Research Assignment 1

Kan Yuenyong (6220131004)

National Institute of Development Administration

Graduate School of Public Administration

PhD, Doctor of Philosophy Program in Governance and Development (International)

**1. Overview of Public Education Policy and Spending[[1]](#footnote-1)**

Redistribution of income and wealth have been widely discussed among policy makers long time ago. There are instruments such as taxation, welfare, and subsidies to achieve such policy. But we can categorize into two types which are (1) a bottom-up approach as discussed in median-voter approach or a result from pressure groups in the society, and (2) a top-down modeling that represents political elites’ interests as discussed in Mancur Olson (Lindert, 2004: 3-4). Contrary to general belief, the top-down approach dominates in poorer countries who have a lot of inequality because “*a main reason why greater inequality fails to tax the rich more is that the inequality discourages the poor from joining the fight for progressive redistributions*.” (Lindert, 2004: 8). We can trace such redistribution policy from governmental expenditure (in this study, we will use government’s education spending as representative for the social expenditure) which can be caused from either demand-side approach (which we consider Wagner’s law is one of it) or supply-side approach. This paper will demonstrate empirically on how we can employ statistical analysis to justify causal relationship between governmental expenditure and driving forces caused from either demand-side approach or supply-side approach. Six models from both approaches will be comparatively discussed in literature review and also in findings from multiple regression analysis part of this paper.

**2. Research Questions**

Given empirical data from fictitious country A (see data in Appendix 3,) what cause an increase in education spending in Country A (CE 1982 – 2007, or BE 2525 - 2550).

**3. Literature Review**

In this literature review part, we will discuss five theories used in this paper, four of which that are (1) Wagner’s law, (2) compensation theory, (3) median voter, and (4) interest group will be considered as demand-side approach; while the last theory, incrementalism, will be considered as only single theory from supply-side approach.

A German economist, Adolph Wagner, discussed in 1880s known later as “Wager’s law” that “*the development of modern industrial society would give rise to increasing political ‘pressure for social progress’ and call for increased allowance for ‘social consideration’ in the conduct of industry*,” therefore to generate an expansion of public sector and its share in the economy (Musgrave and Musgrave, 1989: 114). According to Wagner’s law, it can be assumed that when income increased, people tend to send their children to get more education and people tend to buy cars in order to satisfy their needs on travelling, the government, therefore, needs to build more road, transportation and airport and they need to make an investment in educational services to fulfil those needs from the public. Or we can say that “*industrialization, urbanization, and increased population density would give rise to a need for more provision of public facilities such as hospitals, housing, roads, and other infrastructures*” (Buracom, 2011: 119). The major weakness of this theory is that not all governments to response to those input factors from public demand, some government won’t take into the account because of different in political system.

Provided three different variables, Wagner’s Law will be tested in this paper which are the degree of urbanization, population growth, include the degree of industrialization and the real gross domestic product (Buracom, 2011: 118-119). We can consider Wagner’s law as one of the demand-side theory to observe behavior on governmental expenditure.

Compensation theory, discussed by Dani Rodrik, Robert Kaufman and Alex Segura-Ubiergo; and Geoffrey Garrett and Deborah Mitchell, epics that globalization tend to make the government to intervene the economy and to increase social program in order to absorb global economic vulnability (Buracom, 2011: 120).

Median voter theory and the demand for income distribution can be considered as on variant inside public choice school discussed by Anthony, Downs A. H. Meltzer, and S. F. Richard. According to this theory, in order to win the election, the government must response to the demands of the voters, therefore both competition for votes and the distribution of income will dictate the outcome of the election, therefore, corresponds to the governmental expenditure (Buracom, 2011: 119).

Interest group theory will discuss pressure group such as trade associations that can influence policy legislation process such as taxes, tariffs, price ceilings and regulations via campaign contribution and lobbying efforts. As found by Robert D. McCormick and Robert E. Tollison, economic regulation in the US has related accordingly to number of trade associations registered in this country (Buracom, 2011: 119).

The first four theories are all demand-side approach, here we will discuss our only single supply-side approach that is incrementalism theory. According to this theory, the policymakers will be considered as a bounded-rational agent that they can’t afford perfect information neither to get enough time to digest policy details, therefore, they will avoid possible risks and uncertainties by making a budget expenditure incrementally. A scholar such as Wildavsky sees incrementalism omnipresent in budget practice (Buracom, 2011: 122).

**4. Conceptual Framework**

This part will discuss empirically on testable model of the growth of public expenditure. Such expenditure can be caused from either demand-side or supply-side factors, the details and relationship between dependent and independent variable can be shown as in table 1, mostly adapted from Buracom (2011: 125).

Table 1: Dependent and Independent Variables used in the conceptual framework

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Expect sign | Measurement | Possible Data Source  (i.e. in Thailand) |
| Dependent Variable  GEDU | N/A | Government education spending as % of total expenditure | Bureau of the Budget |
| Independent Variable  GDP  POP  URB  GLOBAL  INEQTY  REV  TRADE  LABOR  DGOV  GEDUtm1  a  b1 – b10 | +  +  +  +  +  +  +  +  N/A  +  N/A  N/A | Gross domestic product at current price (in billion baht)  Population growth rate (%)  Growth rate of total population living in urban areas (%)  Imports and exports as % of GDP  GDP of non-agricultural sectors as a proportion to agricultural sector (a measure of inequality)  Total tax as % of GDP  % change in number of trade association (a measure of the strength of business group)  % change in number of labor union (a measure of labor strength)  Democratic government, dummy variable:   * 1 for the period of government by election * 0 for the period of government not by election, i.e. by coup d’état or revolution   One-year lagged education expenditure as % of total expenditure (measure incrementalism)  Constant  Coefficients of each independent variable with dependent variable | Bank of Thailand  National Statistics Office  Thailand Statistics Yearbook  National Economic and Social Development Boards  National Economic and Social Development Boards  Bureau of the Budget  Bank of Thailand  Ministry of Labor  Sombat Thamrongthanyawong, Politics and Government in Thailand (2005)  Refer to GEDU  N/A  N/A |

The paper will verify regression equations as follows:

Model 1: Examining any relevant independent variables[[2]](#footnote-2)

GEDU = a + b1\*GDP + b2\*POP + b3\*URB + b4\*GLOBAL + b5\*INEQTY + b6\*REV + b7\*TRADE + b8\*LABOR + b9\*DGOV + b10\*GEDUtm1

*Demand-Side Explanation*

Model 2: Wagner’s law[[3]](#footnote-3)

GEDU = a + b1\*GDP + b2\*POP + b3\*URB + b4\*REV

Model 3: Compensation Theory

GEDU = a + b1\*Global + b2\*DGOV

Model 4: Median Voter

GEDU = a + b1\*INEQTY + b2\*DGOV

Model 5: Interest Group Theory

GEDU = a + b1\*TRADE + b2\*LABOR

*Supply-Side Explanation*

Model 6: Incrementalism Theory

GEDU = a + b1\*GEDUtm1

**5. Research Methodology**

The study will verify empirical data whether it will conform to respective theories suggested in literature review. Hypothesis (identified on expected sign in table 1) will be examined thoroughly.

**5.1 Data Collection**

Data collection has been identified in table 1. Because our data is a fictitious data from country ‘A’, in table 1, we have provided some possible data source in the table accordingly.

**5.2 Data Analysis**

We will use R to analyze multiple linear regression analysis[[4]](#footnote-4) over available data extracted from data source (in this study, we will use fictitious data instead.)

**6. Analysis of the Finding from Multiple Regression**

We can identify our candidate models with different techniques, three techniques will be used in this paper, however. Firstly, from table 2, it can be summarized on empirical finding from the multiple regression that an overview comparative multiple regression table of six models range from (1) model#1, (2) model#2\_1, (3) model#3, (4) model#4, (5) model#5 and (6) model#6 via stargazer function in R, it seems candidates model per R2 or fitting method are: model#1 (R2 = 0.993)[[5]](#footnote-5), model#2.1 (R2 = 0.915) [[6]](#footnote-6), model#3 (R2 = 0.905), model#4 (R2 = 0.692), model#5 (R2 = 0.983), and model#6 (R2 = 0.910).

Table 2 (produced by stargazer function in R) and Table 4 (produced by apa.reg.table function in R) are shown as standard APA format table output to be published in standard academic journal. Another possible standard output as shown in Buracom (2011: 127) can be produced by tidy function in R or by manual writing.

The second method instead of checking from multiple model, we will indirectly verify candidate-model via looking in scatterplot matrices with correlation values produced by ggpairs function as seen in table 3. With this method, the diagonal boxes show density plot for each variable, while boxes in the lower left display the scatterplot between each variable, the boxes in the upper right corner display the Pearson correlation coefficient. From here, we can roughly identify that the independent variables such as GDP, POP, GLOBAL, INEQTY, REV, LABOR and GEDUm1 have a significant with our dependent variable GEDU. While TRADE and URB seem to have nonsignificant correlation with GEDU. DGOV has posited some degree of correlation with GEDU. From this information, we can quickly identify models that have been comprised of these variables such as model#2\_1[[7]](#footnote-7) (POP and REV), model#3 (GLOBAL and partly DGOV), model#4 (INEQTY and partly DGOV), model#5 (LABOR) and model#6 (GEDUm1) should be our candidate model. Instead of ggpairs, we may also use crplot in R to see similar result as shown in appendix 3, however the results from these two function will be very much different, albeit give similar conclusion, please consult in document such as Lanlee (2019) for details.

In the third method, we will use function summary in R to directly examine each model as shown in table 5 – 10. We’ll see later that all three techniques will give similar candidate models which are model#2\_1, model#3, model#4, model#5 and model#6.

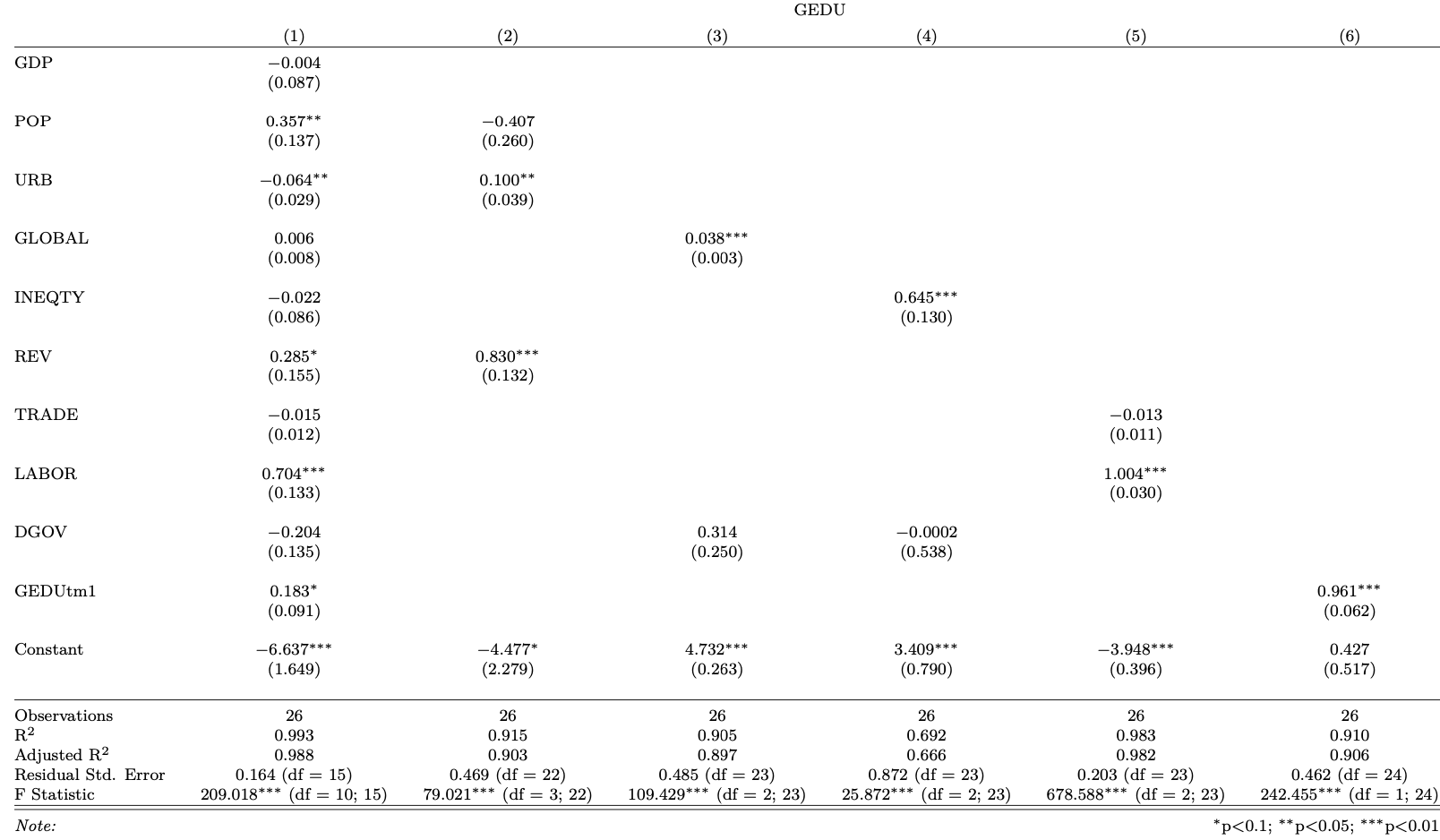
Table 2: A comparative multiple regression table of six models

Table 3: Scatterplot matrices with correlation values

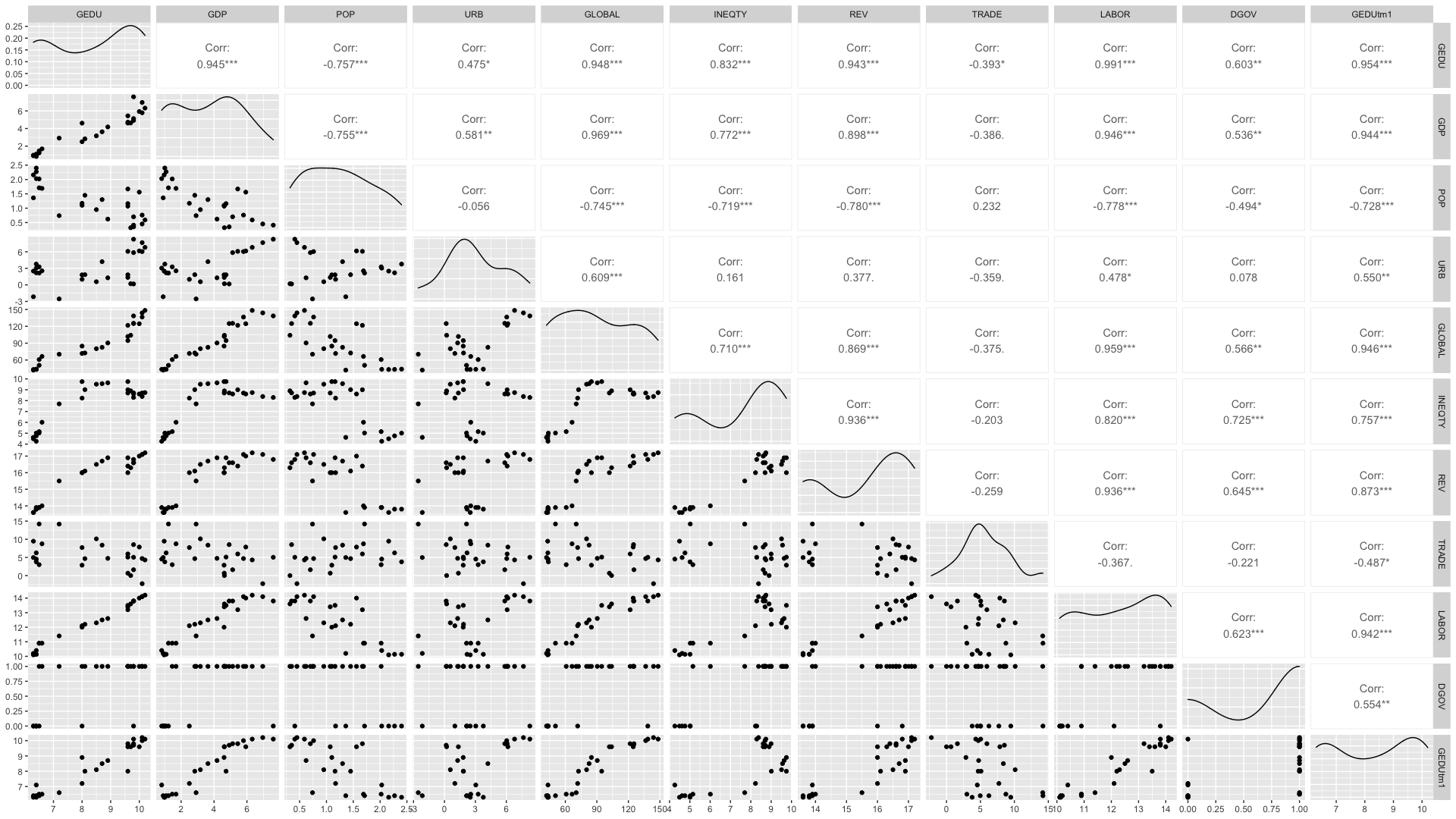


Table 4: Regression model#1 results using GEDU as the criterion

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | *b* | *b*  95% CI  [LL, UL] | *beta* | *beta*  95% CI  [LL, UL] | *sr2* | *sr2*  95% CI  [LL, UL] | *r* | Fit |
| (Intercept) | -6.64\*\* | [-10.15, -3.12] |  |  |  |  |  |  |
| GDP | -0.00 | [-0.19, 0.18] | -0.01 | [-0.26, 0.25] | .00 | [-.00, .00] | .94\*\* |  |
| POP | 0.36\* | [0.07, 0.65] | 0.15 | [0.03, 0.27] | .00 | [-.00, .01] | -.76\*\* |  |
| URB | -0.06\* | [-0.13, -0.00] | -0.12 | [-0.24, -0.00] | .00 | [-.00, .01] | .47\* |  |
| GLOBAL | 0.01 | [-0.01, 0.02] | 0.13 | [-0.26, 0.52] | .00 | [-.00, .00] | .95\*\* |  |
| INEQTY | -0.02 | [-0.21, 0.16] | -0.03 | [-0.27, 0.21] | .00 | [-.00, .00] | .83\*\* |  |
| REV | 0.28 | [-0.05, 0.62] | 0.25 | [-0.04, 0.55] | .00 | [-.00, .00] | .94\*\* |  |
| TRADE | -0.01 | [-0.04, 0.01] | -0.04 | [-0.10, 0.03] | .00 | [-.00, .00] | -.39\* |  |
| LABOR | 0.70\*\* | [0.42, 0.99] | 0.69 | [0.41, 0.96] | .01 | [.00, .03] | .99\*\* |  |
| DGOV | -0.20 | [-0.49, 0.08] | -0.06 | [-0.15, 0.03] | .00 | [-.00, .00] | .60\*\* |  |
| GEDUtm1 | 0.18 | [-0.01, 0.38] | 0.18 | [-0.01, 0.37] | .00 | [-.00, .01] | .95\*\* |  |
|  |  |  |  |  |  |  |  | *R2*  = .993\*\* |
|  |  |  |  |  |  |  |  | 95% CI[.97,.99] |
|  |  |  |  |  |  |  |  |  |

*Note.* A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. *sr2* represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.  
\* indicates *p* < .05. \*\* indicates *p* < .01.

Table 5: Finding from model#1 (Examining any relevant independent variables)

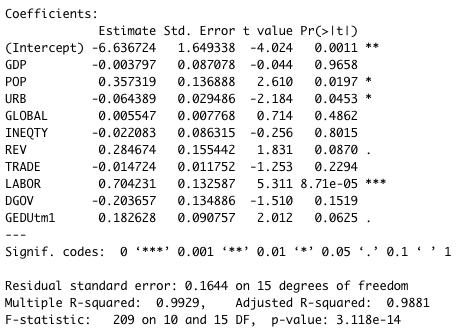


Table 6: Finding from model#2 (Wagner’s law)[[8]](#footnote-8)

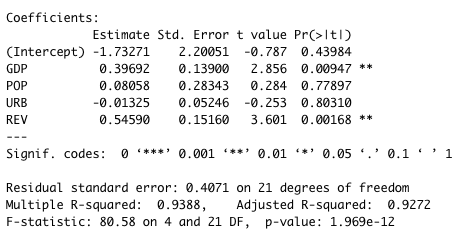
****

Table 7: Finding from model#3 (Compensation Theory)

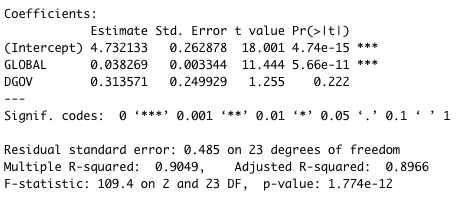
****

Table 8: Finding from model#4 (Median Voter)

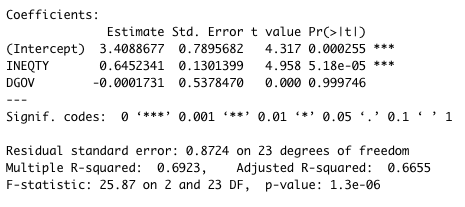
****

Table 9: Finding from model#5 (Interest Group Theory)

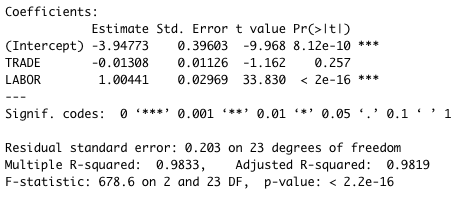
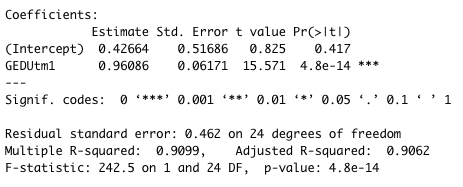
****

Table 10: Finding from model#6 (Incrementalism Theory)

****

According to R-squared, it seems that model#1 is the fittest model to the existing data (R2 = 0.9929), followed by model#5 (R2 = 0.9833), model#2 (R2 = 0.9388)[[9]](#footnote-9), model#6 (R2 = 0.9099), model#3 (R2 = 0.9049), model#4 (R2 = 0.6923) respectively. It means that from the demand-side explanation, interest group theory (model#5), Wagner’s law (model#2) and compensation theory (model#3) can explain this data well enough. Whereas the median voter theory (model#4) can partly explain the data. On the supply-side explanation, incrementalism theory (model#6) can explain the data finely either.

From interest group theory, the independent variable LABOR can explain the data better than TRADE, it might be assumed that this “fictitious country A” has a strong interest group in the labor union more than in the trade association.

From Wagner’s law, the independent variable GDP and REV are the most relevant causal explanation on increasing educational budget of the country, while POP and URB cannot demonstrate significant relate to the increasing of the educational budget. This can be implied to the direct relationship between the economic growth that can cause more tax and revenue to the government, and therefore they can increase educational budget accordingly. An increasing of population and number of people who live in urban area do not pose any impact on budget increasing which can be interpreted that this country’s status might be in developing status with significant inequality. This is in line with an explanation that why the median voter theory cannot explain this data well enough. It can refer to our earlier mention to the discussion that “*the inequality discourages the poor from joining the fight for progressive redistributions*.” (Lindert, 2004: 8).

From the compensation theory, the independent variable GLOBAL poses a strong relationship with an increasing on educational budget, while DGOV doesn’t demonstrate equal impact. It doesn’t matter how democratic regime it is, the country has to depend on global trade tightly. It has to invest in social expenditure such as in our study case, educational service to compensate with the global economic vulnerability.

The median voter theory doesn’t demonstrate explanation power in this available data, it might be, as discussed earlier, this country is still in developing status with high inequality. The poor can’t accumulate strong power enough to endure in political fight. The independent variable, INEQTY that showing significant correlation has confirm our understanding on this. If we have more data on distribution beneficial based on economic strata it can confirm more that this country is a pro-rich or pro-poor country.

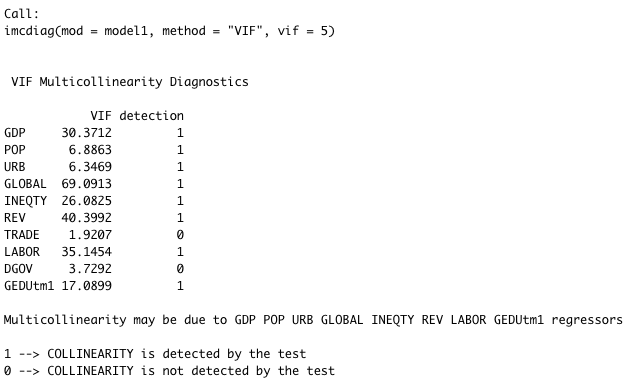
On the supply-side explanation, the incrementalism has demonstrated a strong relationship between an independent variable, GEDUtm1 or one-year lagged education expenditure to the present educational expenditure, hence incremental manner vividly.

From the overall relations of all independent variables on model#1, it seems LABOR can pose the strongest influence toward the educational expenditure. This is again, a confirmation that interest group theory can be the best explanation model to this country.

**6.1 Multicollinearity Problem**

Although our model#1 has a high R-squared per table 1, but it has a problem of multicollinearity. According to Allen (1997: 170), multicollinearity do exist “*whenever an independent variable is highly correlated with one or more of the other independent variables in a multiple regression equation*,” it poses a problem because it “*undermines the statistical significance of an independent variable. Other things being equal, the larger the standard error of a regression coefficient, the less likely it is that this coefficient will be statistically significant*.” We can perform detection of multicollinearity in R by using several methods, such as Farrar – Glauber test (enable to test in function “mctest”,) F-test (provided in function “imcdiag”,) Remedial measures (provided in function “vif”,) for example (Ghosh, 2017; Ullah and Aslam, 2020). As shown in table 8 below, our model#1 has various multicollinearity problem:

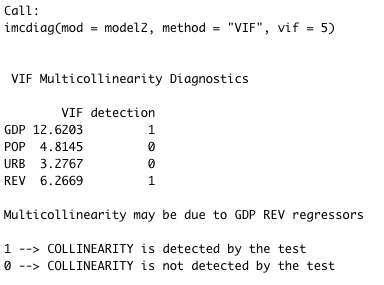
Table 11: multicollinearity detection on Model#1



From table 8, we will see that GDP, POP, URB, GLOBAL, INEQTY, REV, LABOR and GEDUtml have multicollinearity problem. It means that these independent variables have high correlation to each other, and thus they can be explained from different other independent variables. In order to resolve the multicollinearity problem, we must reject model#1 and delete some variable out of our tested model. Therefore, we break model#1 into model#2, model#3, model#4, model#5 and model#6 to test related theories accordingly. The multicollinearity can be also detected in pair-wise correlation among the explanatory variables (see screenshot of R studio in Appendix 2), this can be enable by function ggpairs in R (Ghosh, 2017).

However, we can detect similar multicollinearity problem in our model#2 either per table 9 below:

Table 12: multicollinearity detection on Model#2



From table 9, we can see that GDP and REV have a multicollinearity problem which means that they can be explained from other independent variables either. Therefore, we need to break model#2 into model#2\_1 and model#2\_2 to check GDP and REV separately. Nonetheless, from table 10, 11 and 12 below we will see that model#2\_2 also has a problem of multicollinearity in GDP, whereas in model#2\_1, REV poses no any multicollinearity problem. It means that we can use only POP, URB and REV to explain in our Wagner’s law.

Table 13: Finding from model#2\_1 (Fixing multicollinearity, Wagner’s law)

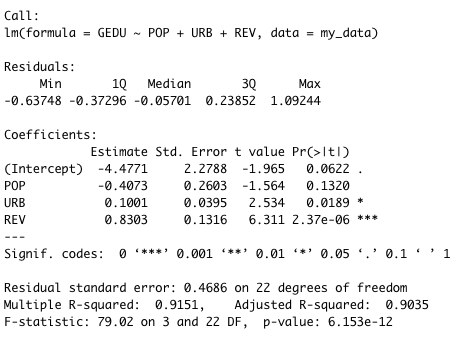
****

Table 14: multicollinearity detection on model#2\_1 (no multicollinearity problem)

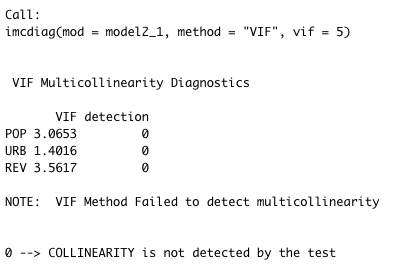
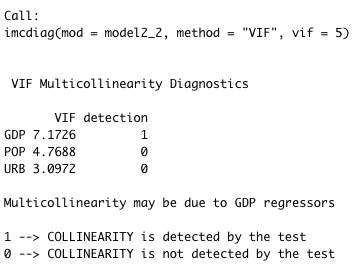
****

Table 15: multicollinearity detection on model#2\_2 (multicollinearity problem on GDP)

****

From table 10 and 11, we see no multicollinearity in our new correction of model#2\_1. REV poses no multicollinearity problem in the model with R-squared at 0.9151. An increasing of government’s income from tax revenue, therefore, can be better explained the Wagner’s law. While an increasing of GDP can’t decisively explain the model, because some other independent variables might correlate to the GDP, hence a multicollinearity problem as shown in table 12. Thus, we will reject the old model#2 and accept the new adjusted model#2\_1 regards Wagner’s law instead as of follow:

Model 2\_1: Wagner’s law (adjusted)

GEDU = a + b1\*POP + b2\*URB + b3\*REV

There is no any multicollinearity problem in model#3, #4, #5, #6.

**7. Policy Recommendation**

From our finding part, it seems that the policymakers in this country has to direct their attention toward the labor union per interest group theory, it has to put more effort to resolve a problem in inequality problem, or else it will discourage the poor from engaging in policy making process per Wagner’s law. It must make investment further in educational service and perhaps on other social expenditure as an absorbed mechanism for global economic vulnerability per compensation theory. Once the inequality has started to be eased, political parties in this country will seek to response to the voting based accordingly and therefore it should include poor people in the policymaking process, and thus to help easing the weaknesses in incrementalism policy that might blind the policymaker from a thorough understanding how the grassroots people can contribute their demand and idea in to the policymaking process and hence a better redistribution of income and wealth as of our earlier discussion on the genuine objective of this study in the end.

**8. References**

Allen, M. P. (1997). Understanding Regression Analysis. Plenum Press. New York, USA.

Buracom, P. (2011). The Determinants and Distributional Effects of Public Education, Health, and Welfare Spending in Thailand. In: Asian Affairs: An American Review, Vol 38, No 3.

Crawley, M. J. (2013). The R Book, 2nd edition. John Wiley & Sons, Ltd. West Sussex, United Kingdom.

Downey, A. B. (2015). Think Stats, 2nd edition. O’Reilly Media, Inc. CA, USA.

Ghosh, B. (2017). Regression Models in R: Multicollinearity in R. In: Data Science Plus website, <https://datascienceplus.com/multicollinearity-in-r/>, accessed on April 29, 2021

Kassambara. (2018). Multiple Linear Regression in R. In: STHDA: Statistical Tools for High-Throughput Data Analysis Website, <http://www.sthda.com/english/articles/40-regression-analysis/168-multiple-linear-regression-in-r/>, accessed on April 26, 2021.

Lanlee. (2019). Modeling and Prediction for Movies. In: RPubs by RStudio Website, <https://rpubs.com/lenlee2019/MLR2019May>, accessed on May 12, 2021.

Lindert, P. H. (2004). Growing Public: Social Spending and Economic Growth Since the Eighteenth Century, Vol 2: Further Evidence. Cambridge University Press. New York, USA.

Long, J. D. & Teetor, P. (2019). R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics, 2nd edition. O’Reilly Media, Inc. CA, USA.

Musgrave, R. A., and Musgrave, P. B. (1989). Public Finance in Theory and Practice (International Edition). McGaw-Hill. Singapore.

Ullah, M. I, and Aslam, M. (2020). Imcdiag: Individual Multicollinearity Diagnostic Measures. In: R Package Documentation Website, <https://rdrr.io/cran/mctest/man/imcdiag.html>, accessed on April 29, 2021.

**9. Appendix**

**Appendix 1: R Code Used in this study[[10]](#footnote-10)**

File: /Users/Admin/Documents/GitHub/Fiscal-Analysis/Public-Education-Policy-and-Spending/R/crPlot\_model1.R

001: # Assignment1

002:

003: # Loading

004: library("tidyverse")

005: library("readxl")

006:

007:

008: #setwd("~/Desktop/workspace/DA8410 Fiscal and Monetary Policy Analysis/DA8410 Part 2")

009: setwd("~/Documents/GitHub/Fiscal-Analysis/Public-Education-Policy-and-Spending/data")

010: # xls files

011: #my\_data <- read\_excel("DA.841,PA8603-DATA.xls")

012: my\_data <- read.csv("DA.841,PA8603-DATA\_CSV.csv")

013:

014: #data check (for debug purpose)

015: #head("my\_data", 4)

016: #print(my\_data)

017:

018: #demand-side explanation

019: #Model1: examining all independent variables

020:

021: ## old version

022: model1 <- lm(GEDU ~ GDP + POP + URB + GLOBAL + INEQTY + REV + TRADE + LABOR + DGOV + GEDUtm1, data = my\_data)

023: #summary(model1) $coefficient

024: summary(model1)

025: confint (model1)

026:

027: ## new version

028: #fitted.model1 <- lm(GEDU ~ GDP + POP + URB + GLOBAL + INEQTY + REV + TRADE + LABOR + DGOV + GEDUtm1, data = my\_data)

029: #summary(model1) $coefficient

030: #confint (model1)

031:

032:

033: #Model2: Wagner's law

034: model2 <- lm(GEDU ~ GDP + POP + URB + REV, data = my\_data)

035: model2\_1 <- lm(GEDU ~ POP + URB + REV, data = my\_data)

036: model2\_2 <- lm(GEDU ~ GDP + POP + URB, data = my\_data)

037:

038: summary(model2)

039: confint(model2)

040:

041: summary(model2\_1)

042: confint(model2\_1)

043:

044: summary(model2\_2)

045: confint(model2\_2)

046:

047:

048: #Model3: Compensation Theory

049: model3 <- lm(GEDU ~ GLOBAL + DGOV, data = my\_data)

050: summary(model3)

051: confint(model3)

052:

053: #Model4: Median Voter

054: model4 <- lm(GEDU ~ INEQTY + DGOV, data = my\_data)

055: summary(model4)

056: confint(model4)

057:

058: #Model5: Interest Group Theory

059: model5 <- lm(GEDU ~ TRADE + LABOR, data = my\_data)

060: summary(model5)

061: confint(model5)

062:

063: #Supply-Side Explanation

064: #Model 6: Incrementalism Theory

065: model6 <- lm(GEDU ~ GEDUtm1, data = my\_data)

066: summary(model6)

067: confint(model6)

068:

069:

070: #library(car)

071: #vif(model1)

072:

073: #check for multicollinearity problem

074: library(corpcor)

075: cor2pcor(cov(model1))

076:

077: library(mctest)

078: imcdiag(mod = model1, method = "VIF", vif = 5)

079: imcdiag(mod = model2, method = "VIF", vif = 5)

080: imcdiag(mod = model2\_1, method = "VIF", vif = 5)

081: imcdiag(mod = model2\_2, method = "VIF", vif = 5)

082: imcdiag(mod = model3, method = "VIF", vif = 5)

083: imcdiag(mod = model4, method = "VIF", vif = 5)

084: imcdiag(mod = model5, method = "VIF", vif = 5)

085: imcdiag(mod = model6, method = "VIF", vif = 5)

086:

087: library(GGally)

088: #ggpairs(mydata)

089: ggpairs(my\_data, c('GEDU', 'GDP', 'POP', 'URB', 'GLOBAL', 'INEQTY', 'REV', 'TRADE', 'LABOR', 'DGOV', 'GEDUtm1'))

090:

091: library(car)

092: crPlots(model1)

093: crPlots(model2\_1)

094: crPlots(model3)

095: crPlots(model4)

096: crPlots(model5)

097: crPlots(model6)

098:

099: library(apaTables)

100: apa.reg.table(model1)

101:

102: library(stargazer)

103: stargazer(model1, model2\_1, model3, model4, model5, model6, type="text")

104:

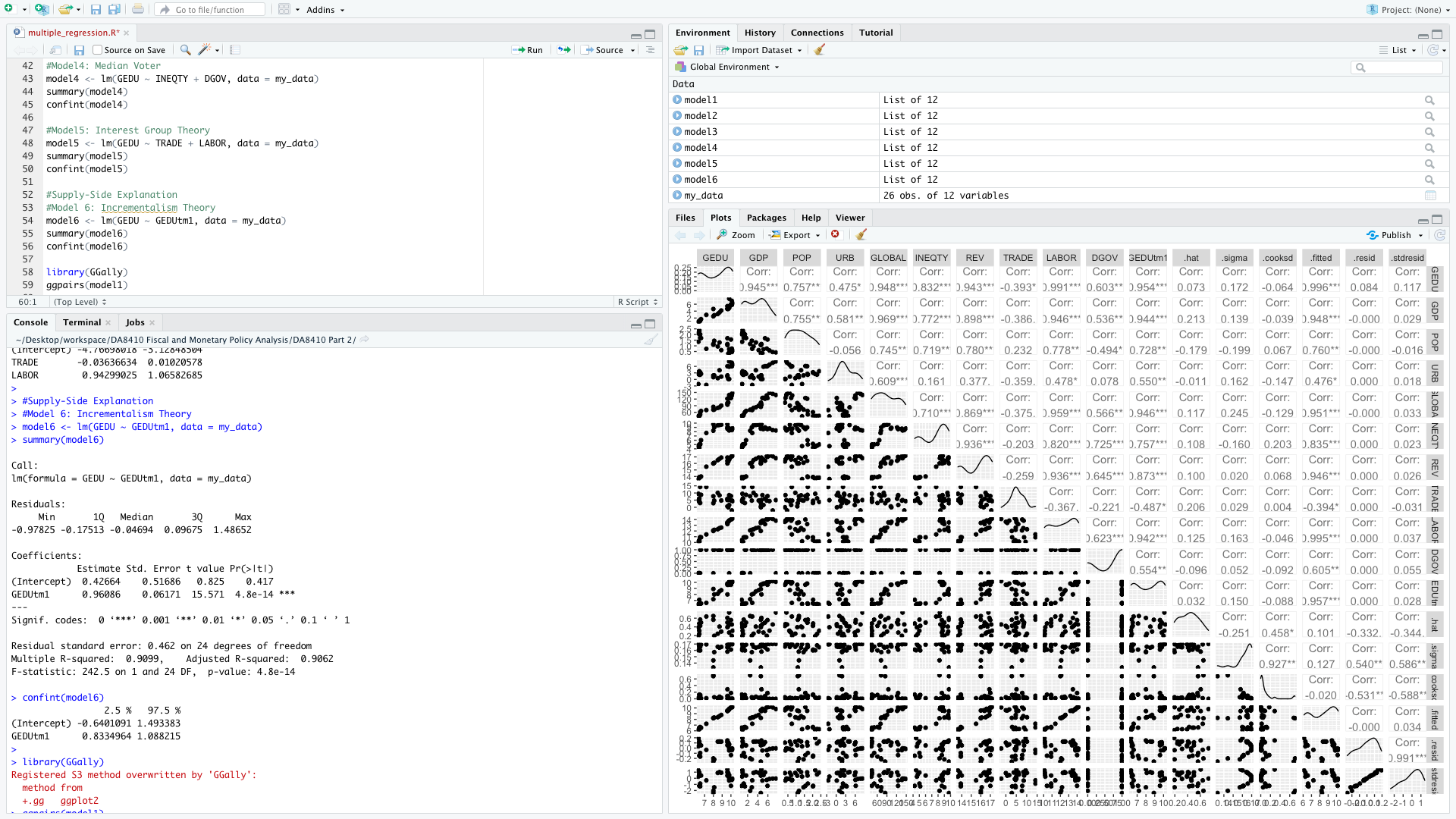
105: library(broom)

106: tidy(model1)

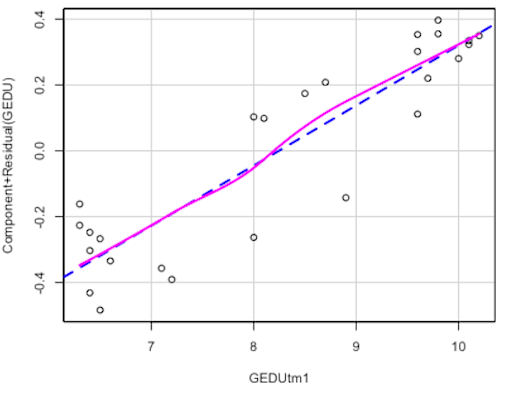
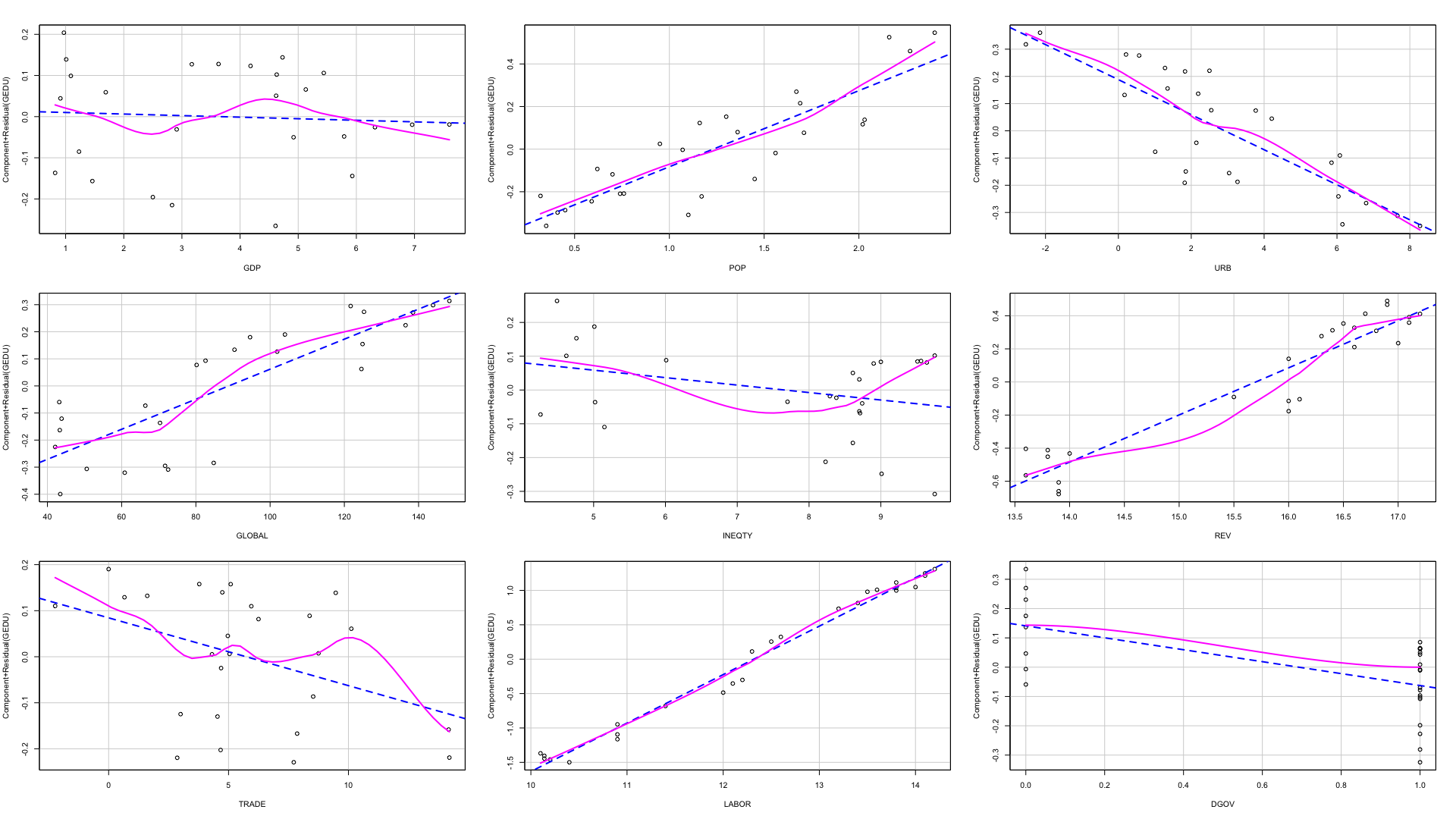
107:

108:

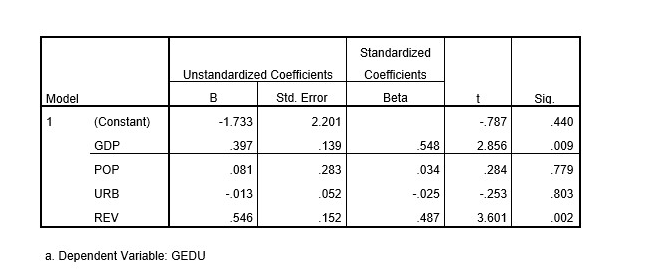
109:

**Appendix 2: R Studio[[11]](#footnote-11)’s Screen on Running the Code**

**Appendix 3: crPlot on model1**

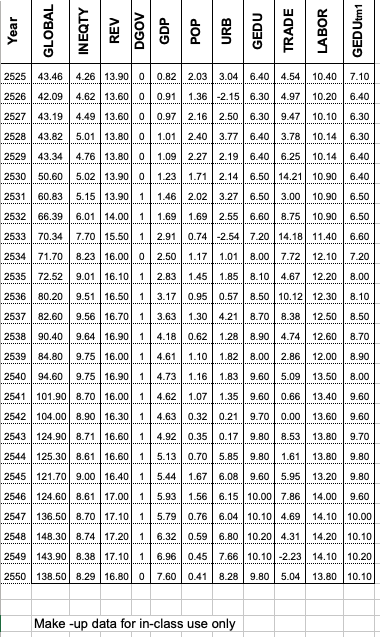


**Appendix 4: SPSS result on model 2**

****

**Appendix 5: Data Used in This Study[[12]](#footnote-12)**

Note: Data label of LOBOR has been corrected into LABOR, and GEDUt-1 has been corrected into GEDUtm1 for appropriately used in the R Studio. The original excel format has been changed into csv format for the similar purpose.



1. Related materials used in this paper whether it be this paper itself, images, source code and data can be accessed via GitHub at <https://github.com/sikkha/Fiscal-Analysis> [↑](#footnote-ref-1)
2. The model will be rejected due to multicollinearity problem see §6.1. [↑](#footnote-ref-2)
3. The model will be rejected due to multicollinearity problem. This model will be adjusted into Model 2\_1, see §6.1. [↑](#footnote-ref-3)
4. Multiple linear regression on R can be found in Long and Teeter (2019: 333-389), Downey (2015: 131-132), Crawley (2013: 489-497), and Kassambara (2018). [↑](#footnote-ref-4)
5. We will drop model#1 because of multicollinearity problem, albeit strong significant correlation, see §6.1. [↑](#footnote-ref-5)
6. We use adjusted model#2\_1 instead of original model#2 due to multicollinearity problem, see §6.1. [↑](#footnote-ref-6)
7. We drop model#2 and model#2\_2 because a dispute in variable GOV seen in multicollinearity problem, albeit a strong candidate shown in Pearson correlation coefficient in the second technique, see §6.1. [↑](#footnote-ref-7)
8. Please note that the result produced in this table is identical to a result produced by the same model in SPSS, please see appendix 4. [↑](#footnote-ref-8)
9. The model will be rejected due to multicollinearity problem. This model will be adjusted into Model 2\_1, see §6.1. [↑](#footnote-ref-9)
10. The R Source code used in this paper can be found at <https://github.com/sikkha/Fiscal-Analysis/blob/main/Public-Education-Policy-and-Spending/R/crPlot_model1.R> [↑](#footnote-ref-10)
11. R Studio’s reading materials can be found for an example at Long and Teetor (2019: 1-26). [↑](#footnote-ref-11)
12. The data used in this paper can be found at [https://github.com/sikkha/Fiscal-Analysis/blob/main/Public-Education-Policy-and-Spending/data/DA.841%2CPA8603-DATA\_CSV.csv](https://github.com/sikkha/Fiscal-Analysis/blob/main/Public-Education-Policy-and-Spending/data/DA.841%2CPA8603-DATA_CVS.csv) [↑](#footnote-ref-12)