This appendix presents the R code I have used to identify and remove outliers from my database.

```
# Packages loading
if (!require("dplyr")) install.packages("dplyr")
library(dplyr)
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
Lag1 <- as.data.frame(read.csv(file = "Analysis/DataBase/DataSynchronization/Lag1.csv",
    header = TRUE, stringsAsFactors = FALSE))
```

```
# Select only variables that I need for my models
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$TotalAssets <- log(ModelLag1$TotalAssets)
# I use the natural log for TobinsQ
ModelLag1$TobinsQ <- log(ModelLag1$TobinsQ)
# I rename some columns
ModelLag1 <- ModelLag1 %>% setnames(old = c("DebtToEquityRatio",
    "TotalAssets", "GicsClassification", "NetMargin"), new = c("FinancialLeverage",
    "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 + FinancialLeverage +
    Industry, data = ModelLag1)
TobinsQ <- lm(TobinsQ ~ SustainabilityPayLink + SustainableThemedCommitment +
    AuditScore + CarbonProductivity + WaterProductivity +
    WasteProductivity + FirmSize + Growth + FinancialLeverage +
    Industry, data = ModelLag1)
# I calculate my cooks distance (i.e. D)
cooksdRoa <- cooks.distance(Roa)
cooksdTobinsQ <- cooks.distance(TobinsQ)
# I extract rows considered as influential (i.e.
# observations whose D > 4 * means)
influentialRoa <- as.numeric(names(cooksdRoa)[(cooksdRoa >
    4 * mean(cooksdRoa, na.rm = T))])
influentialRoa
```

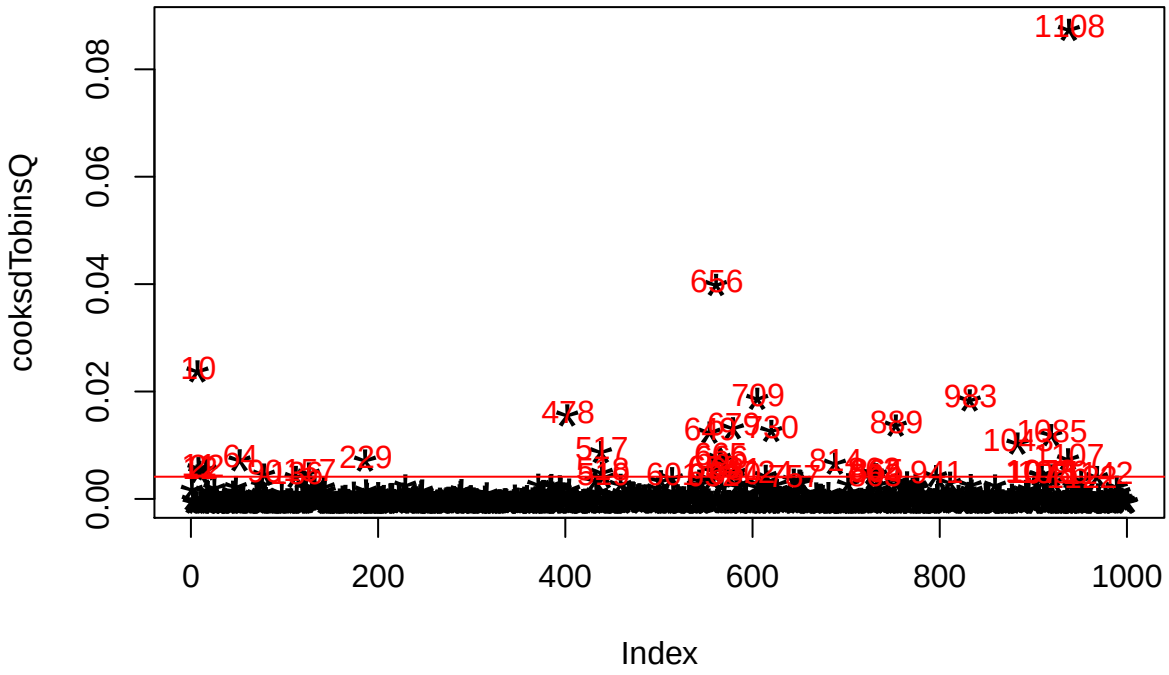
[1] 10 12 25 55 96 244 245 246 381 413 479 480 645 656 [15] 679 684 718 730 777 794 948 949 1106 1107 1108 1122 1123 1156 [29] 1171

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

[1] 10 11 12 22 64 90 136 157 229 478 517 518 519 601 [15] 649 652 653 654 656 665 666 679 680 681 709 724 730 757 [29] 814 862 863 864 865 889 941 983 1043 1073 1074 1075 1085 1086 [43] 1107 1108 1122 1142

```
# I remove outliers and create two new dataframes that I
# write in my folders
TobinsQ_Db <- ModelLag1[-c(influentialTobin), ]
write.csv(TobinsQ_Db, file = "Analysis/DataBase/DataSynchronization/NoOutliersLag1/TobinsQ.csv")
Roa_Db <- ModelLag1[-c(influentialRoa), ]
write.csv(Roa_Db, file = "Analysis/DataBase/DataSynchronization/NoOutliersLag1/Roa.csv")
```

```
# 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")
```