.603	0.000	0.668	1.407
eiu	-0.1702	0.414	-0.411	0.683	-0.997	0.656
Urbanpop	-0.0668	0.067	-0.995	0.323	-0.201	0.067
population_density	0.0306	0.078	0.392	0.696	-0.125	0.186
giniWB	1.4942	0.563	2.652	0.010	0.370	2.618

Table 1

Based on the results (Table1), there is a positive correlation between Secondary School Completion rate and infection rate (ß1=1,1287). This correlation is significant at the 0,01 significance level. GDP and Gini index is also positively correlated with infection rate at 0,01 significance level when Secondary School Completion Rate is taken into account as an indicator for education level. The R^2 value for the model is 0,751.

Although the constant value has a high VIF, the others' are below 5 so there is no problem of multicollinearity.

	VIFS
const	849.826972
SecondarySchoolCompletionRate	2.955858
GDP	4.678869
eiu	2.365332
Urbanpop	1.085240
population_density	1.058832
giniWB	1.329923

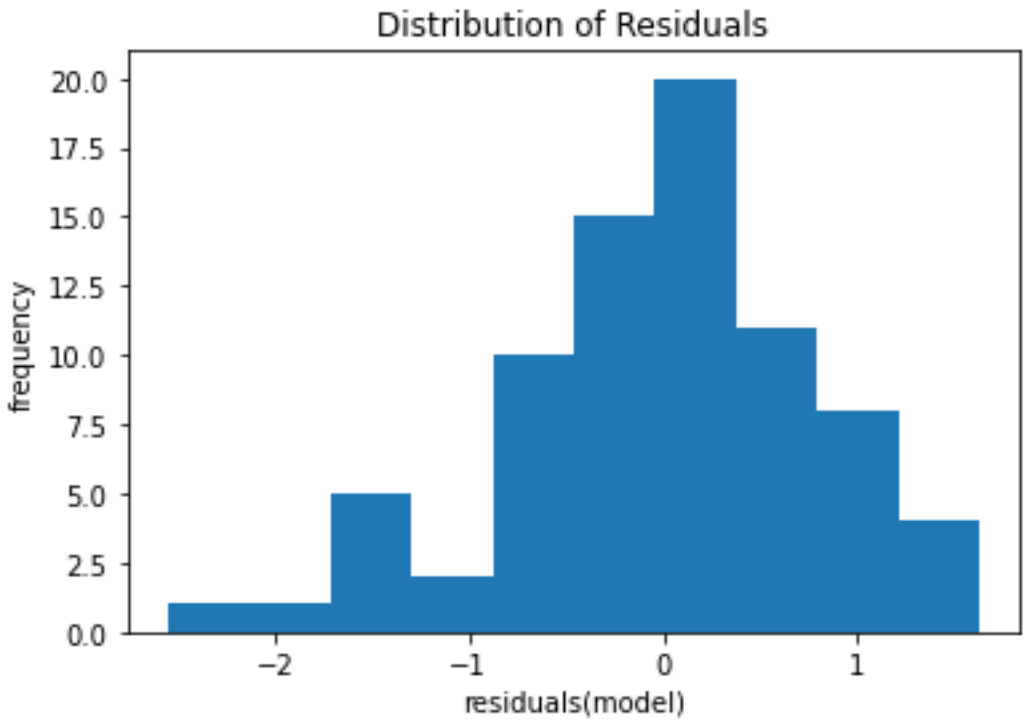


Figure 3

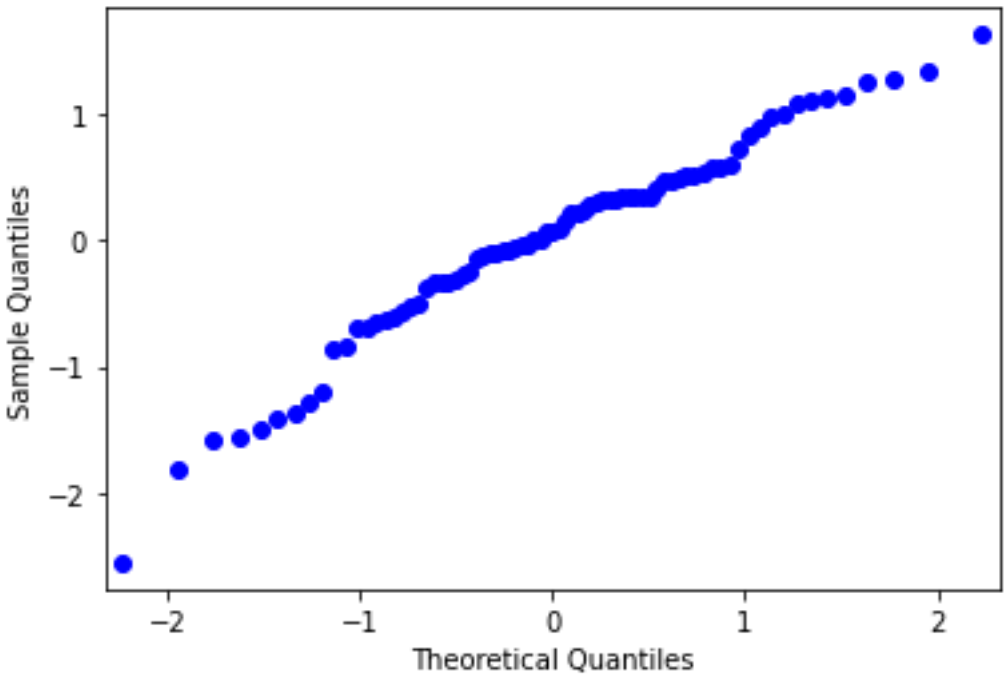


Figure 4

Based on the histogram (Figure 3) and Qqplot(Figure 4) residuals of the model follow approximately a normal distribution. The residual plot right above in Figure-5 indicates that residuals are randomly distributed.(Homoscesdacity)

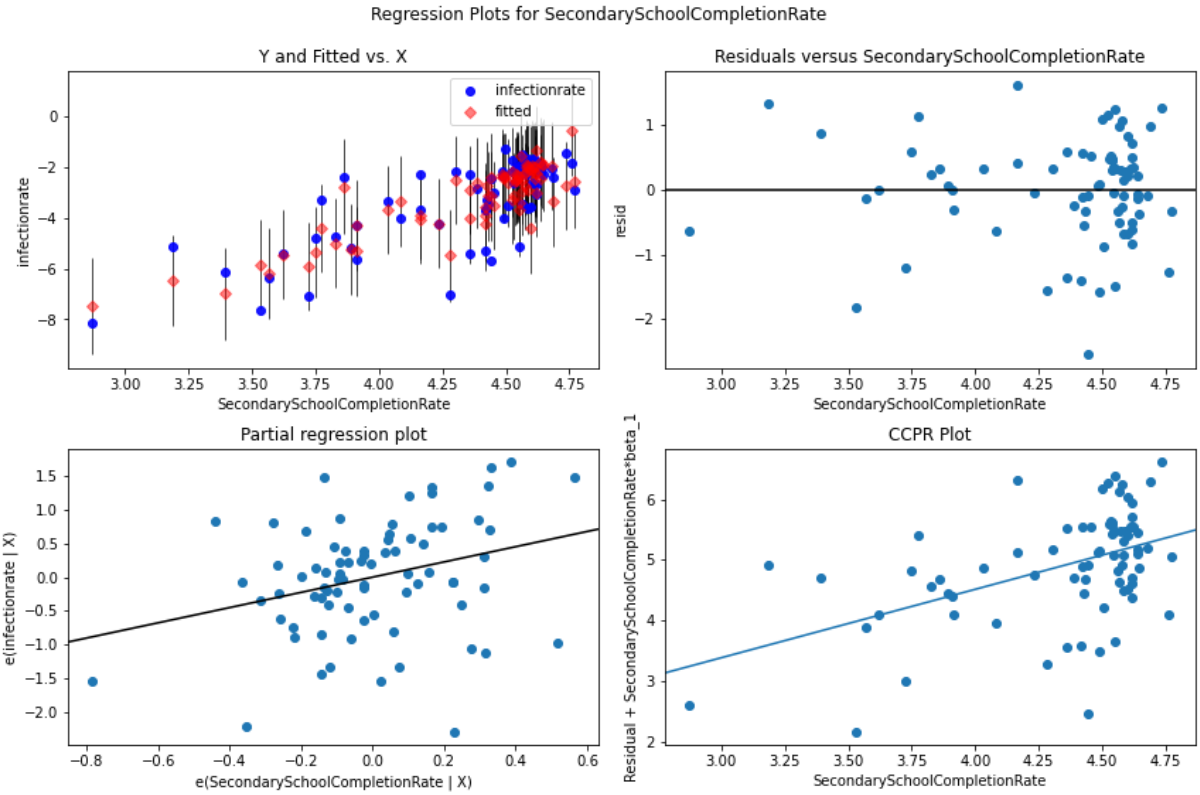


Figure 5

Primary School Completion Rate(model 2):

OLS Regression Results

Dep. Variable:	infectionrate	R-squared:	0.775
Model:	OLS	Adj. R-squared:	0.757
Method:	Least Squares	F-statistic:	43.62
Date:	Fri, 28 Jan 2022	Prob (F-statistic):	1.16e-22
Time:	02:40:47	Log-Likelihood:	-95.517
No. Observations:	83	AIC:	205.0
Df Residuals:	76	BIC:	222.0
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-28.8426	3.543	-8.141	0.000	-35.899	-21.786
primarycompletionrate	2.6074	0.709	3.677	0.000	1.195	4.020
GDP	1.0836	0.144	7.506	0.000	0.796	1.371
eiu	-0.2179	0.363	-0.601	0.550	-0.940	0.504
Urbanpop	-0.0859	0.061	-1.407	0.164	-0.207	0.036
population_density	0.0377	0.072	0.524	0.601	-0.106	0.181
giniWB	1.3945	0.503	2.770	0.007	0.392	2.397

Table 2

There is a positive correlation between Primary School Completion rate and infection rate(ß1=2,6074). This correlation is significant at the 0,01 significance level. GDP and Gini Index is also positively correlated with infection rate at 0,01 significance level when Primary School Completion Rate is included in the model. The value of R^2 is 0,775 for this model.

Again for model 2, there is no multicollinearity, residuals are random and normally distributed. (see Appendix)

	VIFS
const	1630.873902
primarycompletionrate	1.881498
GDP	3.478198
eiu	2.267327
Urbanpop	1.073006
population_density	1.064178
giniWB	1.329578

Tertiary Enrolment (model 3):

OLS Regression Results

Dep. Variable:	infectionrate	R-squared:	0.690
Model:	OLS	Adj. R-squared:	0.665
Method:	Least Squares	F-statistic:	27.45
Date:	Fri, 28 Jan 2022	Prob (F-statistic):	5.47e-17
Time:	03:32:29	Log-Likelihood:	-109.30
No. Observations:	81	AIC:	232.6
Df Residuals:	74	BIC:	249.4
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-11.5086	3.102	-3.710	0.000	-17.690	-5.327
SchoolEnrolment(tertiary)	1.0372	0.231	4.484	0.000	0.576	1.498
GDP	0.3440	0.228	1.509	0.136	-0.110	0.798
eiu	0.6135	0.411	1.493	0.140	-0.205	1.433
Urbanpop	-0.2795	0.073	-3.830	0.000	-0.425	-0.134
population_density	0.1518	0.084	1.813	0.074	-0.015	0.319
giniWB	1.0685	0.667	1.601	0.114	-0.261	2.398

Table 3

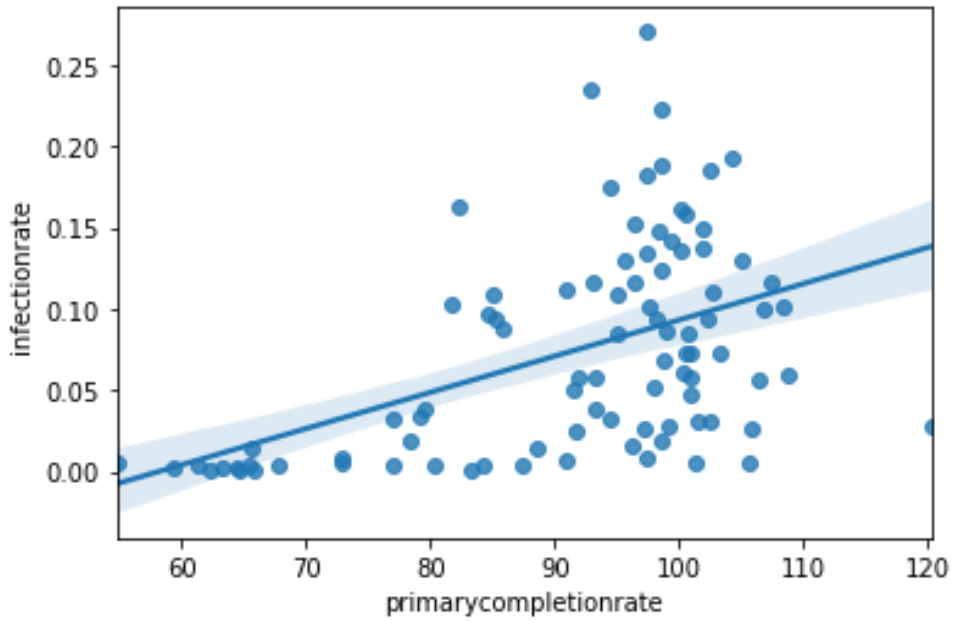
Tertiary enrolment is positively correlated with infection rate at 0,01 significance level(ß1=1,0372). In this model, Urban Population is negatively correlated with infection rate at 0,01 significance level. Population density is positively correlated with infection rate at a 0,1 significance level. The R^2 of the model is slightly lower than the other two that is 0,690.

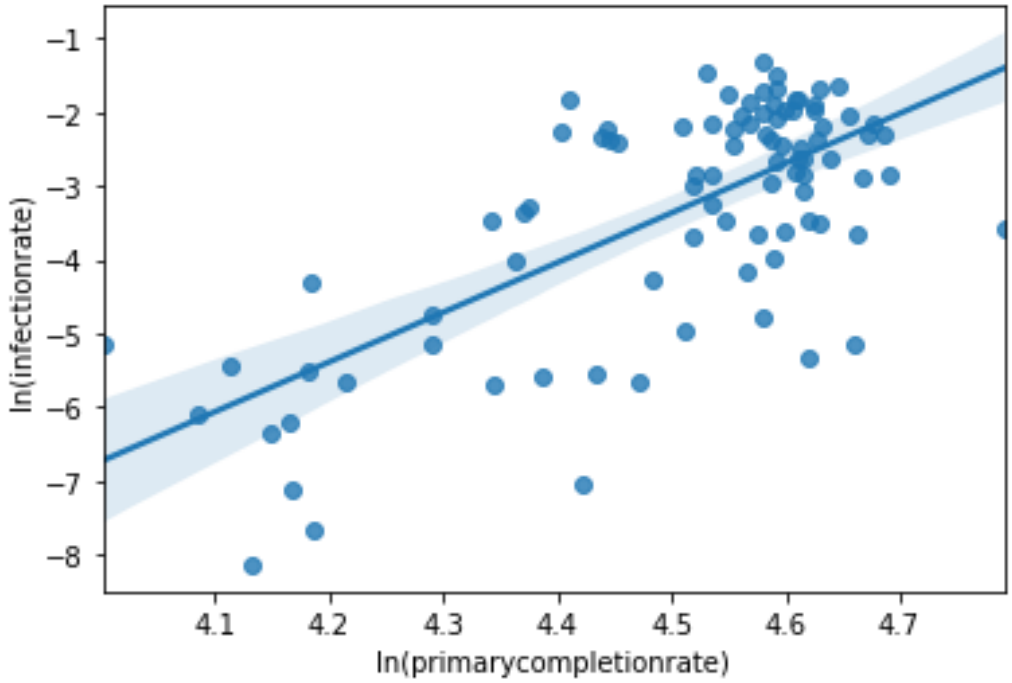
Again for model 3, there is no multicollinearity, residuals are random and normally distributed. (see Appendix)

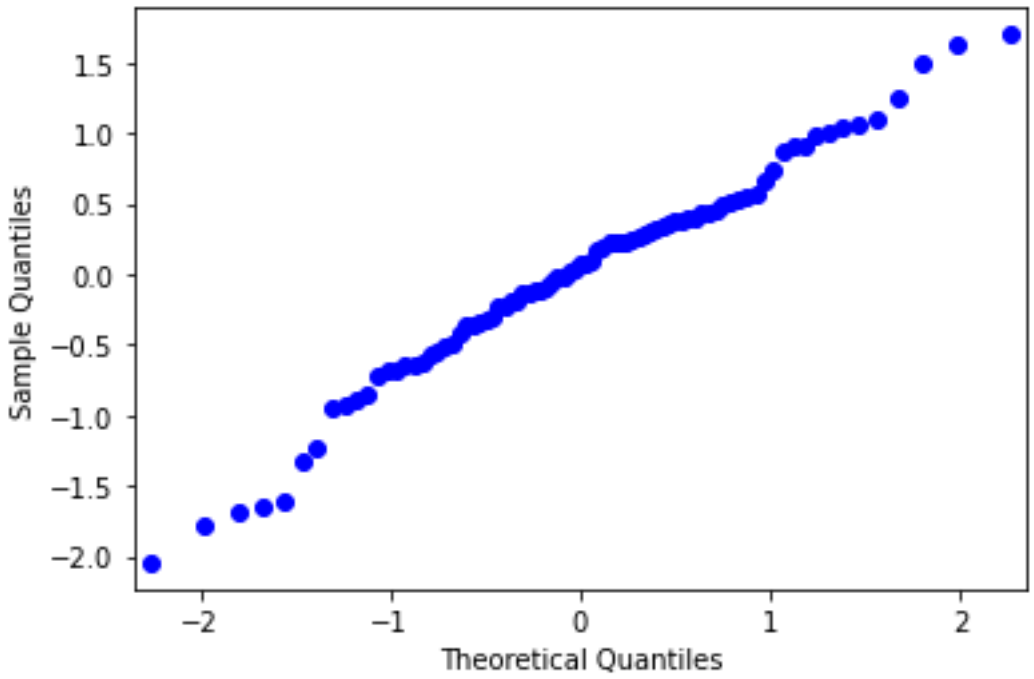
	VIFS
const	818.472805
SchoolEnrolment(tertiary)	3.535087
GDP	4.759586
eiu	2.155844
Urbanpop	1.175448
population_density	1.049660
giniWB	1.347019

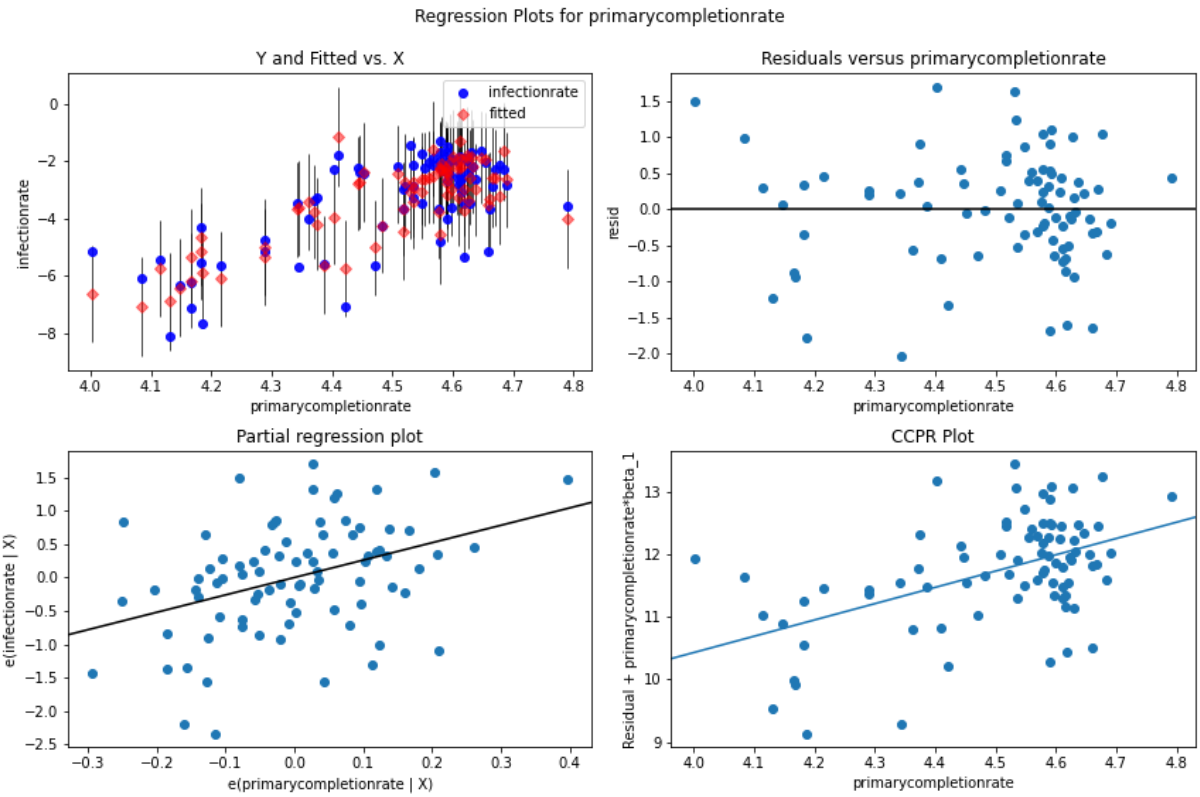
APPENDIX

Graphs-Model 2 :

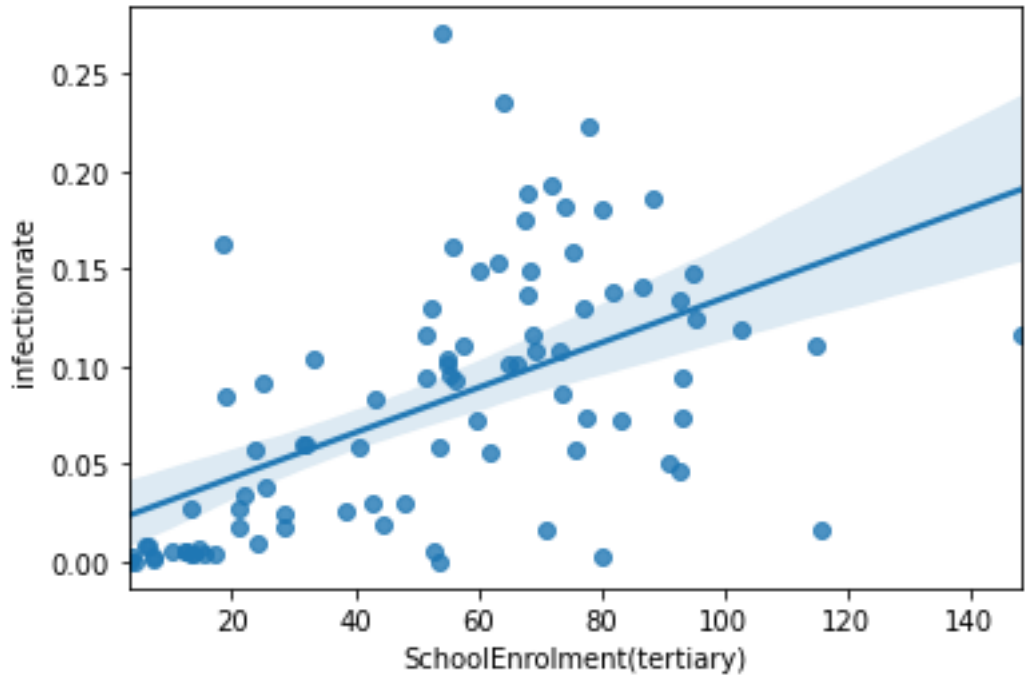


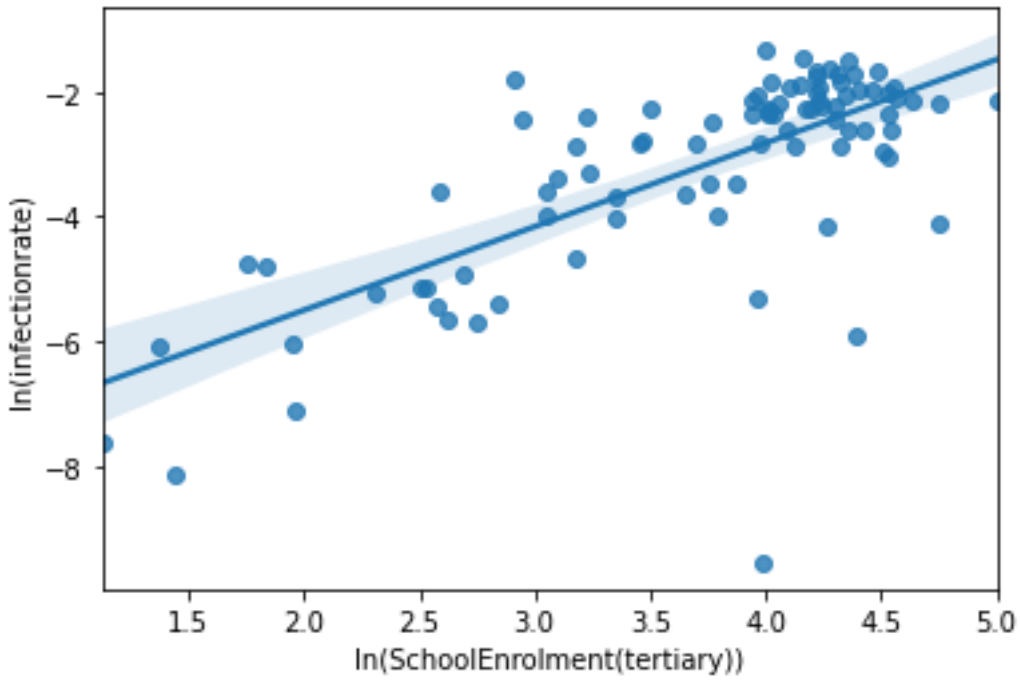

ln(infectionrate)
−1
−2
−3
−4
−5
−6
−7
−8
4.1
4.2
4.3
4.4
4.5
4.6
4.7
ln(primarycompletionrate)

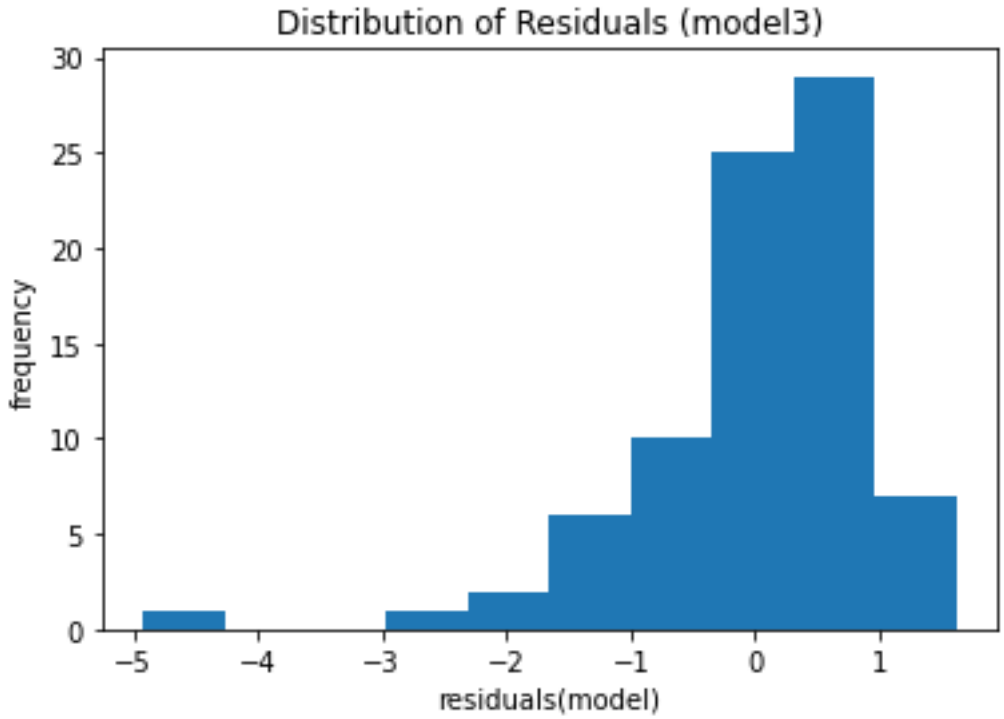

Sample Quantiles
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2
−1
0
1
2
Theoretical Quantiles

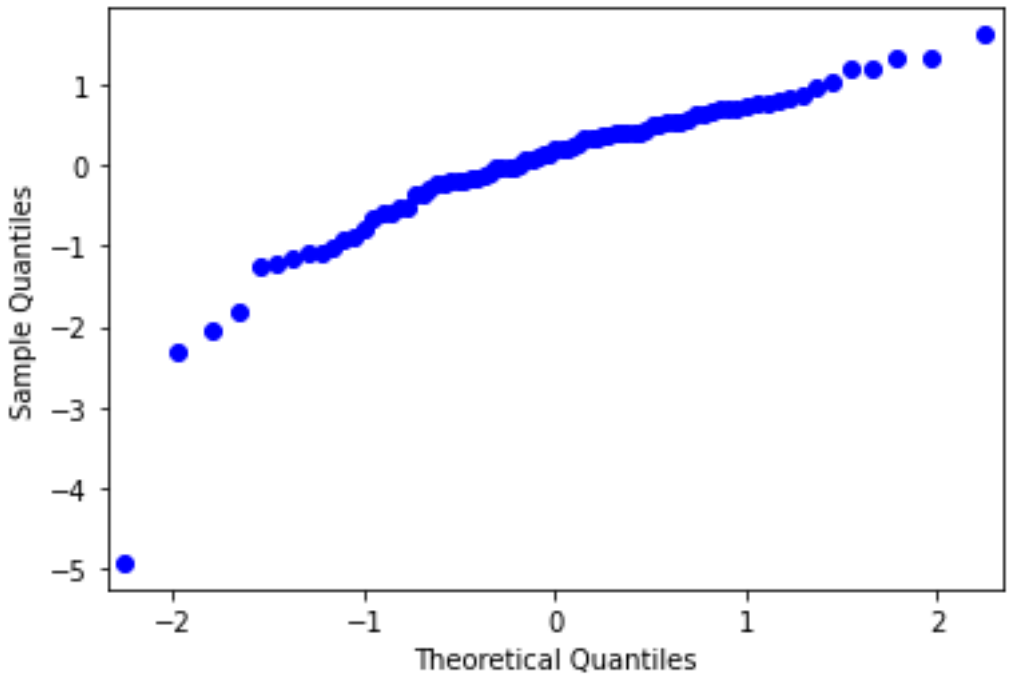


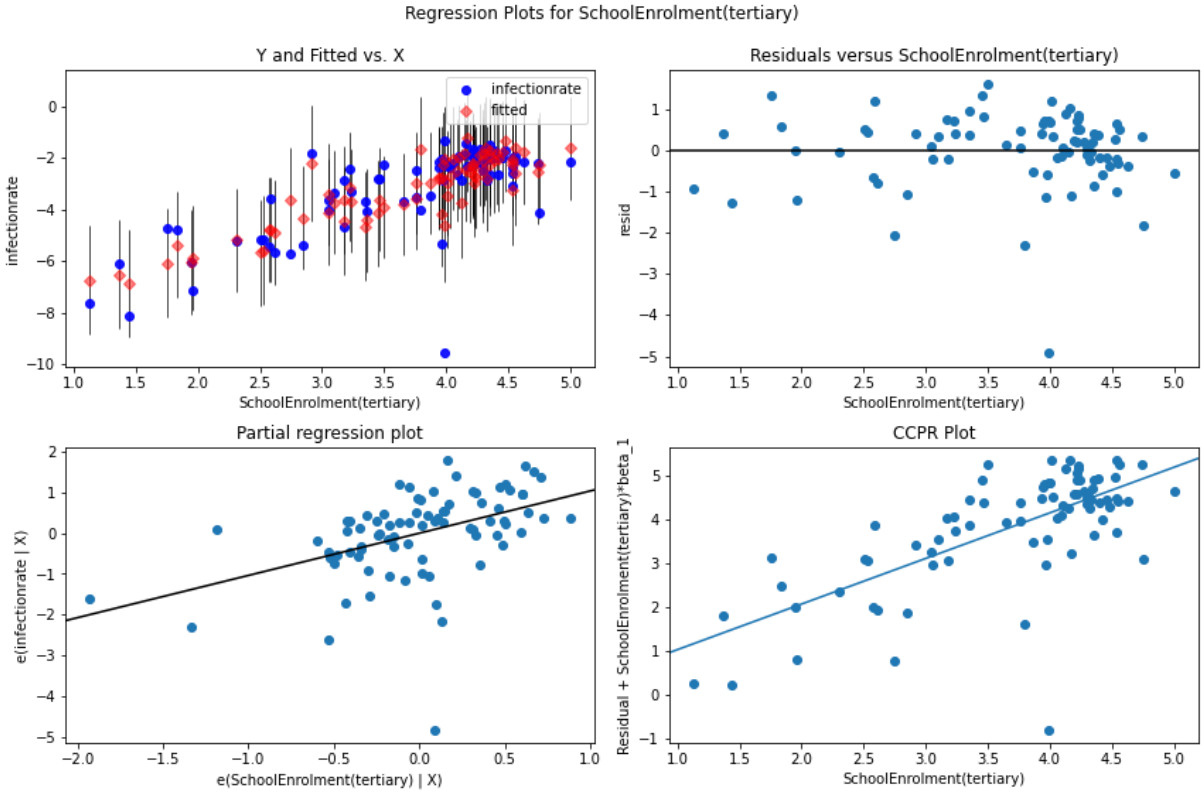
Graphs-Model 3:




ln(infectionrate)
−2
−4
−6
−8
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
ln(SchoolEnrolment(tertiary))


Distribution of Residuals (model3)
frequency
30
25
20
15
10
5
0
−5
−4
−3
−2
−1
0
1
residuals(model)


Sample Quantiles
Theoretical Quantiles


Regression Plots for SchoolEnrolment(tertiary)
Y and Fitted vs. X
infectionrate
fitted
SchoolEnrolment(tertiary)
Residuals versus SchoolEnrolment(tertiary)
resid
Partial regression plot
e(infectionrate | X)
e(SchoolEnrolment(tertiary) | X)
CCPR Plot
Residual + SchoolEnrolment(tertiary)*beta_1