

HDI Analysis

2023-07-02

We aim at seeing if we can use information of HDI from all countries as a proxy for their wealth across the whole period considered.

```
rm(list=ls())
setwd("~/Documents/GitHub/Nonparametric-Statistics-project/src/Markdowns")
#Installing packages
library(roahd)

#Extracting data
rawdata<-read.delim("../data/raw_datasets/HDR21-22_Composite_indices.csv",
                    sep = ",", quote="", header=T)
countries<-read.table("../data/OECD_countries_income_level", header=T)$Country

#Adding manually the countries with potentially slightly different names
countries<-c(countries,
             "Korea (Republic of)",
             "Czechia",
             "Republic of Korea",
             "Korea",
             "Slovak Republic",
             "Slovakia",
             "Türkiye",
             "United States of America",
             "United Kingdom of Great Britain and Northern Ireland",
             "Costa Rica",
             "Japan")

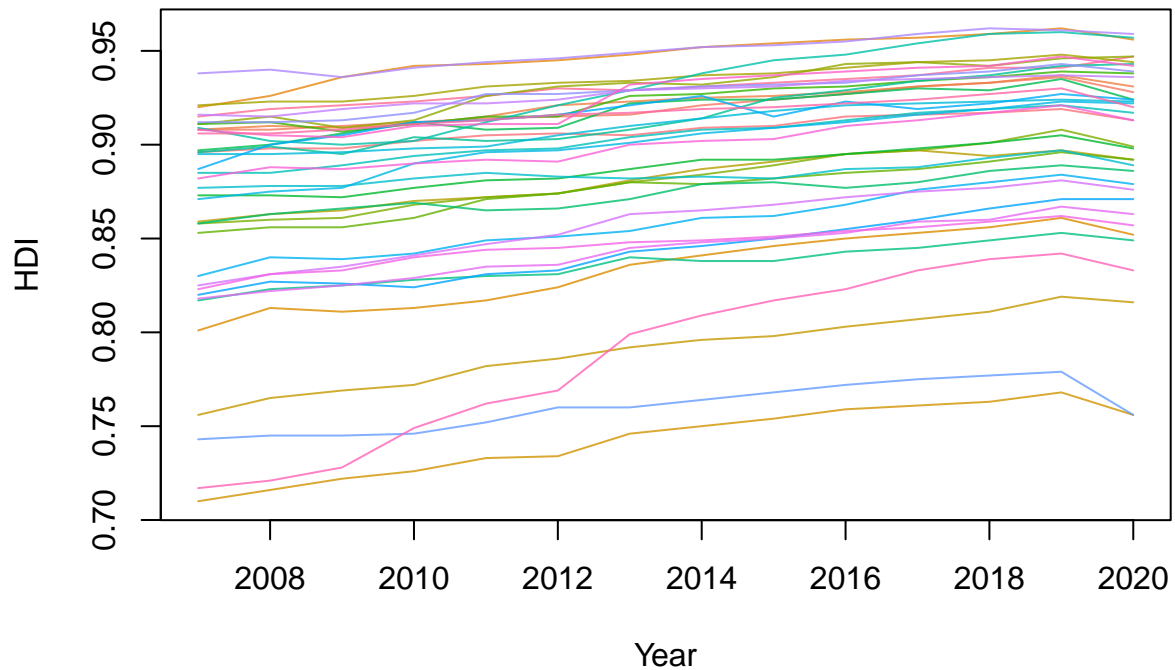
hdi_raw<-rawdata[rawdata$country %in% countries,]
hdi<-hdi_raw[,c(2,23,24,25,26,27,28,29,30,31,32,33,34,35,36)]
```

To do so, we extract raw HDI index data in the period 2007-2020, and select the ones of OECD countries.

We first create the functional datum and plot it

```
#Create functional datum and plot it
grid <- seq(2007, 2020, by=1)
f_data <- fData(grid, hdi[, -c(1)])
par(mfrow=c(1,1))
plot(f_data, xlab="Year",
     ylab="HDI",
     main="HDI index across years")
```

HDI index across years

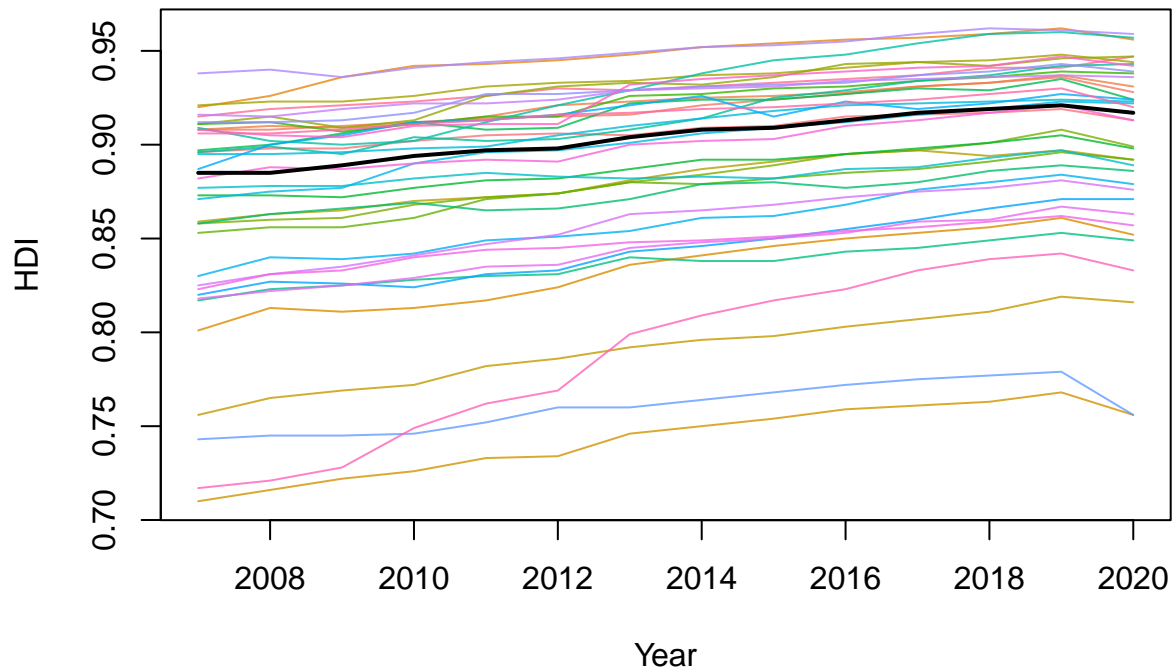


We see that all countries have (luckily) increasing trends. This may turn useful, as allow us to think about classifying and quantifying the “overall height” of the HDI function.

Plotting functional median based on MBD

```
plot(f_data,xlab="Year",
     ylab="HDI",
     main="HDI index across years")
par(cex.axis=1.2, cex.lab=1.2)
median_curve <- median_fData(fData = f_data, type = "MBD",)
lines(grid,median_curve$values,lwd=2.0) # superimpose median
```

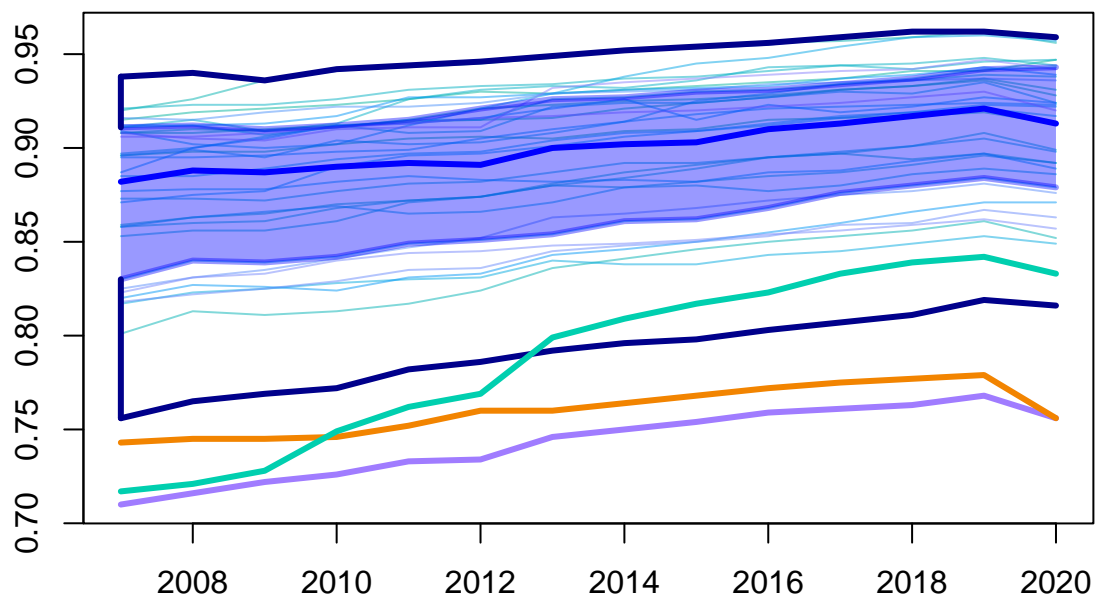
HDI index across years



Plotting magnitude outliers based on MBD

```
# Functional Box Plot with MBD depth
fb_plot_MBD = roahd::fbplot(f_data,
                             main="Magnitude outliers",
                             Depths = "MBD") # Functional Box Plot
```

Magnitude outliers



```
hdi$country[fb_plot_MBD$ID_outliers]
```

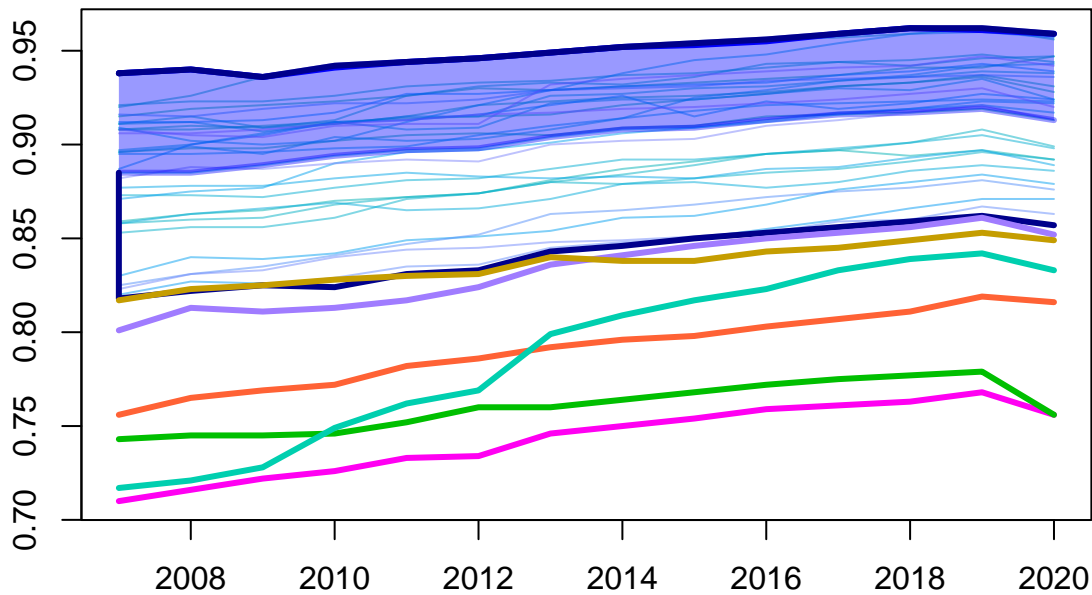
```
## [1] "Colombia" "Mexico" "Turkey"
```

Amplitude outliers according to MBD are Colombia, Mexico and Turkey, with lower amplitude

But we are more interested in the order of curves, to stratify based on increasing values of curves MHI depth

```
fb_plot_MHI = roahd::fbplot(f_data, main="Magnitude outliers", Depths = "MHI")
```

Magnitude outliers



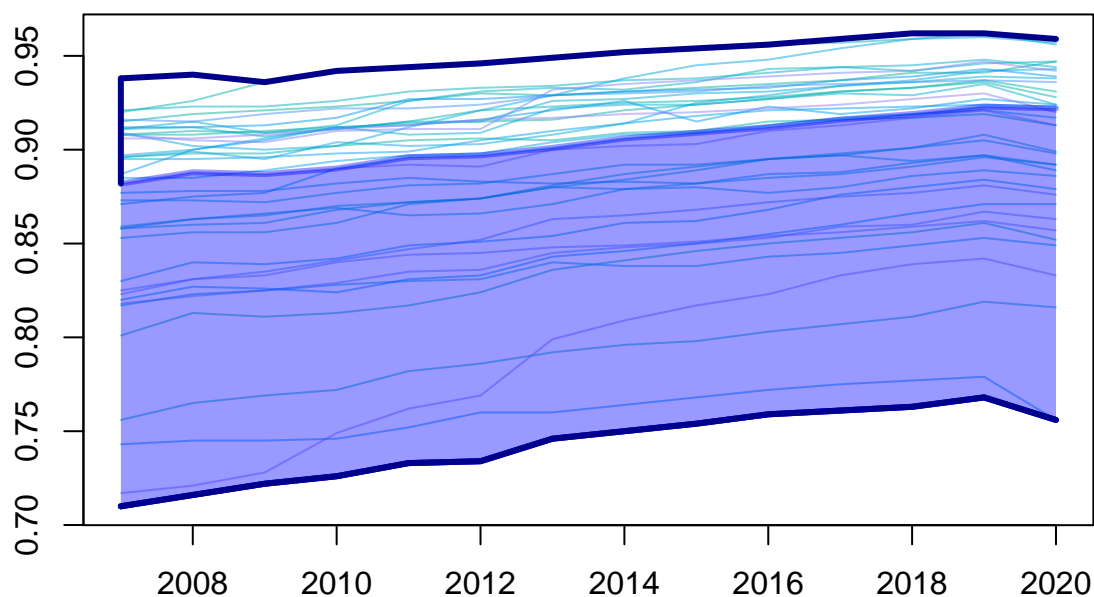
```
hdi$country[fb_plot_MHI$ID_outliers]
```

```
## [1] "Chile" "Colombia" "Costa Rica" "Hungary" "Mexico"
## [6] "Turkey"
```

Amplitude outliers according to MHI are Chile, Colombia, Costa Rica, Hungary, Mexico, Turkey, which are countries with lower income, education level and life expectancy compared with the majority of OECD countries.

```
fb_plot_MEI = roahd::fbplot(f_data, main="Magnitude outliers", Depths = "MEI")
```

Magnitude outliers



```
hdi$country[fb_plot_MEI$ID_outliers]
```

```
## character(0)
```

No outliers according to MEI. We then proceed by using the MHI variable, as it seems more informative.

Now we select the years corresponding to the MPOWER measures (2007,2008,2010,2012, 2014,2016, 2018,2020), and we save - the original HDI value - The MHI value, which can be used as a continuous variable quantifying country health and wealth across the whole period considered
- The HDI clustering induced by the MHI

```
hdi_index_yearselected<-hdi_raw[,c(2,23,24,26,28,30,32,34,35,36)]
```

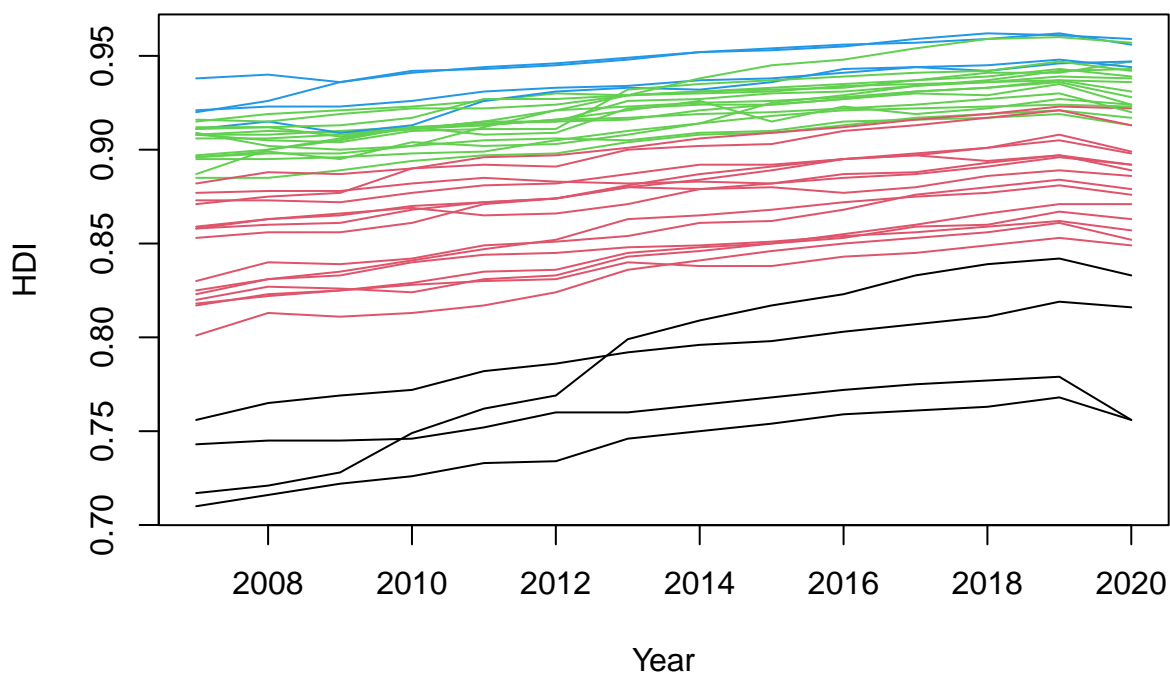
```
MHI_f <- MHI(f_data)
MHI_quantiles<-quantile(MHI_f, probs=c(0.10,0.5,0.90))
MHI_quantiles2<-MHI_quantiles
```

```
hdi_raw$MHI<-0
hdi_raw$MHI[MHI_f<=MHI_quantiles2[1]]<-1
hdi_raw$MHI[MHI_f>MHI_quantiles2[1]]<-2
hdi_raw$MHI[MHI_f>MHI_quantiles2[2]]<-3
hdi_raw$MHI[MHI_f>MHI_quantiles2[3]]<-4
hdi_raw$MHI
```

```
## [1] 3 3 3 3 4 2 1 1 2 4 4 2 2 3 2 3 2 2 3 3 3 2 3 2 2 3 2 1 3 4 3 2 2 2 2 3 1 3
```

And we plot the clustering structure

HDI-based clustering



```
hdi$country[hdi_raw$MHI==1]
```

```
## [1] "Colombia" "Costa Rica" "Mexico" "Turkey"
```

```
hdi$country[hdi_raw$MHI==2]
```

```
## [1] "Chile" "Czechia" "Spain"
## [4] "Estonia" "France" "Greece"
## [7] "Hungary" "Italy" "Korea (Republic of)"
## [10] "Lithuania" "Latvia" "Poland"
## [13] "Portugal" "Slovakia" "Slovenia"
```

```
hdi$country[hdi_raw$MHI==3]
```

```
## [1] "Australia" "Austria" "Belgium" "Canada"
## [5] "Finland" "United Kingdom" "Ireland" "Iceland"
## [9] "Israel" "Japan" "Luxembourg" "Netherlands"
## [13] "New Zealand" "Sweden" "United States"
```

```
hdi$country[hdi_raw$MHI==4]
```

```
## [1] "Switzerland" "Germany" "Denmark" "Norway"
```

The HDI-based clustering seems extremely coherent with the health and wealth of the nations. In particular, the first cluster contains less developed countries from Latin America plus Turkey; the second cluster contains mostly European countries with middle or middle-low economies and healthcare systems, plus some

extraeuropean countries with the similar features. On the other hand, the third and fourth clusters contain richer/healthier countries. In particular, the fourth cluster contain northern European countries and Switzerland, for their extremely advanced healthcare system and their rich economies, alongside Germany.

We will add also this MHI-based clustering structure which may be used for further analysis