This appendix presents the R code used to identify and remove outliers from the database. This R scipt is the one contains in the makefile: $Analysis/DataBase/MakeFile_RemoveOutliers_Lag1.Rmd$. This step had been repeated three times: (i) when dependent variables were lagged one year (see section: [The impact of CEP on CFP]) and (ii) two years behind others variables and (iii) when the GreenScore variables was the only independent variables considered into the econometric model (see section: [Sensitivity Analysis]).

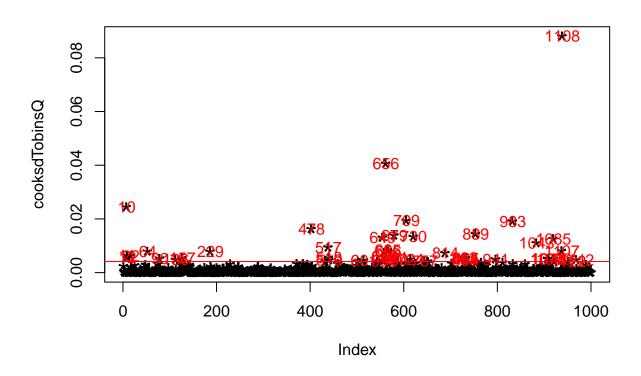
```
# Packages loading
if (!require("dplyr")) install.packages("dplyr")
library(dplyr)
if (!require("grDevices")) install.packages("grDevices")
library(grDevices)
if (!require("data.table")) install.packages("data.table")
library(data.table)
if (!require("formatR")) install.packages("formatR")
library(formatR)
if (!require("highlight")) install.packages("highlight")
library(highlight)
# Database Loading
path <- "Analysis/DataBase/DataSynchronization/Lag1.csv"</pre>
Lag1 <- read.csv(file = path, header = TRUE, stringsAsFactors = FALSE)
# Select only variables that I need for my model
ModelLag1 <- Lag1 %>% select(c(YearIndex, CompaniesIndex,
   Roa, TobinsQ, DebtToEquityRatio, NetMargin, TotalAssets,
    GicsClassification, CarbonProductivity, WaterProductivity,
   WasteProductivity, SustainabilityPayLink, SustainableThemedCommitment,
    AuditScore))
# I transform the 'TotalAssets' column into FirmSize
# using the log of TotalAssets
ModelLag1$TotalAssets <- log10(ModelLag1$TotalAssets)</pre>
# I use the natural log for TobinsQ
ModelLag1$TobinsQ <- log10(ModelLag1$TobinsQ)</pre>
# I rename some columns
ModelLag1 <- ModelLag1 %>% setnames(old = c("DebtToEquityRatio",
    "TotalAssets", "GicsClassification", "NetMargin"), new = c("Leverage".
    "FirmSize", "Industry", "Growth"))
# I define my models in lm as cooks.distance do not
# support plm object
Roa <- lm(Roa ~ SustainabilityPayLink + SustainableThemedCommitment +
   AuditScore + CarbonProductivity + WaterProductivity +
    WasteProductivity + FirmSize + Growth + Leverage + Industry,
    data = ModelLag1)
TobinsQ <- lm(TobinsQ ~ SustainabilityPayLink + SustainableThemedCommitment +
    AuditScore + CarbonProductivity + WaterProductivity +
    WasteProductivity + FirmSize + Growth + Leverage + Industry,
    data = ModelLag1)
# I calculate my cooks distance (i.e. D)
cooksdRoa <- cooks.distance(Roa)</pre>
cooksdTobinsQ <- cooks.distance(TobinsQ)</pre>
# I extract rows considered as influential (i.e.
# observations whose D > 4 * means) and I print them for
# the reader.
influentialRoa <- as.numeric(names(cooksdRoa)[(cooksdRoa >
   4 * mean(cooksdRoa, na.rm = T))])
```

influentialRoa

 $\begin{smallmatrix} 1 \end{smallmatrix} 10 \ 12 \ 25 \ 55 \ 96 \ 244 \ 245 \ 246 \ 381 \ 413 \ 479 \ 480 \ 645 \ 656 \ [15] \ 679 \ 684 \ 718 \ 730 \ 777 \ 794 \ 948 \ 949 \ 1106 \ 1107 \ 1108 \\ 1122 \ 1123 \ 1156 \ [29] \ 1171$

```
influentialTobin <- as.numeric(names(cooksdTobinsQ)[(cooksdTobinsQ >
          4 * mean(cooksdTobinsQ, na.rm = T))])
influentialTobin
```

```
\#\ I remove outliers and create two new dataframes that I
# write in my folders
TobinsQ_Db <- ModelLag1[-c(influentialTobin), ]</pre>
p <- "Analysis/DataBase/DataSynchronization/NoOutliersLag1/TobinsQ.csv"
write.csv(TobinsQ_Db, file = p)
p <- "Analysis/DataBase/DataSynchronization/NoOutliersLag1/Roa.csv"</pre>
Roa_Db <- ModelLag1[-c(influentialRoa), ]</pre>
write.csv(Roa_Db, file = p)
# I report influencial obervations on a graph
## TobinsQ plot cook's distance
plot(cooksdTobinsQ, pch = "*", cex = 2)
### add cutoff line
abline(h = 4 * mean(cooksdTobinsQ, na.rm = T), col = "red")
### add labels
text(x = 1:length(cooksdTobinsQ) + 1, y = cooksdTobinsQ,
    labels = ifelse(cooksdTobinsQ > 4 * mean(cooksdTobinsQ,
        na.rm = T), names(cooksdTobinsQ), ""), col = "red")
```



```
## Roa plot cook's distance
plot(cooksdRoa, pch = "*", cex = 2)
### add cutoff line
abline(h = 4 * mean(cooksdRoa, na.rm = T), col = "red")
### add labels
text(x = 1:length(cooksdRoa) + 1, y = cooksdRoa, labels = ifelse(cooksdRoa >
        4 * mean(cooksdRoa, na.rm = T), names(cooksdRoa), ""),
        col = "red")
```

