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)

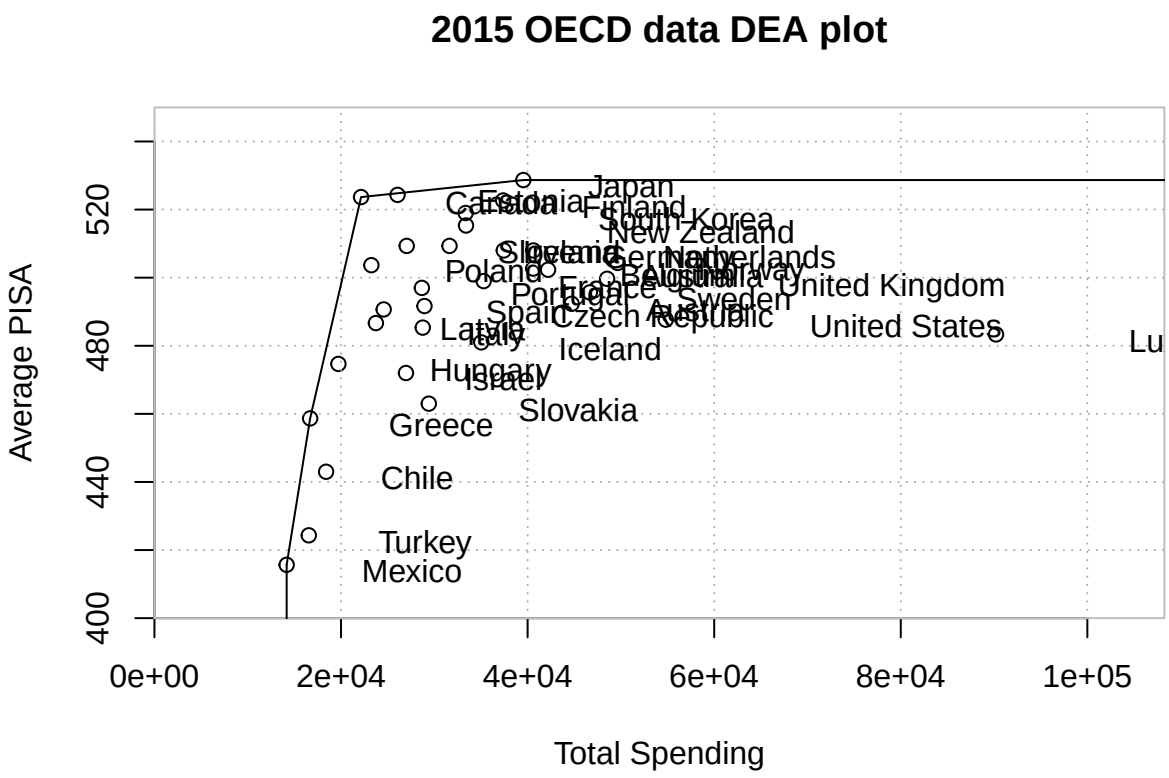
```
#this is single input and single output
x <- DEAdata %>% select(Total.Spending) ## input
#x2 <- data %>% select(Spending.as.percentage.of.GDP) ## ratio input used in next section
y <- DEAdata %>% select(AveragePISA) ## output
row.names(x) <- DEAdata$Countries #%>%
```

```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
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:

```
#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))
```

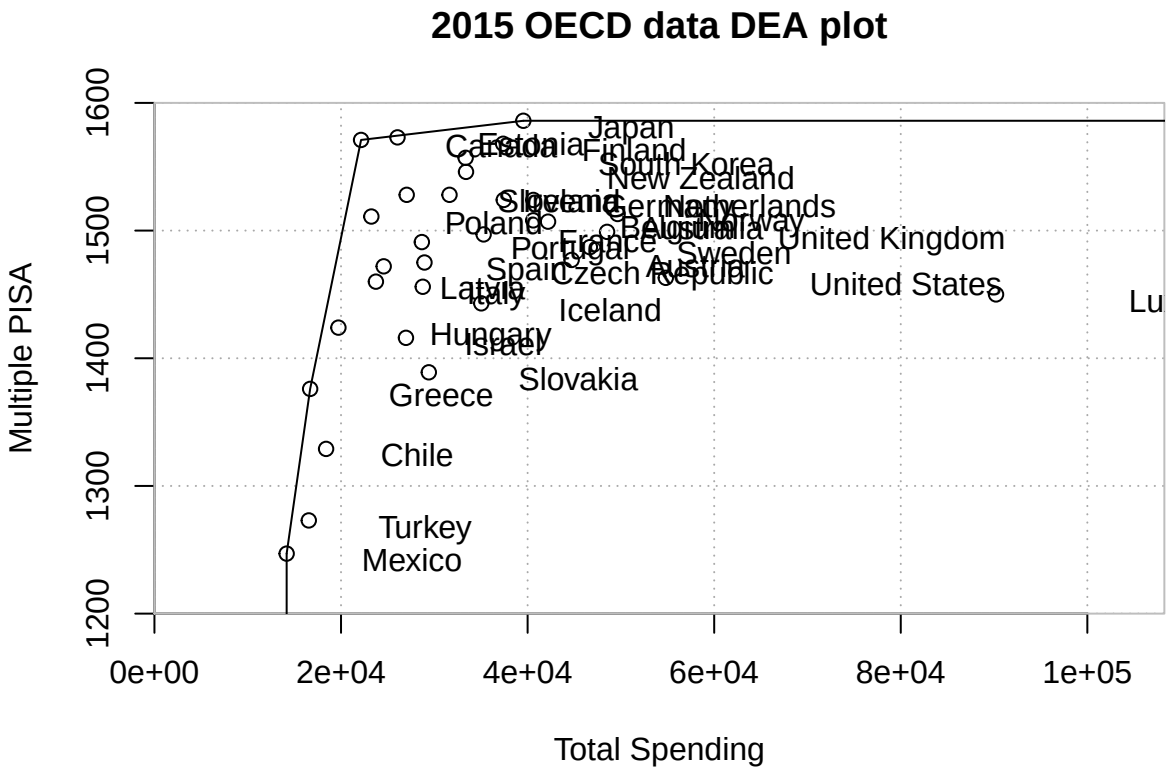
```
pander(head(multEfficiencyTable, 10), caption = "DEA Output Efficency for Multiple Scores")
```

Table 2: DEA Output Efficency for Multiple Scores

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

country	efficiency
South Korea	0.9955
New Zealand	0.9952
Ireland	0.9886
Slovenia	0.9791

```
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
```

```
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
```

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

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

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

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

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

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

```
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
```

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

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

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

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

```
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
```

example change

References

Data Sources

insert links/doi to data sets/info here

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