Efficiency of Public Expenditure in Education: Peru Case

Two Step Analysis: Data Envelopment Analysis (DEA) and Panel Data

Created by: Eduardo Paredes

10/12/2021

Contents

Data and Variables

- tasadesnutricion: Chronic malnutrition, kids under 5 years old, by region, 2009-2019, Source: INEI (Percentage)

- docentes: Teachers of public sector, by region, 2008-2009. (number)

- alumnosmatriculados: Students of secundary education, by region, 2009-2019 (thousands)

- analfabetismo: Rate of analfabetism of population under 15 years old, by region, 2008 -2019 (Percentage)

- gastopublico: Public expenditure by student en regular basic education, by educational level and region, 2008-2018 (nuevos soles, PEN)

- pbiregiones: GDP per capita. by year and region, 2007-2020 (constant prices 2007 thousands of soles), then divided by pobtotal.

- pobtotal: Total population registered and omitted, by census year/inter census years estimated (number)

- poburb: Census population of urban areas, by region and census year (number)

- tasamatricula: School enrollment of secondary education between 12 and 16 years old, by region, 2008-2019 (Rate)

- tasaconclusion: Percentage of people that concluded secondary education, population older than 15 years old, by region, 2008-2019 (percentage)

- complectora: Mean of grade of reading. Secondary level, by region. Educational Census of Peru(Number)

- matematica: Mean of grade of mathematics. Secondary level, by region. Educational Census of Peru (Number)

- alumDoc: Rate of Students per teacher, by region and by year (rate)

- nivelUrb: Percentage of urban population. popUrb divided by popTotal (percentage)

Only 2015, 2016 and 2018 were taken as sample due to Educational Census limitations.

Logic Behind

The study aims to rank the regions of Peru by its efficiency expenditure in Education. This variable is calculated by DEA procedure, where the output is compLectora, matematica, tasaMatricula, tasaConclusion and the input is gastoPublico.

After DEA, the paper seeks to find the impact of some variables on the efficiency using a Panel Data.

Further econometric, theory explanation and details were explained in a meeting.

Data Envelopment Analysis (DEA) - First Step

## Descriptive Analysis

Table 1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
gastoPublico	78	3,896.834	984.882	2,470.946	3,167.269	4,544.244	7,172.360
compLectora	78	12.675	7.130	3.354	7.227	15.011	31.107
matematica	78	10.493	6.809	1.327	5.978	12.600	31.340
tasaMatricula	78	84.063	5.438	69.916	80.838	88.476	92.904
tasaConclusion	78	43.930	4.974	31.936	40.524	47.512	56.687

Calculus of Efficiency (DEA)

Table 2: Efficiency by region and year

Departamento	2015	2016	2018	Promedio	Ranking
Tumbes	0.952	0.949	0.954	0.952	1
Región Lima	0.954	0.934	0.962	0.950	2
Ica	0.946	0.913	0.961	0.940	3
Puno	0.906	0.931	0.970	0.936	4
Arequipa	0.917	0.930	0.955	0.934	5
Madre de Dios	0.919	0.938	0.942	0.933	6
Moquegua	0.921	0.907	0.953	0.927	7
Provincia de Lima	0.935	0.897	0.950	0.927	8
Piura	0.919	0.901	0.939	0.920	9
Tacna	0.922	0.900	0.938	0.920	10
Callao	0.917	0.898	0.940	0.918	11
Cusco	0.938	0.901	0.916	0.918	12
Áncash	0.938	0.890	0.919	0.916	13
Huancavelica	0.914	0.898	0.921	0.911	14
Loreto	0.886	0.890	0.907	0.894	15
Apurímac	0.882	0.859	0.933	0.891	16
Ayacucho	0.888	0.868	0.912	0.889	17
Pasco	0.864	0.843	0.925	0.877	18
Ucayali	0.822	0.898	0.892	0.871	19
Lambayeque	0.845	0.874	0.887	0.869	20
Junín	0.860	0.840	0.905	0.868	21
La Libertad	0.873	0.810	0.879	0.854	22
San Martín	0.871	0.802	0.862	0.845	23
Cajamarca	0.854	0.798	0.856	0.836	24
Amazonas	0.815	0.728	0.889	0.811	25
Huánuco	0.806	0.792	0.836	0.811	26

Expenditure vs Efficiency
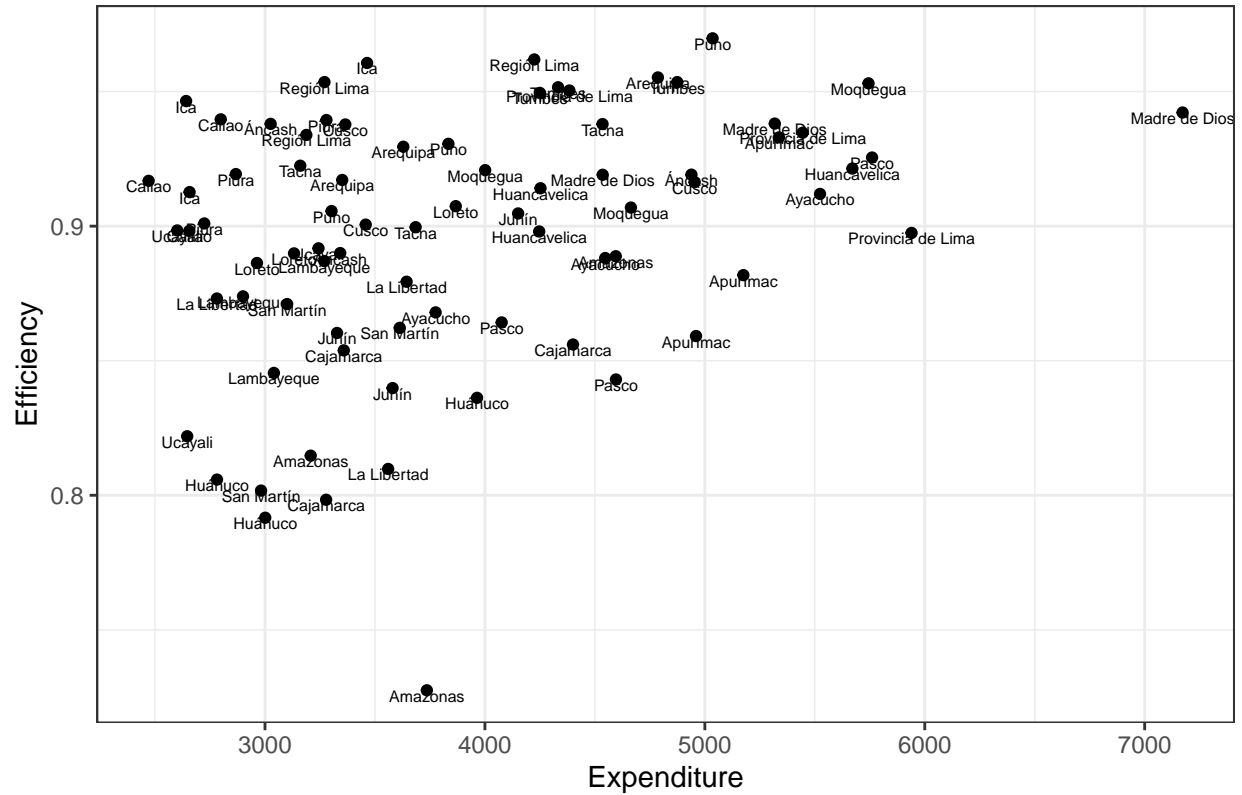
Panel Data - Second Step

Table 3: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
deaRobust	78	0.897	0.048	0.728	0.872	0.934	0.970
analfabetismo	78	7.112	3.714	1.763	4.408	9.454	16.254
pbiRegiones	78	14.907	8.807	6.488	8.600	18.998	50.487
tasaDesnutricion	78	14.831	7.967	1.300	7.134	20.294	33.979
alumDoc	78	10.947	2.061	5.947	9.552	12.243	16.347
nivelUrb	78	70.780	19.750	27.602	57.325	85.719	100.000

Analysis of Correlations

Table 4: Pearson Correlation

	deaRobust	analfabetismo	pbiRegiones	tasaDesnutricion	alumDoc	nivelUrb
deaRobust	1.00	-0.40	0.41	-0.48	-0.05	0.47
analfabetismo	-0.40	1.00	-0.42	0.75	-0.08	-0.91
pbiRegiones	0.41	-0.42	1.00	-0.58	-0.37	0.47

	deaRobust	analfabetismo	pbiRegiones	tasaDesnutricion	alumDoc	nivelUrb
tasaDesnutricion	-0.48	0.75	-0.58	1.00	-0.05	-0.84
alumDoc	-0.05	-0.08	-0.37	-0.05	1.00	0.15
nivelUrb	0.47	-0.91	0.47	-0.84	0.15	1.00
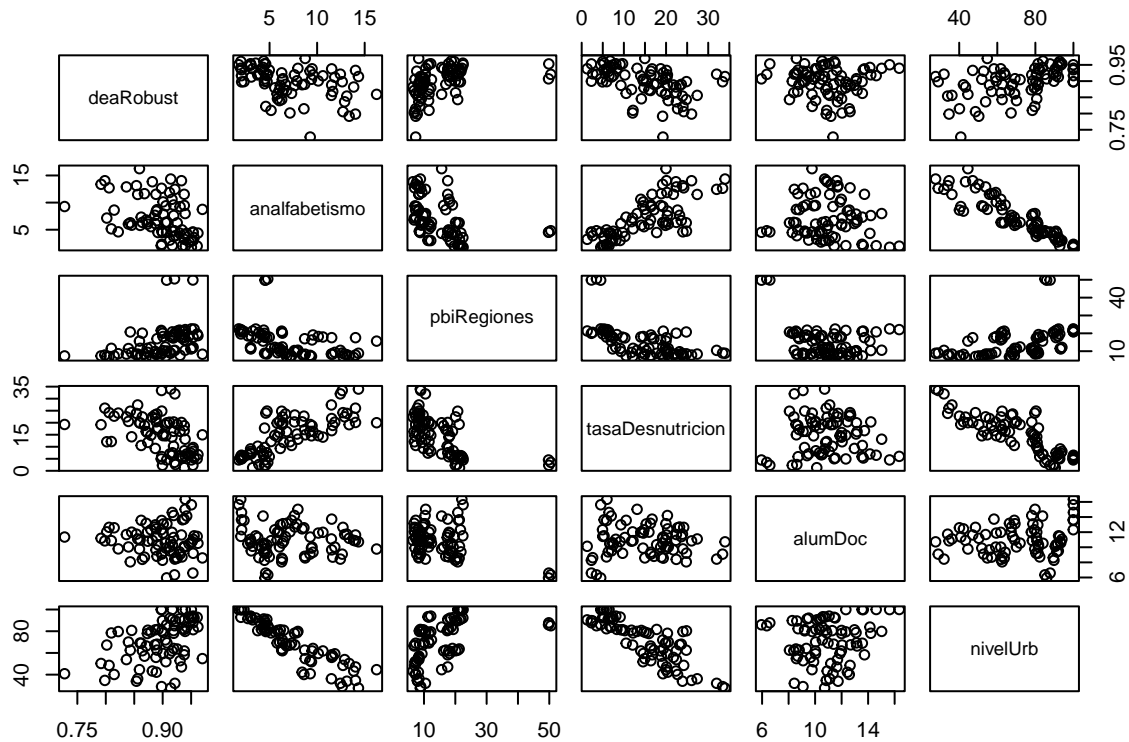
Variance Inflation Factor

With nivelUrb

```
##     data$analfabetismo         data$nivelUrb        data$alumDoc
##           6.124719              9.605255            1.407076
## data$tasaDesnutricion     data$pbiRegiones
##           4.125170              2.020256
```

Withouth nivelUrb

```
##     data$analfabetismo     data$alumDoc data$tasaDesnutricion
##           2.285715          1.331700            2.986365
##     data$pbiRegiones
##           2.008506
```

Fixed Effects

Table 5: Panel Data Fixed Effects

	Dependent variable:
	deaRobust
alumDoc	0.015***
	(0.004)
analfabetismo	−0.012*
	(0.006)
pbiRegiones	0.006**
	(0.003)
tasaDesnutricion	0.00004
	(0.002)
Observations	78
R^2	0.363
Adjusted R^2	−0.022
F Statistic	6.841*** (df = 4; 48)
Note:	*p<0.1; **p<0.05; ***p<0.01

Random Effects

Table 6: Panel Data Random Effects

	Dependent variable:
	deaRobust
alumDoc	0.008***
	(0.003)
analfabetismo	−0.003
	(0.003)
pbiRegiones	0.003**
	(0.001)
tasaDesnutricion	−0.0003
	(0.001)
Constant	0.796***
	(0.052)
Observations	78
R^2	0.218
Adjusted R^2	0.175
F Statistic	20.307***
Note:	*p<0.1; **p<0.05; ***p<0.01

Tests

```
##
##  Hausman Test
##
## data:  data$deaRobust ~ data$alumDoc + data$analfabetismo + data$pbiRegiones +  ...
## chisq = 20.31, df = 4, p-value = 0.0004338
## alternative hypothesis: one model is inconsistent


##
##  Breusch-Pagan LM test for cross-sectional dependence in panels
##
## data:  data$deaRobust ~ data$alumDoc + data$analfabetismo + data$pbiRegiones +    data$tasaDesnutric
## chisq = 475.41, df = 325, p-value = 9.875e-08
## alternative hypothesis: cross-sectional dependence


##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  data$deaRobust ~ data$alumDoc + data$analfabetismo + data$pbiRegiones +    data$tasaDesnutric
## z = 0.27844, p-value = 0.7807
## alternative hypothesis: cross-sectional dependence
```

```
## 
##  studentized Breusch-Pagan test
## 
## data:  plm_within_2
## BP = 5.0503, df = 4, p-value = 0.2822


## 
##  studentized Breusch-Pagan test
## 
## data:  plm_random_2
## BP = 5.0503, df = 4, p-value = 0.2822
```