ETM 540 Group Project

Samira Akther
Jordan Hilton
Andey Nunes
Aparna Raghuram
Mohammed Sheikh

March 12, 2019

Efficiency of Public Expenditure on Education

In this section, the paper aims to study the efficiency of the public expenditures on education of OECD (Organization for Economic Development) countries. DEA (Data Envelopement Analysis), which is a vastly used tool for measuring efficiency, has been used to measure the efficiency of the public expenditures on education by the countries. The analysis provides an efficient frontier which is a linear combination of the data point with efficiency scores of 1. The analysis also allows us to know possible reduction in the inputs to generate the output of the countries with scores less than 1.

Optimization Model Definition

In our paper, we used output oriented model. Here, only one input is considered and defined as $x_{i,j}$ to be the amount of the i'th input used by unit j . Here j is the country we are analysing. Two outputs have been considered and defined as $y_{r,j}$ to be the amount of the r'th output produced by unit j. n is the number of countries we analysed.

The following optimization model has been solved to obtain the efficiency score while minimizing the value of

$$\theta$$

.

$$\begin{aligned}
& \text{Min } \theta \\
\text{subject to } & \sum_{j=1}^{n} x_{i,j}\lambda_j \leq x_{i,k} \forall\, i \quad i = 1, \ldots, n \\
& \sum_{j=1}^{N^D} y_{r,j}\lambda_j \geq y_{r,k} \forall\, r \quad j = 1, \ldots, r \\
& \lambda_j \geq 0 \;\forall\, j
\end{aligned}$$

Here, Theta provides the efficiency scores of the countries we analyzed. The country with score of 1 lies in the efficient frontier and shows that any input reduction is not possible for generating the output. The countries with scores less than 1 will be considered as inefficient within the efficient frontier. Vector Lambda he specific amount of a unit j used in setting the target for for performance for unit k.

Data Description

Data has been obtained from secondary sources. Data on PISA scores were obtained from OECD website. Data on public education expenditures as percentage of GDP were obtained from UNESCO website. We considered 36 countries of OECD. The input variable is the public education spending as percentage of GDP of all 36 countries for the year of 2015. The output variables are the average PISA score of all 36 countries and the graduation rates for the year of 2015.

Here is a summary of the data.

```
## load our data, downloaded from the google sheet
data <- read.csv("OECDdata.csv", stringsAsFactors = F)

glimpse(data)
```

```
## Observations: 36
## Variables: 16
## $ Countries                     <chr> "Australia", "Austria", "Belgium...
## $ GDP                           <dbl> 1349.00, 382.10, 455.00, 1560.00...
## $ SciencePISA                   <int> 510, 495, 502, 528, 447, 493, 50...
## $ ReadingPISA                   <int> 503, 485, 499, 527, 459, 487, 50...
## $ MathematicsPISA               <int> 494, 497, 507, 516, 423, 492, 51...
## $ AveragePISA                   <dbl> 502.33, 492.33, 502.67, 523.67, ...
## $ HDI                           <dbl> 0.936, 0.903, 0.913, 0.920, 0.84...
## $ Primary.education             <chr> "9,546", "11,689", "10,211", "9,...
## $ Secondary.education           <chr> "12,303", "15,477", "13,070", "1...
## $ Tertiary.education            <chr> "20,344", "17,555", "17,320", "0...
## $ Total.Spending                <chr> "42,193", "44,721", "40,601", "2...
## $ Tertiary.Graduation.Rate      <dbl> NA, 86.079, 31.350, 93.005, 91.1...
## $ Bachelors.Graduation.rates    <dbl> 59.767, 25.013, 43.903, 37.594, ...
## $ Masters                       <dbl> 20.531, 20.287, 26.758, 11.777, ...
## $ Doctorates                    <dbl> 2.618, 1.862, 0.639, 1.559, 0.26...
## $ Spending.as.percentage.of.GDP <chr> "5.32%", "5.45%", "6.55%", "", "...
```

We can see that there are three observations missing from several of the population education level fields and three zero values for `Total Spending`. These countries (listed below) will be excluded from the analysis.

```
data$Total.Spending <- as.numeric(gsub(",", "", data$Total.Spending))
data$Spending.as.percentage.of.GDP <- as.numeric(gsub("%", "", data$Spending.as.percentage.of.GDP))

# remove 0 spending countries
DEAdata <- data %>%
   filter(Total.Spending > 0)

setdiff(data[,1], DEAdata[,1])
```

```
## [1] "Denmark"     "Lithuania"   "Switzerland"
```

DEA Model

Part A: Single input, single output

For this first part, we will set up and solve a DEA model using the single GDP data point for the input and the aggregated test score average for the single output.

Run a DEA with a single input (total education spending) and a single output (Average PISA score)

```
#this is single input and single output
x <- DEAdata %>% select(Total.Spending) ## input
#x2 <- data %>% select(Spending.as.percentage.of.GDP) ## ratio input used in next section
y <- DEAdata %>% select(AveragePISA) ## output
row.names(x) <- DEAdata$Countries #%>%
```

```r
row.names(y) <- DEAdata$Countries


ressingle <- DeaMultiplierModel(x, y, rts = "vrs", orientation = "output")

efficiencyTable <- tibble(
   country = as.character(dimnames(ressingle$Efficiency)[[1]]),
   efficiency = as.numeric(ressingle$Efficiency)) %>%
   arrange(desc(efficiency))

pander(head(efficiencyTable), caption = "DEA Output Efficiency for Aggregate Scores")
```

Table 1: DEA Output Efficiency for Aggregate Scores

country	efficiency
Canada	1
Greece	1
Japan	1
Mexico	1
Estonia	0.9991
Finland	0.9898

```r
pander(head(ressingle$Lambda), caption = "DEA lambda values for Aggregate Scores")
```

Table 2: DEA lambda values for Aggregate Scores (continued below)

	Australia	Austria	Belgium	Canada	Chile
Australia	0	0	0	0	0
Austria	0	0	0	0	0
Belgium	0	0	0	0	0
Canada	0	0	0	1	0
Chile	0	0	0	0.3131	0
Czech Republic	0	0	0	0.8606	0

Table 3: Table continues below

	Czech Republic	Estonia	Finland	France	Germany
Australia	0	0	0	0	0
Austria	0	0	0	0	0
Belgium	0	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0	0	0	0	0

Table 4: Table continues below

	Greece	Hungary	Iceland	Ireland	Israel	Italy
Australia	0	0	0	0	0	0

	Greece	Hungary	Iceland	Ireland	Israel	Italy
Austria	0	0	0	0	0	0
Belgium	0	0	0	0	0	0
Canada	0	0	0	0	0	0
Chile	0.6869	0	0	0	0	0
Czech Republic	0	0	0	0	0	0

Table 5: Table continues below

	Japan	South Korea	Latvia	Luxembourg	Mexico
Australia	1	0	0	0	0
Austria	1	0	0	0	0
Belgium	1	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0.1394	0	0	0	0

Table 6: Table continues below

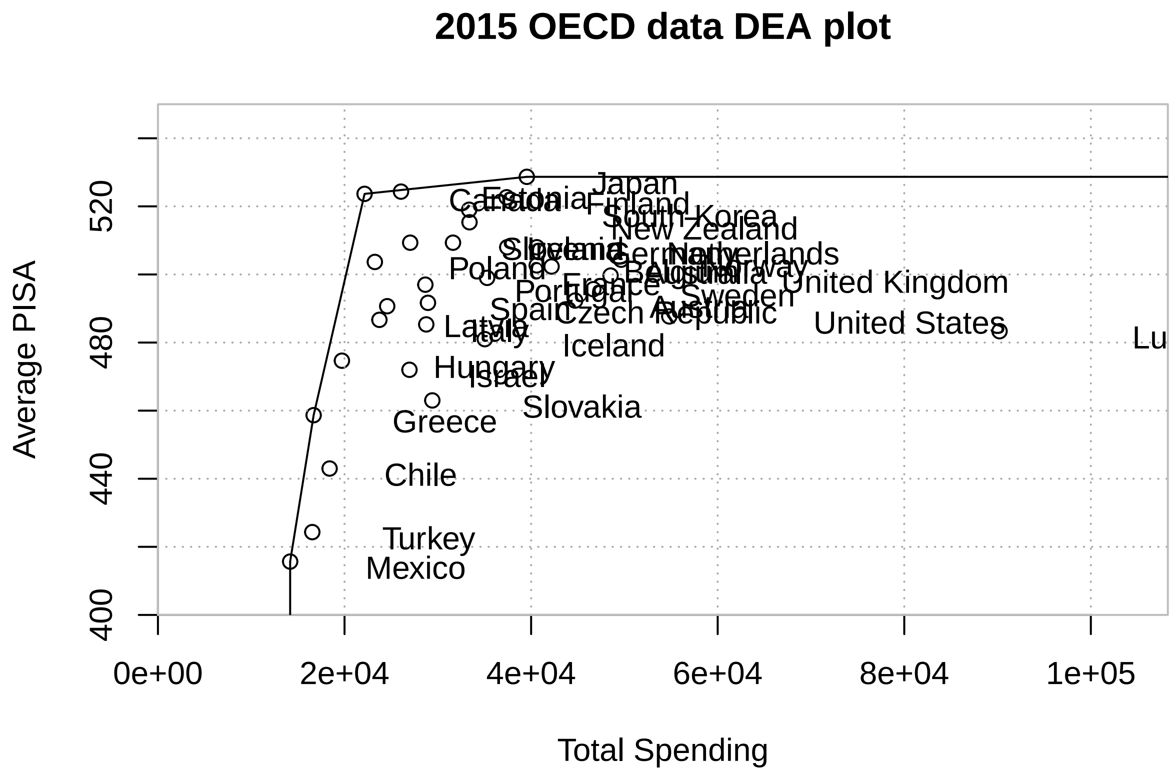
	Netherlands	New Zealand	Norway	Poland	Portugal
Australia	0	0	0	0	0
Austria	0	0	0	0	0
Belgium	0	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0	0	0	0	0

Table 7: Table continues below

	Slovakia	Slovenia	Spain	Sweden	Turkey
Australia	0	0	0	0	0
Austria	0	0	0	0	0
Belgium	0	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0	0	0	0	0

	United Kingdom	United States
Australia	0	0
Austria	0	0
Belgium	0	0
Canada	0	0
Chile	0	0
Czech Republic	0	0

```r
dea.plot(x = x, y = y, txt = dimnames(x)[[1]], GRID = T,
         xlab = "Total Spending", ylab = "Average PISA",
         ylim = c(400,550), main = "2015 OECD data DEA plot")
```



Part B:

Run a DEA with a single input, but multiple outputs:

```r
#this is single input and multiple output
y <- DEAdata %>% select(SciencePISA, ReadingPISA, MathematicsPISA) ## output
row.names(y) <- DEAdata[,1]

resmult <- DeaMultiplierModel(x, y, rts = "vrs", orientation = "output")

multEfficiencyTable <- tibble(
   country = as.character(dimnames(resmult$Efficiency)[[1]]),
   efficiency = as.numeric(resmult$Efficiency)) %>%
   arrange(desc(efficiency))
```

```r
pander(head(multEfficiencyTable, 10), caption = "DEA Output Efficency for Multiple Scores")
```

Table 9: DEA Output Efficency for Multiple Scores

country	efficiency
Mexico	1
Finland	1

country	efficiency
Greece	1
Japan	1
Canada	1
Estonia	1
South Korea	0.9955
New Zealand	0.9952
Ireland	0.9886
Slovenia	0.9791

```r
pander(head(resmult$Lambda), caption = "DEA lambda values for Multiple Scores")
```

Table 10: DEA lambda values for Multiple Scores (continued below)

	Australia	Austria	Belgium	Canada	Chile
Australia	0	0	0	0.1261	0
Austria	0	0	0	0.1185	0
Belgium	0	0	0	0.2843	0
Canada	0	0	0	1	0
Chile	0	0	0	0.3131	0
Czech Republic	0	0	0	0.3792	0

Table 11: Table continues below

	Czech Republic	Estonia	Finland	France	Germany
Australia	0	0	0.7091	0	0
Austria	0	0	0	0	0
Belgium	0	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0	0.6208	0	0	0

Table 12: Table continues below

	Greece	Hungary	Iceland	Ireland	Israel	Italy
Australia	0	0	0	0	0	0
Austria	0	0	0	0	0	0
Belgium	0	0	0	0	0	0
Canada	0	0	0	0	0	0
Chile	0.6869	0	0	0	0	0
Czech Republic	0	0	0	0	0	0

Table 13: Table continues below

	Japan	South Korea	Latvia	Luxembourg	Mexico
Australia	0.1649	0	0	0	0
Austria	0.8815	0	0	0	0

	Japan	South Korea	Latvia	Luxembourg	Mexico
Belgium	0.7157	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0	0	0	0	0

Table 14: Table continues below

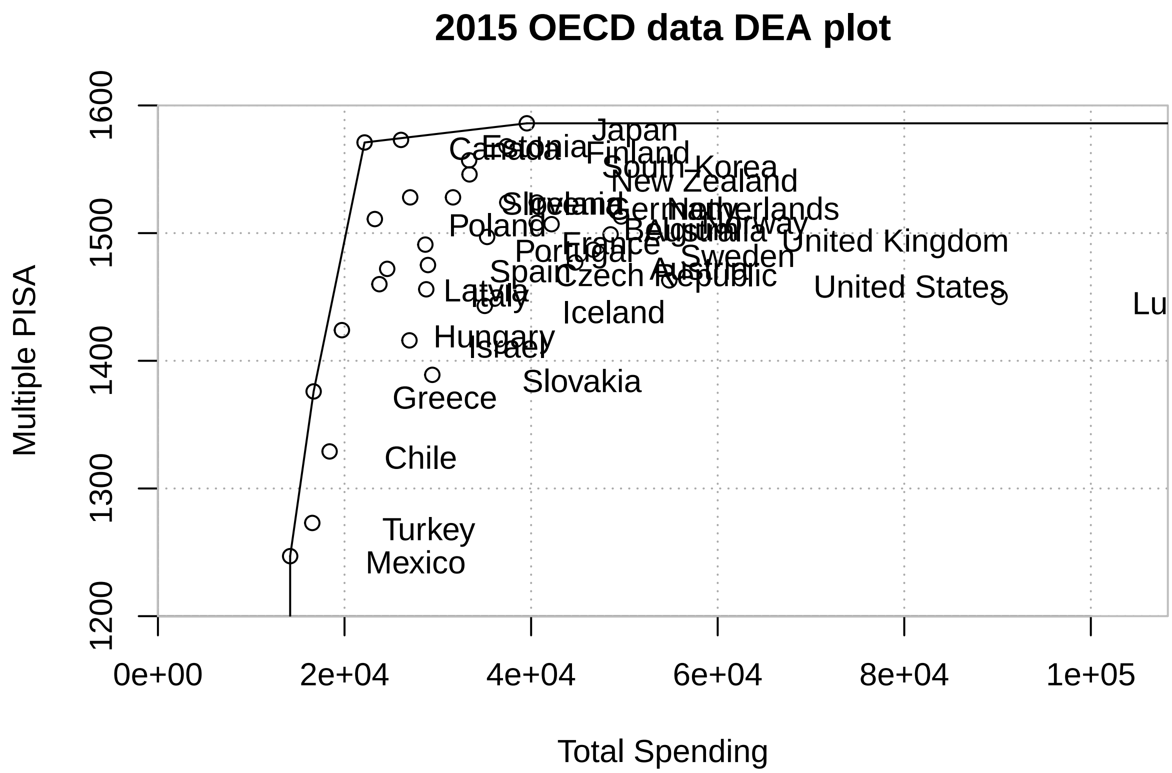
	Netherlands	New Zealand	Norway	Poland	Portugal
Australia	0	0	0	0	0
Austria	0	0	0	0	0
Belgium	0	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0	0	0	0	0

Table 15: Table continues below

	Slovakia	Slovenia	Spain	Sweden	Turkey
Australia	0	0	0	0	0
Austria	0	0	0	0	0
Belgium	0	0	0	0	0
Canada	0	0	0	0	0
Chile	0	0	0	0	0
Czech Republic	0	0	0	0	0

	United Kingdom	United States
Australia	0	0
Austria	0	0
Belgium	0	0
Canada	0	0
Chile	0	0
Czech Republic	0	0

```r
dea.plot(x = x, y = y, txt = dimnames(x)[[1]], GRID = T,
         xlab = "Total Spending", ylab = "Multiple PISA",
         ylim = c(1200,1600), main = "2015 OECD data DEA plot")
```



Part C: Results and discussion

Some quick analysis of the results:

```
max(ressingle$Efficiency) ## the maximum efficiency for one output
```

```
## [1] 1
```

```
which.max(ressingle$Efficiency) ## the row value of the maximum efficiency
```

```
## [1] 4
```

```
data[33, ] ## the row from the data matching the answers above
```

```
##      Countries   GDP SciencePISA ReadingPISA MathematicsPISA AveragePISA
## 33 Switzerland 679.3         506         492             521      506.33
##      HDI Primary.education Secondary.education Tertiary.education
## 33 0.942
##    Total.Spending Tertiary.Graduation.Rate Bachelors.Graduation.rates
## 33              0                       NA                     47.376
##    Masters Doctorates Spending.as.percentage.of.GDP
## 33  18.201      3.345                           5.1
```

```
min(ressingle$Efficiency) ##same for mins
```

```
## [1] 0.8806337
```

```
which.min(ressingle$Efficiency)
```

```
## [1] 27
```

```r
data[22,]
```

```
##     Countries   GDP SciencePISA ReadingPISA MathematicsPISA AveragePISA
## 22 Luxembourg 57.78         483         481             486      483.33
##      HDI Primary.education Secondary.education Tertiary.education
## 22 0.899            20,892              20,413             48,907
##    Total.Spending Tertiary.Graduation.Rate Bachelors.Graduation.rates
## 22          90212                   79.373                      9.725
##    Masters Doctorates Spending.as.percentage.of.GDP
## 22    7.85      1.224                          3.92
```

```r
ressingle$Efficiency["United States",] ## US efficiency
```

```
## [1] 0.9224469
```

```r
max(resmult$Efficiency) ## the maximum efficiency for one output
```

```
## [1] 1
```

```r
which.max(resmult$Efficiency) ## the row value of the maximum efficiency
```

```
## [1] 21
```

```r
data[4, ]
```

```
##   Countries  GDP SciencePISA ReadingPISA MathematicsPISA AveragePISA  HDI
## 4    Canada 1560         528         527             516      523.67 0.92
##   Primary.education Secondary.education Tertiary.education Total.Spending
## 4             9,249              12,900                  0          22149
##   Tertiary.Graduation.Rate Bachelors.Graduation.rates Masters Doctorates
## 4                   93.005                     37.594  11.777      1.559
##   Spending.as.percentage.of.GDP
## 4                            NA
```

```r
min(resmult$Efficiency) ##same for mins
```

```
## [1] 0.9082457
```

```r
which.min(resmult$Efficiency)
```

```
## [1] 27
```

```r
data[14,]
```

```
##    Countries   GDP SciencePISA ReadingPISA MathematicsPISA AveragePISA
## 14   Iceland 16.94         473         482             488         481
##      HDI Primary.education Secondary.education Tertiary.education
## 14 0.927            11,215              11,149             12,671
##    Total.Spending Tertiary.Graduation.Rate Bachelors.Graduation.rates
## 14          35035                   88.968                     50.991
##    Masters Doctorates Spending.as.percentage.of.GDP
## 14  29.324      1.611                          7.71
```

Linear Program

For our linear program, we're going to pretend for a moment that all these countries have equal student populations, and that if we look at the current ratio of spending per student and test score achievement, that will tell us how efficient it is to spend additional money in each country.

Here's our question for analysis: let's say we have a grant of size $\$5,000$ student, and that we can't spend more than $\$1,000$ per student in any single country. How should we spend our grant to most efficiently improve test scores?

We have 36 countries in the sample, so we have 36 decision variables x_i; how much money to spend in each country. We should not spend less than \$0 in any country.

For this formulation let's define the decison variable as x_i and the spending efficiency as E_i:

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{i=1}^{\#countries} \quad E_i * x_i \\
\text{subject to} \quad & \sum_{i=1}^{\#countries} \quad x_i \leq 5000 \\
& x_i \leq 1000, \\
& x_i \geq 0 \; \forall \; i
\end{aligned}
$$

Here's some code to adapt the data values from our multiple-output DEA for use in our linear program:

```
# eddata <- eddata[-c(7,21,33),] ##drop rows missing data
countrynames <- DEAdata[,1]
countrycount <- length(countrynames) #the total number of countries
spendingefficiency <- resmult$Efficiency
```

Now let's implement our program in OMPR, implicitly since we don't want to write out more than 60 constraints::

```
educationLP <- MIPModel() %>%
  add_variable(country[i], i = 1:countrycount, type = "continuous", lb = 0) %>%
  set_objective(sum_expr(spendingefficiency[i] * country[i] , i = 1:countrycount), "max") %>% #maximize
  add_constraint(sum_expr(country[i], i = 1:countrycount) <= 5000) %>%  # total spending constraint
  add_constraint(country[i] <= 1000,  i = 1:countrycount) %>%  # per country spending restraint
  solve_model(with_ROI(solver = "glpk"))

educationLP
```

```
## Status: optimal
## Objective value: 5000
```

Let's take a look at our solution:

```
edLPsolution <- educationLP$solution
names(edLPsolution) <- countrynames
table<-cbind(edLPsolution, DEAdata$AveragePISA, DEAdata$Total.Spending)
colnames(table)<-c("Grant", "AveragePISA", "Current Spending")
kable(table)
```

	Grant	AveragePISA	Current Spending
Australia	0	502.33	42193
Austria	0	492.33	44721
Belgium	0	502.67	40601
Canada	1000	523.67	22149
Chile	0	443.00	18400
Czech Republic	0	490.67	24574

	Grant	AveragePISA	Current Spending
Estonia	0	524.33	26055
Finland	1000	522.67	37378
France	0	499.00	35287
Germany	0	508.00	37446
Greece	1000	458.67	16691
Hungary	0	474.67	19720
Iceland	0	481.00	35035
Ireland	0	509.33	31628
Israel	0	472.00	26961
Italy	0	485.33	28762
Japan	1000	528.67	39541
South Korea	0	519.00	33358
Latvia	0	486.67	23739
Luxembourg	0	483.33	90212
Mexico	1000	415.67	14173
Netherlands	0	508.00	40614
New Zealand	0	515.33	33398
Norway	0	504.33	49649
Poland	0	503.67	23250
Portugal	0	497.00	28664
Slovakia	0	463.00	29411
Slovenia	0	509.33	27040
Spain	0	491.67	28945
Sweden	0	495.67	46672
Turkey	0	424.33	16546
United Kingdom	0	499.67	48519
United States	0	487.67	54814

Shadow Prices Row Duals

In this study the row duals indicate how much benefit we could obtain from additional grant money given our current constraints, which could be used in a request for further grant money.

```
shadow_prices <- educationLP$solution_row_duals()
names(shadow_prices) <- countrynames
shadow_prices
```

```
##      Australia        Austria        Belgium         Canada          Chile
##   1.000000e+00   0.000000e+00   0.000000e+00   0.000000e+00   0.000000e+00
## Czech Republic        Estonia        Finland         France        Germany
##   0.000000e+00   0.000000e+00   0.000000e+00   2.273737e-16   0.000000e+00
##         Greece        Hungary        Iceland        Ireland         Israel
##   0.000000e+00   2.273737e-16   0.000000e+00   0.000000e+00   0.000000e+00
##          Italy          Japan    South Korea         Latvia     Luxembourg
##   0.000000e+00   0.000000e+00   2.273737e-16   0.000000e+00   0.000000e+00
##         Mexico    Netherlands    New Zealand         Norway         Poland
##   0.000000e+00   4.547474e-16   0.000000e+00   0.000000e+00   0.000000e+00
##       Portugal       Slovakia       Slovenia          Spain         Sweden
##   0.000000e+00   0.000000e+00   0.000000e+00   0.000000e+00   0.000000e+00
##         Turkey United Kingdom  United States           <NA>
##   0.000000e+00   0.000000e+00   0.000000e+00   0.000000e+00
```

References

Data Sources

All data sourced from https://data.oecd.org/

Anderson, T. R. (2019) Operations Research in R

Anderson, T. R. (2019) Data Envelope Analysis in R

Bogetoft, P., Otto, L. (2011) Benchmarking with DEA, SFS, and R. Springer.

Appendix