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]

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).

[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/tree/main/Public-Education-Policy-and-Spending, the whole series can be accessed at https://github.com/sikkha/Fiscal-Analysis.

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 vulnerability (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	+	Gross domestic product at current price (in billion baht)	Bank of Thailand
POP	+	Population growth rate (%)	National Statistics Office
URB	+	Growth rate of total population living in urban areas (%)	Thailand Statistics Yearbook
GLOBAL	+	Imports and exports as % of GDP	National Economic and Social Development Boards
INEQTY	+	GDP of non-agricultural sectors as a proportion to agricultural sector (a measure of inequality)	National Economic and Social Development Boards
REV	+	Total tax as % of GDP	Bureau of the Budget
TRADE	+	% change in number of trade association (a measure of the strength of business group)	Bank of Thailand
LABOR	+	% change in number of labor union (a measure of labor strength)	Ministry of Labor
DGOV	N/A	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	Sombat Thamrongthanyawong, Politics and Government in Thailand (2005)
GEDUtm1	+	One-year lagged education expenditure as % of total expenditure (measure incrementalism)	Refer to GEDU
α	N/A	Constant	N/A
$\beta_1 - \beta_{10}$	N/A	Coefficients of each independent variable with dependent variable	N/A
ε	N/A	Error term	N/A

The paper will verify regression equations as follows:

Model 1: Examining any relevant independent variables[2]

$$GEDU = \alpha + \beta_1(GDP) + \beta_2(POP) + \beta_3(URB) + \beta_4(GLOBAL) + \beta_5(INEQTY) + \beta_6(REV) + \beta_7(TRADE) + \beta_8(LABOR) + \beta_9(DGOV) + \beta_{10}(GEDUtm1) + \varepsilon$$

Demand-Side Explanation

Model 2: Wagner's law[3]
$$GEDU = \alpha + \beta_1(GDP) + \beta_2(POP) + \beta_3(URB) + \beta_4(REV) + \varepsilon$$

Model 3: Compensation Theory
$$GEDU = \alpha + \beta_1(Global) + \beta_2(DGOV) + \varepsilon$$

Model 4: Median Voter
$$GEDU = \alpha + \beta_1(INEQTY) + \beta_2(DGOV) + \varepsilon$$

Model 5: Interest Group Theory
$$GEDU = \alpha + \beta_1(TRADE) + \beta_2(LABOR) + \varepsilon$$

Supply-Side Explanation

Model 6: Incrementalism Theory
$$GEDU = \alpha + \beta_1(GEDUtm1) + \varepsilon$$

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 and data source have been identified in table 1. Because our data is a fictitious data from country 'A' shown in appendix 5.

[2] The model will be rejected due to multicollinearity problem see §6.1.
[3] The model will be rejected due to multicollinearity problem. This model will be adjusted into Model 2_1, see §6.1.

5.2 Data Analysis

We will use R to analyze multiple linear regression analysis[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 R^2 or fitting method are: model#1 (R^2 = 0.993)[5], model#2.1 (R^2 = 0.915)[6], model#3 (R^2 = 0.905), model#4 (R^2 = 0.692), model#5 (R^2 = 0.983), and model#6 (R^2 = 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] (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.

---

[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).

[5] We will drop model#1 because of multicollinearity problem, albeit strong significant correlation, see §6.1.

[6] We use adjusted model#2_1 instead of original model#2 due to multicollinearity problem, see §6.1.

[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.

Table 2: A comparative multiple regression table of six models

	GEDU					
	(1)	(2)	(3)	(4)	(5)	(6)
GDP	−0.004					
	(0.087)					
POP	0.357**	−0.407				
	(0.137)	(0.260)				
URB	−0.064**	0.100**				
	(0.029)	(0.039)				
GLOBAL	0.006		0.038***			
	(0.008)		(0.003)			
INEQTY	−0.022			0.645***		
	(0.086)			(0.130)		
REV	0.285*	0.830***				
	(0.155)	(0.132)				
TRADE	−0.015				−0.013	
	(0.012)				(0.011)	
LABOR	0.704***				1.004***	
	(0.133)				(0.030)	
DGOV	−0.204		0.314	−0.0002		
	(0.135)		(0.250)	(0.538)		
GEDUtm1	0.183*					0.961***
	(0.091)					(0.062)
Constant	−6.637***	−4.477*	4.732***	3.409***	−3.948***	0.427
	(1.649)	(2.279)	(0.263)	(0.790)	(0.396)	(0.517)
Observations	26	26	26	26	26	26
R^2	0.993	0.915	0.905	0.692	0.983	0.910
Adjusted R^2	0.988	0.903	0.897	0.666	0.982	0.906
Residual Std. Error	0.164 (df = 15)	0.469 (df = 22)	0.485 (df = 23)	0.872 (df = 23)	0.203 (df = 23)	0.462 (df = 24)
F Statistic	209.018*** (df = 10; 15)	79.021*** (df = 3; 22)	109.429*** (df = 2; 23)	25.872*** (df = 2; 23)	678.588*** (df = 2; 23)	242.455*** (df = 1; 24)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 3: Scatterplot matrices with correlation values

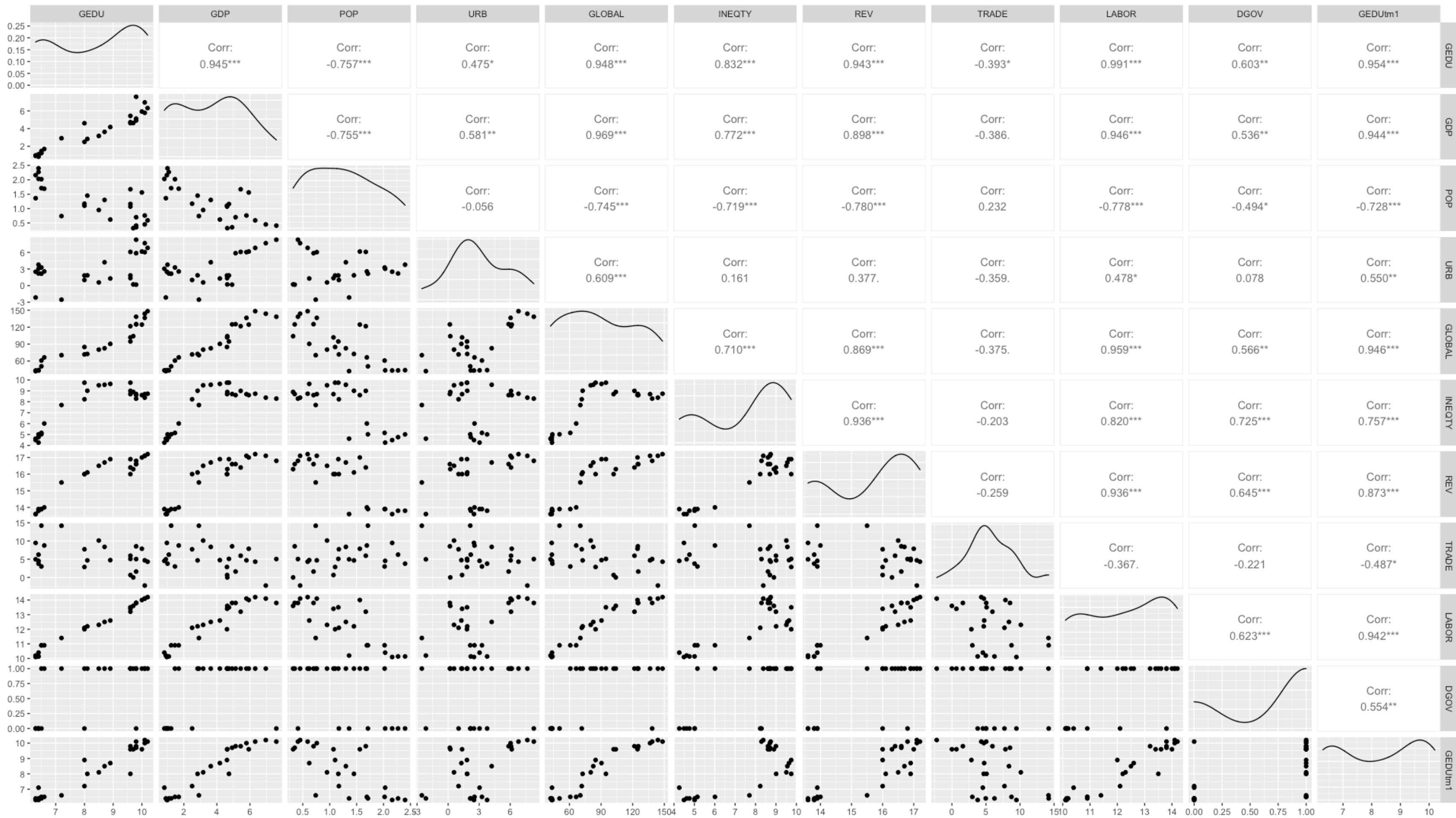


Table 4: Regression model#1 results using GEDU as the criterion

Predictor	b	b 95% CI [LL, UL]	$beta$	$beta$ 95% CI [LL, UL]	sr^2	sr^2 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**	
								R^2 = .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. sr^2 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)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.636724   1.649338  -4.024   0.0011 **
GDP         -0.003797   0.087078  -0.044   0.9658
POP          0.357319   0.136888   2.610   0.0197 *
URB         -0.064389   0.029486  -2.184   0.0453 *
GLOBAL       0.005547   0.007768   0.714   0.4862
INEQTY      -0.022083   0.086315  -0.256   0.8015
REV          0.284674   0.155442   1.831   0.0870 .
TRADE       -0.014724   0.011752  -1.253   0.2294
LABOR        0.704231   0.132587   5.311 8.71e-05 ***
DGOV        -0.203657   0.134886  -1.510   0.1519
GEDUtm1      0.182628   0.090757   2.012   0.0625 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1644 on 15 degrees of freedom
Multiple R-squared:  0.9929,    Adjusted R-squared:  0.9881
F-statistic:   209 on 10 and 15 DF,  p-value: 3.118e-14
```

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

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.73271    2.20051  -0.787  0.43984
GDP          0.39692    0.13900   2.856  0.00947 **
POP          0.08058    0.28343   0.284  0.77897
URB         -0.01325    0.05246  -0.253  0.80310
REV          0.54590    0.15160   3.601  0.00168 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4071 on 21 degrees of freedom
Multiple R-squared:  0.9388,    Adjusted R-squared:  0.9272
F-statistic: 80.58 on 4 and 21 DF,  p-value: 1.969e-12
```

[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.

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

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.732133   0.262878  18.001 4.74e-15 ***
GLOBAL      0.038269   0.003344  11.444 5.66e-11 ***
DGOV        0.313571   0.249929   1.255    0.222
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.485 on 23 degrees of freedom
Multiple R-squared:  0.9049,    Adjusted R-squared:  0.8966
F-statistic: 109.4 on 2 and 23 DF,  p-value: 1.774e-12
```

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

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.4088677  0.7895682   4.317 0.000255 ***
INEQTY       0.6452341  0.1301399   4.958 5.18e-05 ***
DGOV        -0.0001731  0.5378470   0.000 0.999746
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8724 on 23 degrees of freedom
Multiple R-squared:  0.6923,    Adjusted R-squared:  0.6655
F-statistic: 25.87 on 2 and 23 DF,  p-value: 1.3e-06
```

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

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.94773    0.39603  -9.968 8.12e-10 ***
TRADE       -0.01308    0.01126  -1.162    0.257
LABOR        1.00441    0.02969  33.830  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.203 on 23 degrees of freedom
Multiple R-squared:  0.9833,    Adjusted R-squared:  0.9819
F-statistic: 678.6 on 2 and 23 DF,  p-value: < 2.2e-16
```

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

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.42664    0.51686   0.825    0.417
GEDUtm1      0.96086    0.06171  15.571  4.8e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.462 on 24 degrees of freedom
Multiple R-squared:  0.9099,    Adjusted R-squared:  0.9062
F-statistic: 242.5 on 1 and 24 DF,  p-value: 4.8e-14
```

According to R-squared, it seems that model#1 is the fittest model to the existing data ($R^2 = 0.9929$), followed by model#5 ($R^2 = 0.9833$), model#2 ($R^2 = 0.9388$)[9], model#6 ($R^2 = 0.9099$), model#3 ($R^2 = 0.9049$), model#4 ($R^2 = 0.6923$) 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.

[9] The model will be rejected due to multicollinearity problem. This model will be adjusted into Model 2_1, see §6.1.

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 2, 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

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


 VIF Multicollinearity Diagnostics

            VIF detection
GDP     30.3712         1
POP      6.8863         1
URB      6.3469         1
GLOBAL  69.0913         1
INEQTY  26.0825         1
REV     40.3992         1
TRADE    1.9207         0
LABOR   35.1454         1
DGOV     3.7292         0
GEDUtm1 17.0899         1

Multicollinearity may be due to GDP POP URB GLOBAL INEQTY REV LABOR GEDUtm1 regressors

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

From table 11, 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 table 3 and 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 12 below:

Table 12: multicollinearity detection on Model#2

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


 VIF Multicollinearity Diagnostics

        VIF detection
GDP 12.6203         1
POP  4.8145         0
URB  3.2767         0
REV  6.2669         1

Multicollinearity may be due to GDP REV regressors

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

From table 12 above, 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 13, 14 and 15 below we will see that model#2_2 also has a problem of multicollinearity in GDP (per table 15,) whereas in model#2_1 (per table 14,) 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)

```
Call:
lm(formula = GEDU ~ POP + URB + REV, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.63748 -0.37296 -0.05701  0.23852  1.09244 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -4.4771     2.2788  -1.965   0.0622 .  
POP          -0.4073     0.2603  -1.564   0.1320    
URB           0.1001     0.0395   2.534   0.0189 *  
REV           0.8303     0.1316   6.311 2.37e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4686 on 22 degrees of freedom
Multiple R-squared:  0.9151,	Adjusted R-squared:  0.9035 
F-statistic: 79.02 on 3 and 22 DF,  p-value: 6.153e-12
```

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

```
Call:
imcdiag(mod = model2_1, method = "VIF", vif = 5)


 VIF Multicollinearity Diagnostics

       VIF detection
POP 3.0653         0
URB 1.4016         0
REV 3.5617         0

NOTE:  VIF Method Failed to detect multicollinearity


0 --> COLLINEARITY is not detected by the test
```

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

```
Call:
imcdiag(mod = model2_2, method = "VIF", vif = 5)


 VIF Multicollinearity Diagnostics

       VIF detection
GDP 7.1726         1
POP 4.7688         0
URB 3.0972         0

Multicollinearity may be due to GDP regressors

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

From table 13 and 14, 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 15. 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 = $\alpha + \beta_1(POP) + \beta_2(URB) + \beta_3(REV) + \varepsilon$

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]

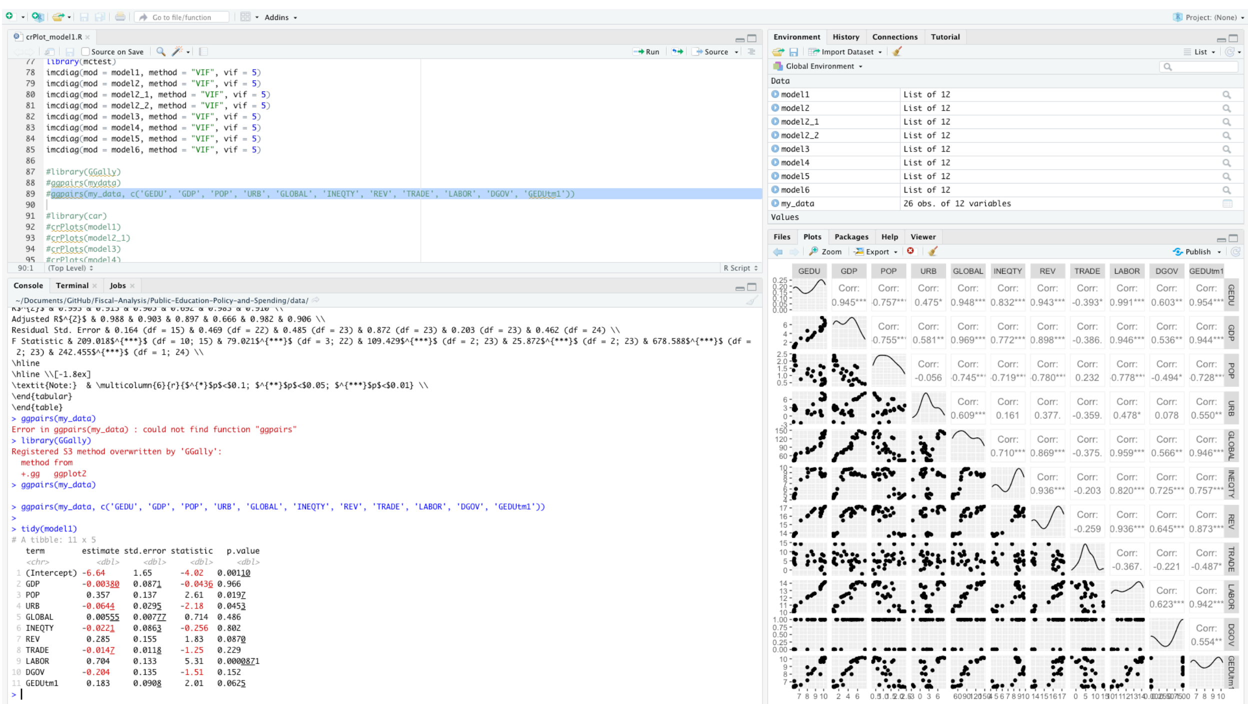
```
File: /Users/Admin/Documents/GitHub/Fiscal-Analysis/Public-Education-Policy-and-Spending/R/crPlot_model1.R
# Assignment1

# Loading
library("tidyverse")
library("readxl")


#setwd("~/Desktop/workspace/DA8410 Fiscal and Monetary Policy Analysis/DA8410 Part 2")
setwd("~/Documents/GitHub/Fiscal-Analysis/Public-Education-Policy-and-Spending/data")
# xls files
#my_data <- read_excel("DA.841,PA8603-DATA.xls")
my_data <- read.csv("DA.841,PA8603-DATA_CSV.csv")

#data check (for debug purpose)
#head("my_data", 4)
#print(my_data)

#demand-side explanation
#Model1: examining all independent variables

## old version
model1 <- lm(GEDU ~ GDP + POP + URB + GLOBAL + INEQTY + REV + TRADE + LABOR + DGOV + GEDUtm1, data = my_data)
#summary(model1) $coefficient
summary(model1)
confint (model1)

## new version
#fitted.model1 <- lm(GEDU ~ GDP + POP + URB + GLOBAL + INEQTY + REV + TRADE + LABOR + DGOV + GEDUtm1, data = my_data)
#summary(model1) $coefficient
#confint (model1)


#Model2: Wagner's law
model2 <- lm(GEDU ~ GDP + POP + URB + REV, data = my_data)
model2_1 <- lm(GEDU ~ POP + URB + REV, data = my_data)
model2_2 <- lm(GEDU ~ GDP + POP + URB, data = my_data)

summary(model2)
confint(model2)

```

[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

```
summary(model2_1)
confint(model2_1)
summary(model2_2)
confint(model2_2)
#Model3: Compensation Theory
model3 <- lm(GEDU ~ GLOBAL + DGOV, data = my_data)
summary(model3)
confint(model3)
#Model4: Median Voter
model4 <- lm(GEDU ~ INEQTY + DGOV, data = my_data)
summary(model4)
confint(model4)
#Model5: Interest Group Theory
model5 <- lm(GEDU ~ TRADE + LABOR, data = my_data)
summary(model5)
confint(model5)
#Supply-Side Explanation
#Model 6: Incrementalism Theory
model6 <- lm(GEDU ~ GEDUtm1, data = my_data)
summary(model6)
confint(model6)
#library(car)
#vif(model1)
#check for multicollinearity problem
library(corpcor)
cor2pcor(cov(model1))
library(mctest)
imcdiag(mod = model1, method = "VIF", vif = 5)
imcdiag(mod = model2, method = "VIF", vif = 5)
imcdiag(mod = model2_1, method = "VIF", vif = 5)
imcdiag(mod = model2_2, method = "VIF", vif = 5)
imcdiag(mod = model3, method = "VIF", vif = 5)
imcdiag(mod = model4, method = "VIF", vif = 5)
imcdiag(mod = model5, method = "VIF", vif = 5)
imcdiag(mod = model6, method = "VIF", vif = 5)
#library(GGally)
#ggpairs(mydata)
#ggpairs(my_data, c('GEDU', 'GDP', 'POP', 'URB', 'GLOBAL', 'INEQTY', 'REV', 'TRADE', 'LABOR', 'DGOV', 'GEDUtm1'))
#library(car)
#crPlots(model1)
#crPlots(model2_1)
#crPlots(model3)
#crPlots(model4)
```

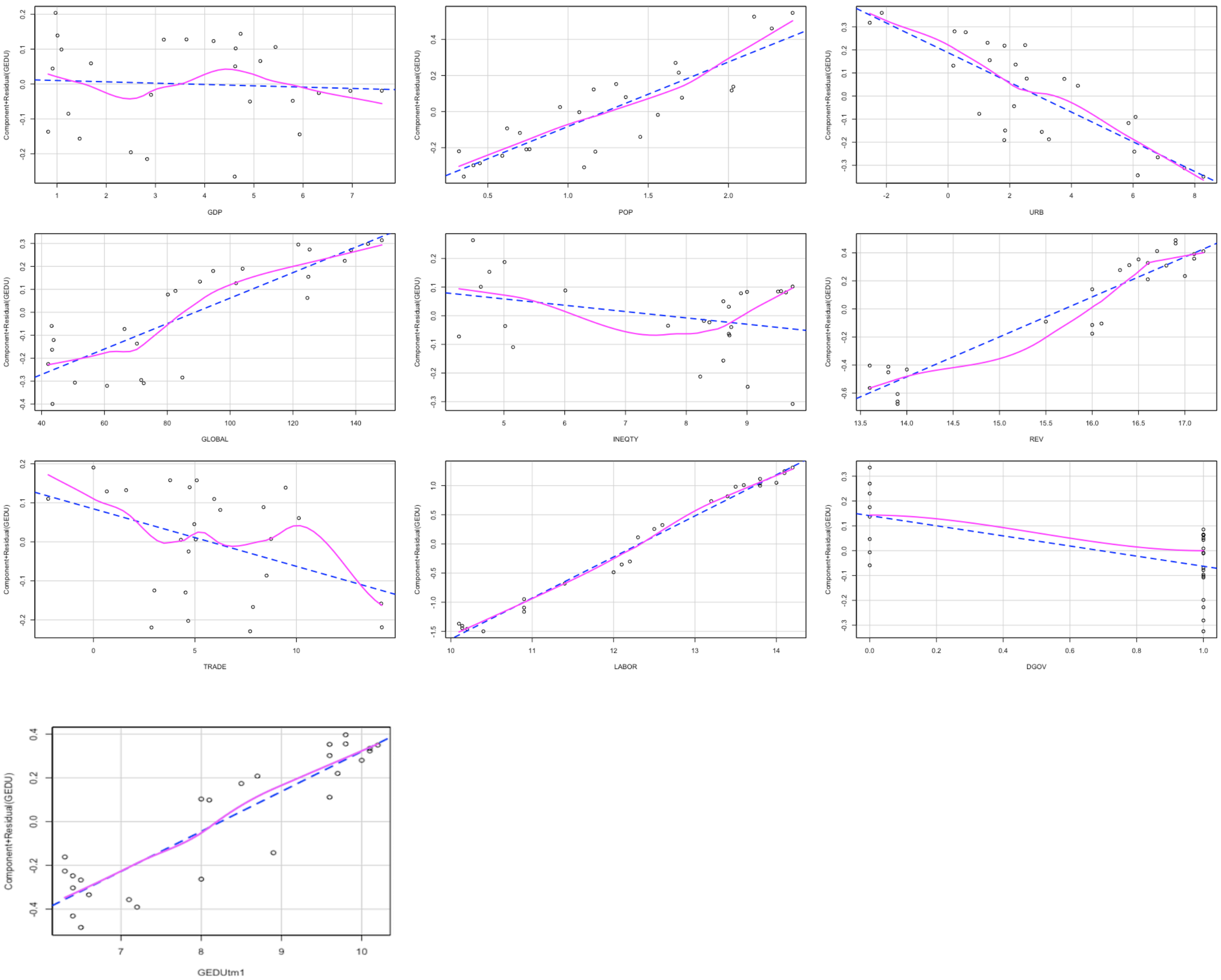
```
#crPlots(model5)
#crPlots(model6)

library(apaTables)
apa.reg.table(model1)

library(stargazer)
stargazer(model1, model2_1, model3, model4, model5, model6, type="text")

library(broom)
tidy(model1)

#mathmatical derived by equatiamatic
library(equatiomatic)
cat(extract_eq(model1))
```

Appendix 2: R Studio[11]'s Screen on Running the Code



[11] R Studio's reading materials can be found for an example at Long and Teetor (2019: 1-26).

Appendix 3: crPlot on model1



Appendix 4: SPSS result on model 2[12]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.733	2.201		-.787	.440
	GDP	.397	.139	.548	2.856	.009
	POP	.081	.283	.034	.284	.779
	URB	-.013	.052	-.025	-.253	.803
	REV	.546	.152	.487	3.601	.002

a. Dependent Variable: GEDU

[12] Please compare output of model#2 from this SPSS with R output in table 6, their result are identical.

Appendix 5: Data Used in This Study[13]

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.

Year	GLOBAL	INEQTY	REV	DGOV	GDP	POP	URB	GEDU	TRADE	LABOR	GEDUtm1
2525	43.46	4.26	13.90	0	0.82	2.03	3.04	6.40	4.54	10.40	7.10
2526	42.09	4.62	13.60	0	0.91	1.36	-2.15	6.30	4.97	10.20	6.40
2527	43.19	4.49	13.60	0	0.97	2.16	2.50	6.30	9.47	10.10	6.30
2528	43.82	5.01	13.80	0	1.01	2.40	3.77	6.40	3.78	10.14	6.30
2529	43.34	4.76	13.80	0	1.09	2.27	2.19	6.40	6.25	10.14	6.40
2530	50.60	5.02	13.90	0	1.23	1.71	2.14	6.50	14.21	10.90	6.40
2531	60.83	5.15	13.90	1	1.46	2.02	3.27	6.50	3.00	10.90	6.50
2532	66.39	6.01	14.00	1	1.69	1.69	2.55	6.60	8.75	10.90	6.50
2533	70.34	7.70	15.50	1	2.91	0.74	-2.54	7.20	14.18	11.40	6.60
2534	71.70	8.23	16.00	0	2.50	1.17	1.01	8.00	7.72	12.10	7.20
2535	72.52	9.01	16.10	1	2.83	1.45	1.85	8.10	4.67	12.20	8.00
2536	80.20	9.51	16.50	1	3.17	0.95	0.57	8.50	10.12	12.30	8.10
2537	82.60	9.56	16.70	1	3.63	1.30	4.21	8.70	8.38	12.50	8.50
2538	90.40	9.64	16.90	1	4.18	0.62	1.28	8.90	4.74	12.60	8.70
2539	84.80	9.75	16.00	1	4.61	1.10	1.82	8.00	2.86	12.00	8.90
2540	94.60	9.75	16.90	1	4.73	1.16	1.83	9.60	5.09	13.50	8.00
2541	101.90	8.70	16.00	1	4.62	1.07	1.35	9.60	0.66	13.40	9.60
2542	104.00	8.90	16.30	1	4.63	0.32	0.21	9.70	0.00	13.60	9.60
2543	124.90	8.71	16.60	1	4.92	0.35	0.17	9.80	8.53	13.80	9.70
2544	125.30	8.61	16.60	1	5.13	0.70	5.85	9.80	1.61	13.80	9.80
2545	121.70	9.00	16.40	1	5.44	1.67	6.08	9.60	5.95	13.20	9.80
2546	124.60	8.61	17.00	1	5.93	1.56	6.15	10.00	7.86	14.00	9.60
2547	136.50	8.70	17.10	1	5.79	0.76	6.04	10.10	4.69	14.10	10.00
2548	148.30	8.74	17.20	1	6.32	0.59	6.80	10.20	4.31	14.20	10.10
2549	143.90	8.38	17.10	1	6.96	0.45	7.66	10.10	-2.23	14.10	10.20
2550	138.50	8.29	16.80	0	7.60	0.41	8.28	9.80	5.04	13.80	10.10

Make -up data for in-class use only

[13] 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