library(car)

library(lmtest)

library(ggplot2)

library(gridExtra)

library(quantmod)

library(dplyr)

library(xts)

library(tidyr)

library(reshape)

library(readxl)

#загрузим данные

data <- read\_excel("C:/Users/valer/OneDrive/Desktop/data.xlsx")

View(data)

as.data.frame(data)

summary(data)

#создадим переменную NIPS

data$NIPS <- data$Net\_income/data$Shares

#отберем необходимые переменные

data1 <- data[,-c(10,11,13)]

View(data1)

data1$Industry <- as.factor(data1$Industry)

data1$`Company Name` <- as.factor(data1$`Company Name`)

data1$`Identifier (RIC)` <- as.factor(data1$`Identifier (RIC)`)

data1$Date <- as.factor(data1$Date)

#выведем описательные статистики

#для начала отчистим данные от выбросов

low\_b\_ESG <- quantile(data1$ESG\_score,0.005)

low\_b\_ESG

up\_b\_ESG <- quantile(data1$ESG\_score,0.995)

up\_b\_ESG

ind\_ESG <- which(data1$ESG\_score < low\_b\_ESG | data1$ESG\_score > up\_b\_ESG)

ind\_ESG

low\_b\_S <- quantile(data1$Social\_score,0.005)

low\_b\_S

up\_b\_S <- quantile(data1$Social\_score,0.995)

up\_b\_S

ind\_S <- which(data1$Social\_score < low\_b\_S | data1$Social\_score > up\_b\_S)

ind\_S

low\_b\_G <- quantile(data1$Gov\_score,0.005)

low\_b\_G

up\_b\_G <- quantile(data1$Gov\_score,0.995)

up\_b\_G

ind\_G <- which(data1$Gov\_score < low\_b\_G | data1$Gov\_score > up\_b\_G)

ind\_G

low\_b\_E <- quantile(data1$Env\_score,0.005)

low\_b\_E

up\_b\_E <- quantile(data1$Env\_score,0.995)

up\_b\_E

ind\_E <- which(data1$Env\_score < low\_b\_E | data1$Env\_score > up\_b\_E)

ind\_E

low\_b\_BVPS <- quantile(data1$BVPS,0.005)

low\_b\_BVPS

up\_b\_BVPS <- quantile(data1$BVPS,0.995)

up\_b\_BVPS

ind\_BVPS <- which(data1$BVPS < low\_b\_BVPS | data1$BVPS > up\_b\_BVPS)

ind\_BVPS

low\_b\_FL <- quantile(data1$FNCL\_LVRG,0.005, na.rm = TRUE)

low\_b\_FL

up\_b\_FL <- quantile(data1$FNCL\_LVRG,0.995, na.rm = TRUE)

up\_b\_FL

ind\_FL <- which(data1$FNCL\_LVRG < low\_b\_FL | data1$FNCL\_LVRG > up\_b\_FL)

ind\_FL

low\_roa <- quantile(data1$RETURN\_ON\_ASSET,0.005, na.rm = TRUE)

low\_roa

up\_roa <- quantile(data1$RETURN\_ON\_ASSET,0.995, na.rm = TRUE)

up\_roa

ind\_roa <- which(data1$RETURN\_ON\_ASSET < low\_roa | data1$RETURN\_ON\_ASSET > up\_roa)

ind\_roa

low\_ag <- quantile(data1$ASSET\_GROWTH,0.005, na.rm = TRUE)

low\_ag

up\_ag <- quantile(data1$ASSET\_GROWTH,0.995, na.rm = TRUE)

up\_ag

ind\_ag <- which(data1$ASSET\_GROWTH < low\_ag | data1$ASSET\_GROWTH > up\_ag)

ind\_ag

low\_qr <- quantile(data1$QUICK\_RATIO,0.005, na.rm = TRUE)

low\_qr

up\_qr <- quantile(data1$QUICK\_RATIO,0.995, na.rm = TRUE)

up\_qr

ind\_qr <- which(data1$QUICK\_RATIO < low\_qr | data1$QUICK\_RATIO > up\_qr)

ind\_qr

low\_mc <- quantile(data1$Market\_cap,0.005, na.rm = TRUE)

low\_mc

up\_mc <- quantile(data1$Market\_cap,0.995, na.rm = TRUE)

up\_mc

ind\_mc <- which(data1$Market\_cap < low\_mc | data1$Market\_cap > up\_mc)

ind\_mc

low\_pe <- quantile(data1$PE\_RATIO,0.005, na.rm = TRUE)

low\_pe

up\_pe <- quantile(data1$PE\_RATIO,0.995, na.rm = TRUE)

up\_pe

ind\_pe <- which(data1$PE\_RATIO < low\_pe | data1$PE\_RATIO > up\_pe)

ind\_pe

low\_nips <- quantile(data1$NIPS,0.005, na.rm = TRUE)

low\_nips

up\_nips <- quantile(data1$NIPS,0.995, na.rm = TRUE)

up\_nips

ind\_nips <- which(data1$NIPS < low\_nips | data1$NIPS > up\_nips)

ind\_nips

outlier <- c(ind\_BVPS,ind\_mc, ind\_pe,ind\_E,ind\_ESG,ind\_FL,ind\_G,ind\_nips,ind\_S, ind\_roa, ind\_qr, ind\_ag)

outlier2 <- unique(outlier)

data\_out <- data1[-c(outlier2),]

View(data\_out)

summary(data\_out)

data\_out <- as.data.frame(data\_out)

unique(data\_out$`Identifier (RIC)`)

library(stargazer)

stargazer(data\_out, type = "text", column.labels=names(data\_out[,-c(1,2,3,11)]),

df=FALSE, digits=3, out = "summary1.doc")

getwd()

#построим корреляционную матрицу регрессоров

data\_cl <- na.omit(data\_out)

View(data\_cl)

mat1 <- as.dist(round(cor(data\_cl[,-c(1,2,3,11)]),3))

library(GGally)

ggcorr(data\_cl[,-c(1,2,3,10)], nbreaks = 6,

low = "steelblue",

mid = "white",

high = "darkred",

label = TRUE,

label\_size = 3,

legend.size = 9,

legend.position = "right",

nudge\_x=-1.2,

layout.exp = 2)

#построим модель

View(data\_out)

mod <- lm(data = data\_out,

formula = data\_out$Market\_cap~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$ESG\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod)

plot(mod, which = 1)

boxCox(mod)

summary(powerTransform(mod))

mod2 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$ESG\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod2)

plot(mod2, which = 1)

crPlots(mod2)

#тест Рамсея для пропуска степени, Н0 - степени не пропущены

library(lmtest)

resettest(mod2, power=2)

resettest(mod2, power=3)

#тест рамсея показывает, что возможно нужна третья степень

mod3 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

log(data\_out$FNCL\_LVRG)+data\_out$ESG\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod3)

crPlots(mod3)

resettest(mod3, power=2)

resettest(mod3, power=3)

mod4 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

log(data\_out$FNCL\_LVRG)+data\_out$ESG\_score+I(data\_out$RETURN\_ON\_ASSET^3)+I(data\_out$QUICK\_RATIO^3)+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod4)

crPlots(mod4)

#добавление степени не улучшает модель, оставляем mod3

plot(mod3, which = 1)

#посмотрим нормальность остатков

plot(mod3,2)

#мультиколлинеарность

vif(mod3)

plot(mod3,3)

#тест Бреуша-Пагана

#нулевая гипотеза - гетероскедастичности нет

bptest(mod3)

#нулевая гипотеза подтверждается

#удалим влиятельные наблюдения

plot(mod3, 4)

cook\_mod <- cooks.distance(mod3)

K <- which(cook\_mod > 4/133)

K

e <- resid(mod3)

g <- abs(e/sd(e))

barplot(g)

N <- which(g>3)

N

OUT <- c(K,N)

OUT <- unique(OUT)

data\_out2 <- data\_out[-OUT,]

View(data\_out2)

#модель на новой выборке

mod5 <- lm(data = data\_out2,

formula = log(data\_out2$Market\_cap)~data\_out2$BVPS+data\_out2$NIPS +

log(data\_out2$FNCL\_LVRG)+data\_out2$ESG\_score+data\_out2$RETURN\_ON\_ASSET+data\_out2$QUICK\_RATIO+data\_out2$ASSET\_GROWTH+data\_out2$PE\_RATIO+ data\_out2$Date)

summary(mod3)

plot(mod5,1)

crPlots(mod5)

#модель 5 наилучшая

#аналогично построим модели с коэффициентами E, S, G

mod\_e <- lm(data = data\_out,

formula = data\_out$Market\_cap~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$Env\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_e)

plot(mod\_e, which = 1)

boxCox(mod\_e)

summary(powerTransform(mod\_e))

mod\_e\_2 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$Env\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_e\_2)

plot(mod\_e\_2, which = 1)

crPlots(mod\_e\_2)

#тест Рамсея для пропуска степени, Н0 - степени не пропущены

library(lmtest)

resettest(mod\_e\_2, power=2)

resettest(mod\_e\_2, power=3)

#тест рамсея показывает, что возможно нужна третья степень

mod\_e\_3 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

log(data\_out$FNCL\_LVRG)+data\_out$Env\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_e\_3)

crPlots(mod\_e\_3)

#добавление степени не улучшает модель, оставляем mod3

plot(mod\_e\_3, which = 1)

#посмотрим нормальность остатков

plot(mod\_e\_3,2)

#мультиколлинеарность

vif(mod\_e\_3)

plot(mod\_e\_3,3)

#тест Бреуша-Пагана

#нулевая гипотеза - гетероскедастичности нет

bptest(mod\_e\_3)

#нулевая гипотеза подтверждается

#удалим влиятельные наблюдения

plot(mod\_e\_3, 4)

cook\_mod <- cooks.distance(mod\_e\_3)

K <- which(cook\_mod > 4/134)

K

e <- resid(mod\_e\_3)

g <- abs(e/sd(e))

barplot(g)

N <- which(g>3)

N

OUT <- c(K,N)

OUT <- unique(OUT)

data\_out2 <- data\_out[-OUT,]

View(data\_out2)

#модель на новой выборке

mod\_e\_5 <- lm(data = data\_out2,

formula = log(data\_out2$Market\_cap)~data\_out2$BVPS+data\_out2$NIPS +

log(data\_out2$FNCL\_LVRG)+data\_out2$Env\_score+data\_out2$RETURN\_ON\_ASSET+data\_out2$QUICK\_RATIO+data\_out2$ASSET\_GROWTH+data\_out2$PE\_RATIO+ data\_out2$Date)

summary(mod\_e\_5)

plot(mod\_e\_5,1)

crPlots(mod\_e\_5)

#построим модель для фактора S

mod\_s <- lm(data = data\_out,

formula = data\_out$Market\_cap~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$Social\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_s)

plot(mod\_s, which = 1)

boxCox(mod\_s)

mod\_s\_2 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$Social\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_s\_2)

plot(mod\_s\_2, which = 1)

crPlots(mod\_s\_2)

#тест Рамсея для пропуска степени, Н0 - степени не пропущены

library(lmtest)

resettest(mod\_s\_2, power=2)

resettest(mod\_s\_2, power=3)

#тест рамсея показывает, что возможно нужна третья степень

mod\_s\_3 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

log(data\_out$FNCL\_LVRG)+data\_out$Social\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_s\_3)

crPlots(mod\_s\_3)

#добавление степени не улучшает модель, оставляем mod3

plot(mod\_s\_3, which = 1)

#посмотрим нормальность остатков

plot(mod\_s\_3,2)

#мультиколлинеарность

vif(mod\_s\_3)

plot(mod\_s\_3,3)

#тест Бреуша-Пагана

#нулевая гипотеза - гетероскедастичности нет

bptest(mod\_s\_3)

#нулевая гипотеза подтверждается

#удалим влиятельные наблюдения

plot(mod\_s\_3, 4)

cook\_mod <- cooks.distance(mod\_s\_3)

K <- which(cook\_mod > 4/134)

K

e <- resid(mod\_e\_3)

g <- abs(e/sd(e))

barplot(g)

N <- which(g>3)

N

OUT <- c(K,N)

OUT <- unique(OUT)

data\_out2 <- data\_out[-OUT,]

View(data\_out2)

#модель на новой выборке

mod\_s\_5 <- lm(data = data\_out2,

formula = log(data\_out2$Market\_cap)~data\_out2$BVPS+data\_out2$NIPS +

log(data\_out2$FNCL\_LVRG)+data\_out2$Social\_score+data\_out2$RETURN\_ON\_ASSET+data\_out2$QUICK\_RATIO+data\_out2$ASSET\_GROWTH+data\_out2$PE\_RATIO+ data\_out2$Date)

summary(mod\_s\_5)

plot(mod\_s\_5,1)

crPlots(mod\_s\_5)

#построим модель для фактора G

mod\_g <- lm(data = data\_out,

formula = data\_out$Market\_cap~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$Gov\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_g)

plot(mod\_g, which = 1)

boxCox(mod\_g)

mod\_g\_2 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

data\_out$FNCL\_LVRG+data\_out$Gov\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_g\_2)

plot(mod\_g\_2, which = 1)

crPlots(mod\_g\_2)

#тест Рамсея для пропуска степени, Н0 - степени не пропущены

library(lmtest)

resettest(mod\_g\_2, power=2)

resettest(mod\_g\_2, power=3)

#тест рамсея показывает, что возможно нужна третья степень

mod\_g\_3 <- lm(data = data\_out,

formula = log(data\_out$Market\_cap)~data\_out$BVPS+data\_out$NIPS +

log(data\_out$FNCL\_LVRG)+data\_out$Gov\_score+data\_out$RETURN\_ON\_ASSET+data\_out$QUICK\_RATIO+data\_out$ASSET\_GROWTH+data\_out$PE\_RATIO+ data\_out$Date)

summary(mod\_g\_3)

crPlots(mod\_g\_3)

#добавление степени не улучшает модель, оставляем mod3

plot(mod\_g\_3, which = 1)

#посмотрим нормальность остатков

plot(mod\_g\_3,2)

#мультиколлинеарность

vif(mod\_g\_3)

plot(mod\_g\_3,3)

#тест Бреуша-Пагана

#нулевая гипотеза - гетероскедастичности нет

bptest(mod\_g\_3)

#нулевая гипотеза не подтверждается, добавим стандартные ошибки в форме Уайта

#удалим влиятельные наблюдения

plot(mod\_g\_3, 4)

cook\_mod <- cooks.distance(mod\_g\_3)

K <- which(cook\_mod > 4/134)

K

e <- resid(mod\_g\_3)

g <- abs(e/sd(e))

barplot(g)

N <- which(g>3)

N

OUT <- c(K,N)

OUT <- unique(OUT)

data\_out2 <- data\_out[-OUT,]

View(data\_out2)

#модель на новой выборке

mod\_g\_5 <- lm(data = data\_out2,

formula = log(data\_out2$Market\_cap)~data\_out2$BVPS+data\_out2$NIPS +

log(data\_out2$FNCL\_LVRG)+data\_out2$Gov\_score+data\_out2$RETURN\_ON\_ASSET+data\_out2$QUICK\_RATIO+data\_out2$ASSET\_GROWTH+data\_out2$PE\_RATIO+ data\_out2$Date)

summary(mod\_g\_5)

plot(mod\_g\_5,1)

crPlots(mod\_g\_5)

#стандартные ошибки в форме Уайта

library(sandwich)

cov\_wtite <- vcovHC(mod\_g\_5, type="HC0")

coeftest(mod\_g\_5,cov\_wtite)

stargazer(mod5,mod\_e\_5,mod\_s\_5,coeftest(mod\_g\_5,cov\_wtite),type = "text",column.labels=c("Модель 1", "Модель 2","Модель 3", "Модель 4"),

df=FALSE, digits=3, out = "models1.doc" )

stargazer(mod\_g\_5,type = "text",column.labels=c("Модель 1", "Модель 2","Модель 3", "Модель 4"),

df=FALSE, digits=3, out = "models2.doc" )