**Лабораторна робота №1**

**Аналіз впливу у лінійній регресії**

|  |  |
| --- | --- |
| [ABT](https://www.nyse.com/quote/XNYS:ABT) | Медицинское оборудование |
| [ABBV](https://www.nyse.com/quote/XNYS:ABBV) | Фармацевтические препараты |
| [AFL](https://www.nyse.com/quote/XNYS:AFL) | Страхование жизни и здоровья |
| [AIG](https://www.nyse.com/quote/XNYS:AIG) | Страхование имущества и несчастных случаев |
| [AON](https://www.nyse.com/quote/XNYS:AON) | Страховые брокеры |
| [BAX](https://www.nyse.com/quote/XNYS:BAX) | Медицинское оборудование |
| [BSX](https://www.nyse.com/quote/XNYS:BSX) | Медицинское оборудование |
| [CB](https://www.nyse.com/quote/XNYS:CB) | Страхование имущества и несчастных случаев |
| [CINF](http://www.nasdaq.com/symbol/cinf) | Страхование имущества и несчастных случаев |
| [EW](https://www.nyse.com/quote/XNYS:EW) | Медицинское оборудование |
| [HIG](https://www.nyse.com/quote/XNYS:HIG) | Страхование имущества и несчастных случаев |
| [ISRG](http://www.nasdaq.com/symbol/isrg) | Медицинское оборудование |
| [MMC](https://www.nyse.com/quote/XNYS:MMC) | Страховые брокеры |
| [MDT](https://www.nyse.com/quote/XNYS:MDT) | Медицинское оборудование |
| [PBCT](http://www.nasdaq.com/symbol/pbct) | Страхование и ипотечное кредитование |
| [PFG](https://www.nyse.com/quote/XNYS:PFG) | Страхование жизни и здоровья |
| [PRU](https://www.nyse.com/quote/XNYS:PRU) | Страхование жизни и здоровья |
| [TMK](https://www.nyse.com/quote/XNYS:TMK) | Страхование жизни и здоровья |

**Log returns: lg(наступний/попередній)**

1. Показати таблицю даних логарифмічні норми прибутку з лагом 1;
2. Обрання компанії AFL
3. **Побудова лінійної регресії з використання всіх наявних даних крім останніх 10 сесій:**

model1<-lm(afl1~.-afl1,data=Data[1:(nn-10),21-40])

summary(model1)

> model1<-lm(afl1~.-afl1,data=Data[1:(nn-10),21:40])

> summary(model1)

Call:

lm(formula = afl1 ~ . - afl1, data = Data[1:(nn - 10), 21:40])

Residuals:

Min 1Q Median 3Q Max

-0.039188 -0.005835 0.000701 0.007020 0.020925

Coefficients:

Estimate Std. Error t value

(Intercept) -0.0007105 0.0009690 -0.733

abbv1 0.0723111 0.0591408 1.223

abt1 -0.0671446 0.0907262 -0.740

aig1 -0.0031777 0.0853086 -0.037

aon1 0.1883285 0.1166101 1.615

bax1 -0.0181378 0.1204554 -0.151

bsx1 -0.0670633 0.0562482 -1.192

cb1 0.1609738 0.1746612 0.922

cinf1 -0.2428909 0.1859045 -1.307

dhr1 0.1096713 0.1282296 0.855

ew1 -0.0194219 0.0378240 -0.513

hig1 -0.1571621 0.1000430 -1.571

isrg1 0.0050859 0.0356973 0.142

mdt1 0.0842791 0.1315460 0.641

mmc1 -0.0107435 0.1797901 -0.060

pbct1 -0.1005166 0.1336433 -0.752

pfg1 0.2435157 0.1435640 1.696

pru1 0.2808231 0.1006709 2.790

tmk1 0.1492877 0.1951752 0.765

wlp1 0.1667038 0.0890027 1.873

Pr(>|t|)

(Intercept) 0.46480

abbv1 0.22382

abt1 0.46069

aig1 0.97035

aon1 0.10891

bax1 0.88056

bsx1 0.23549

cb1 0.35855

cinf1 0.19385

dhr1 0.39409

ew1 0.60855

hig1 0.11881

isrg1 0.88694

mdt1 0.52294

mmc1 0.95245

pbct1 0.45344

pfg1 0.09242 .

pru1 0.00613 \*\*

tmk1 0.44583

wlp1 0.06348 .

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01064 on 121 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.4076, Adjusted R-squared: 0.3146

F-statistic: 4.382 on 19 and 121 DF, p-value: 2.333e-07

Лінійна регресія має коефіцієнт детермінації (0.4076), рівень значущості

гіпотези про залежність майже нульовий.

Проведемо аналіз залишків:

> resr1<-lm(afl1~.-afl1,data=Data[1:(nn-10),21:40])

> resr1

Call:

lm(formula = afl1 ~ . - afl1, data = Data[1:(nn - 10), 21:40])

Coefficients:

(Intercept) abbv1 abt1

-0.0007105 0.0723111 -0.0671446

aig1 aon1 bax1

-0.0031777 0.1883285 -0.0181378

bsx1 cb1 cinf1

-0.0670633 0.1609738 -0.2428909

dhr1 ew1 hig1

0.1096713 -0.0194219 -0.1571621

isrg1 mdt1 mmc1

0.0050859 0.0842791 -0.0107435

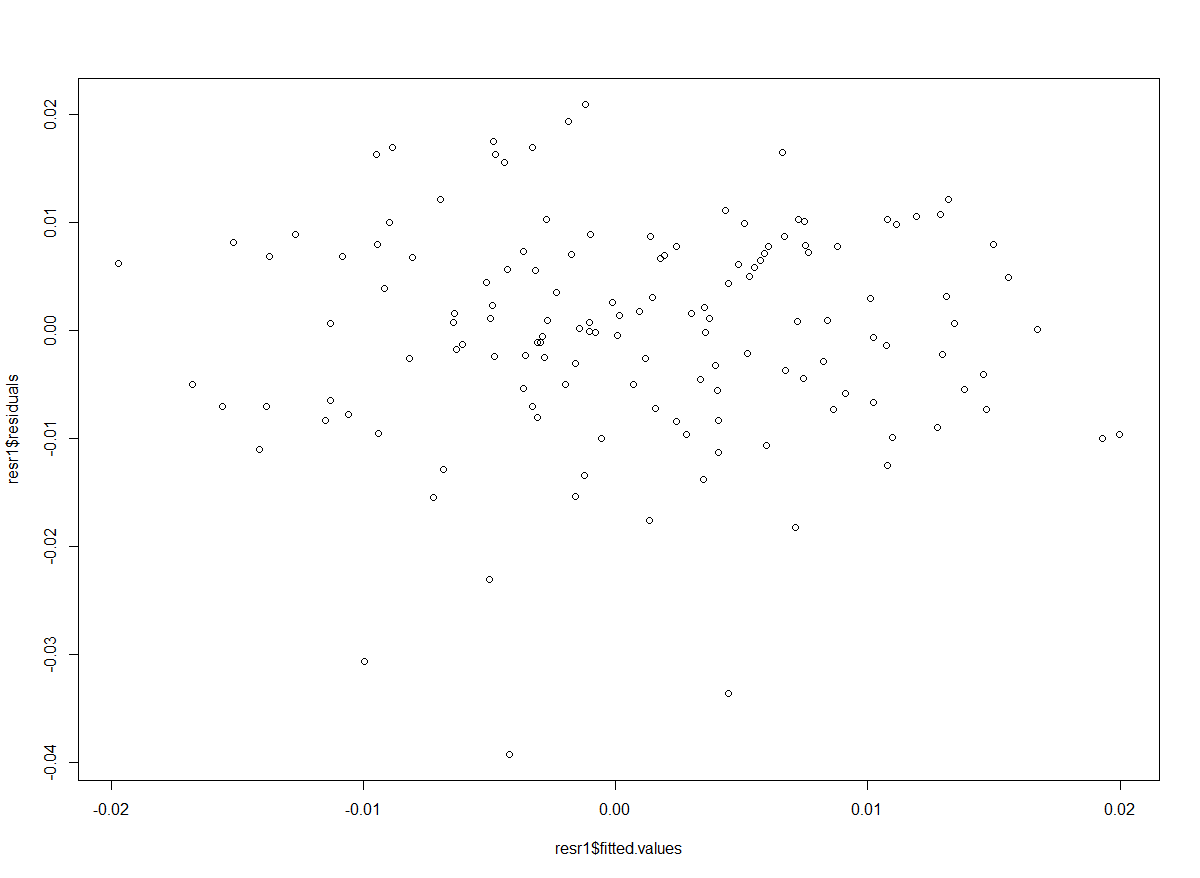
pbct1 pfg1 pru1

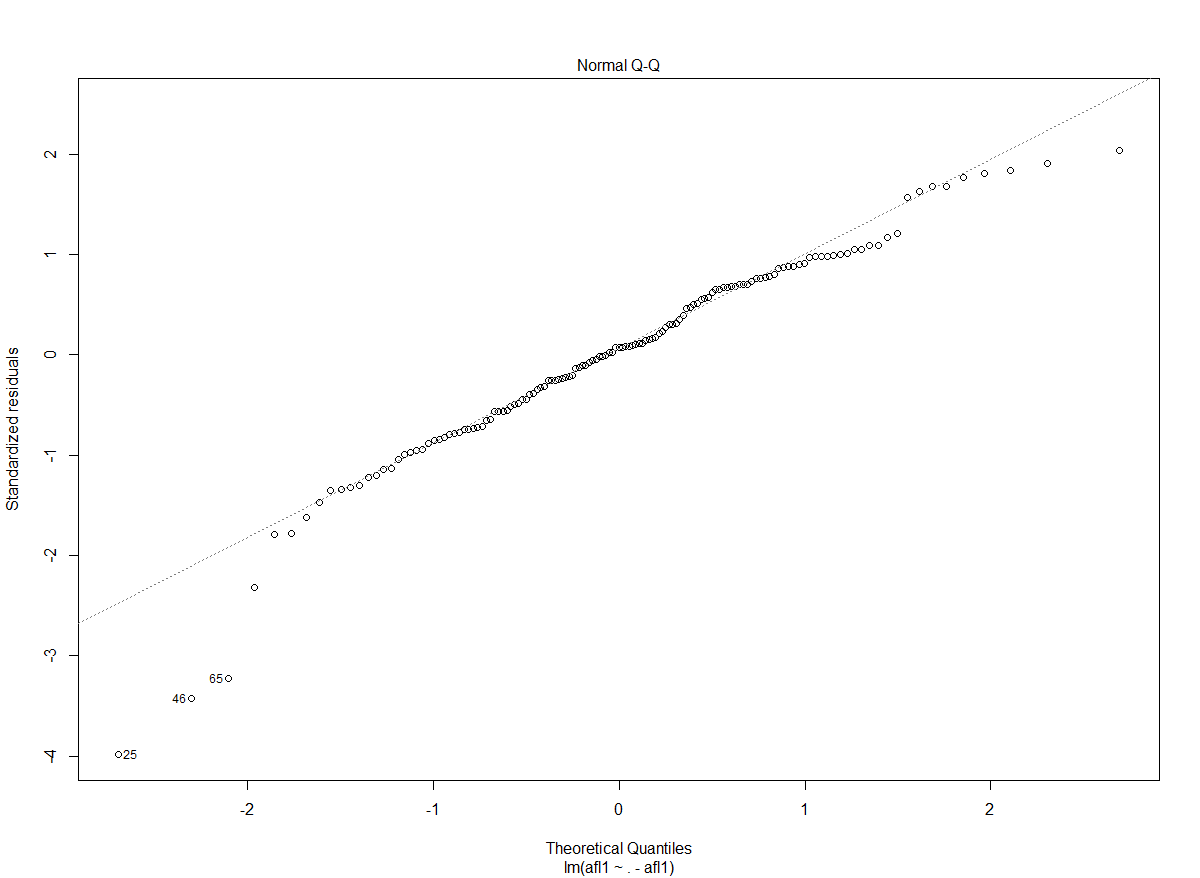
-0.1005166 0.2435157 0.2808231

tmk1 wlp1

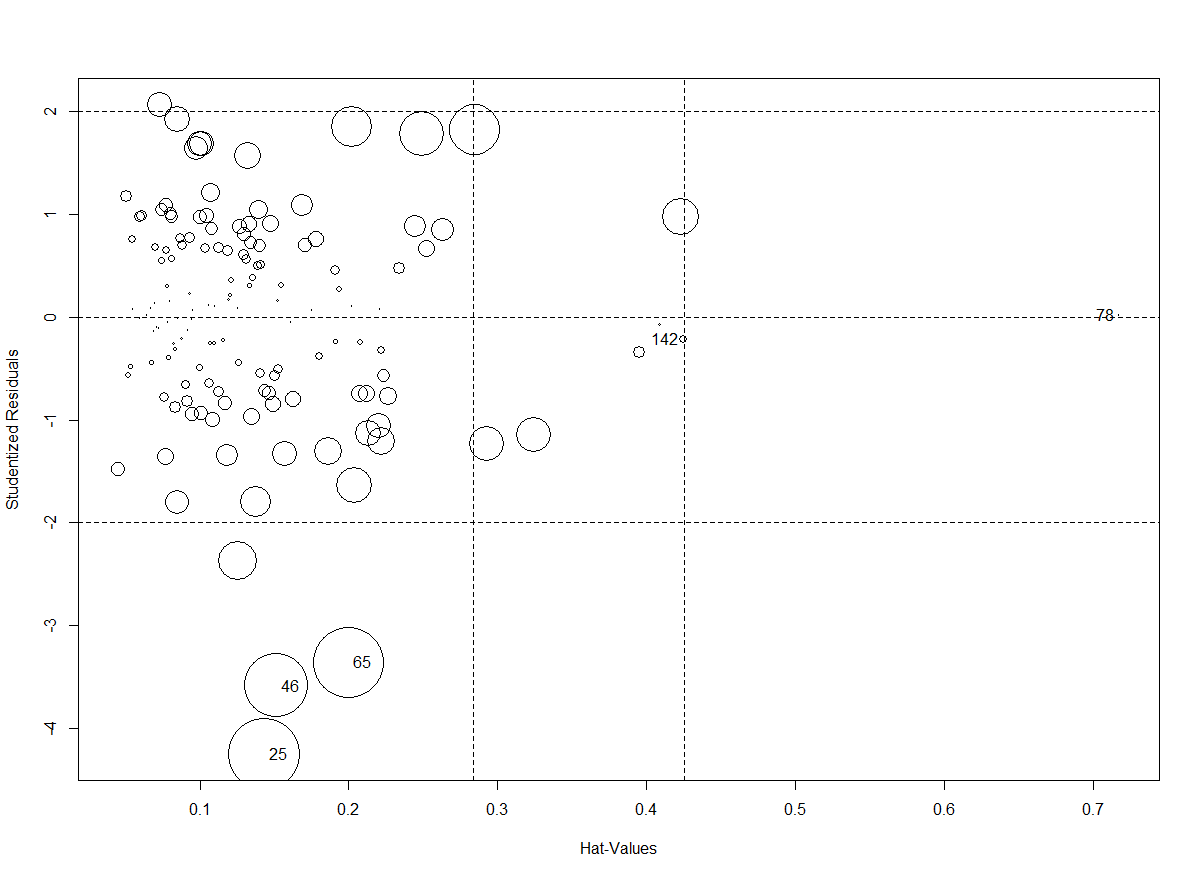
0.1492877 0.1667038

plot(resr1$fitted.values,resr1$residuals)





Схоже, що данні мають викиди. Дослідемо їх



StudRes Hat CookD

25 -4.24856270 0.1432091 0.132220224

46 -3.58509062 0.1512627 0.104314121

65 -3.35298038 0.1996297 0.129263882

78 0.02187125 0.7173928 0.000061220

142 -0.21766591 0.4243545 0.001760181

Зробимо такий же аналіз, але вилучимо викиди:

> model11<-lm(afl1~.-afl1,data=data1[-c(25,42,65,78,142),])

> summary(model11)

Call:

lm(formula = afl1 ~ . - afl1, data = data1[-c(25, 42, 65, 78,

142), ])

Residuals:

Min 1Q Median 3Q Max

-0.031270 -0.005829 0.000021 0.006650 0.021426

Coefficients:

Estimate Std. Error t value

(Intercept) 1.313e-05 8.891e-04 0.015

abbv1 4.881e-02 5.463e-02 0.893

abt1 -2.476e-02 8.345e-02 -0.297

aig1 1.080e-01 8.221e-02 1.313

aon1 8.926e-02 1.080e-01 0.826

bax1 -1.119e-03 1.101e-01 -0.010

bsx1 -7.139e-02 6.322e-02 -1.129

cb1 9.996e-02 1.597e-01 0.626

cinf1 -2.106e-01 1.687e-01 -1.248

dhr1 8.909e-02 1.183e-01 0.753

ew1 -2.721e-02 5.494e-02 -0.495

hig1 -2.533e-01 9.299e-02 -2.724

isrg1 -7.737e-04 3.307e-02 -0.023

mdt1 -4.345e-02 1.220e-01 -0.356

mmc1 2.449e-01 1.690e-01 1.449

pbct1 -9.033e-02 1.224e-01 -0.738

pfg1 2.492e-01 1.317e-01 1.893

pru1 1.914e-01 9.372e-02 2.043

tmk1 2.669e-01 1.837e-01 1.453

wlp1 1.568e-01 8.569e-02 1.829

Pr(>|t|)

(Intercept) 0.98824

abbv1 0.37346

abt1 0.76718

aig1 0.19168

aon1 0.41034

bax1 0.99191

bsx1 0.26109

cb1 0.53247

cinf1 0.21440

dhr1 0.45276

ew1 0.62128

hig1 0.00745 \*\*

isrg1 0.98138

mdt1 0.72224

mmc1 0.14999

pbct1 0.46185

pfg1 0.06091 .

pru1 0.04334 \*

tmk1 0.14896

wlp1 0.06992 .

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.009599 on 116 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.4397, Adjusted R-squared: 0.3479

F-statistic: 4.791 on 19 and 116 DF, p-value: 4.586e-08

Як ми бачимо коефіцієнт детермінації став трохи вище,але не набагато, а

рівень значущості гіпотези про залежність ще більш прямую до 0.

1. **Побудова лінійної регресії з використання даних по 50 сесіях, що передують 10-ти останнім.**

model2<-lm(afl1~.-afl1,data=Data[(nn-60):(nn-10),21-40])

summary(model2)

Call:

lm(formula = afl1 ~ . - afl1, data = Data[(nn - 60):(nn - 10),

21:40])

Residuals:

Min 1Q Median 3Q

-0.0160192 -0.0025693 -0.0006083 0.0031179

Max

0.0201407

Coefficients:

Estimate Std. Error t value

(Intercept) 0.0009178 0.0014055 0.653

abbv1 -0.0993462 0.0982574 -1.011

abt1 0.0694902 0.1157361 0.600

aig1 0.3387709 0.1602300 2.114

aon1 -0.2990625 0.2174060 -1.376

bax1 0.0295591 0.2028458 0.146

bsx1 -0.0573560 0.0668038 -0.859

cb1 0.2356683 0.2528376 0.932

cinf1 -0.4579704 0.2491982 -1.838

dhr1 0.1823505 0.1709662 1.067

ew1 -0.1174990 0.0862957 -1.362

hig1 -0.2126511 0.1833901 -1.160

isrg1 0.0221482 0.0488468 0.453

mdt1 -0.1777310 0.1756037 -1.012

mmc1 0.4927278 0.2971665 1.658

pbct1 -0.0206812 0.1624055 -0.127

pfg1 0.3404977 0.2233394 1.525

pru1 -0.0707477 0.1821972 -0.388

tmk1 0.5612484 0.3954091 1.419

wlp1 0.1529316 0.1649669 0.927

Pr(>|t|)

(Intercept) 0.5186

abbv1 0.3198

abt1 0.5526

aig1 0.0426 \*

aon1 0.1788

bax1 0.8851

bsx1 0.3972

cb1 0.3585

cinf1 0.0757 .

dhr1 0.2944

ew1 0.1831

hig1 0.2551

isrg1 0.6534

mdt1 0.3193

mmc1 0.1074

pbct1 0.8995

pfg1 0.1375

pru1 0.7004

tmk1 0.1658

wlp1 0.3611

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.008244 on 31 degrees of freedom

Multiple R-squared: 0.6292, Adjusted R-squared: 0.4019

F-statistic: 2.769 on 19 and 31 DF, p-value: 0.005731

Лінійна регресія має коефіцієнт вижче коефіцієнт детермінації (0.6292), ніж у першії моделі, але нажаль для аналізу не підходить;

рівень значущості гіпотези про залежність 0,005731, що менше 0,05.

> resr2<-lm(afl1~.-afl1,data=Data[(nn-60):(nn-10),21:40])

> resr2

Call:

lm(formula = afl1 ~ . - afl1, data = Data[(nn - 60):(nn - 10),

21:40])

Coefficients:

(Intercept) abbv1 abt1

0.0009178 -0.0993462 0.0694902

aig1 aon1 bax1

0.3387709 -0.2990625 0.0295591

bsx1 cb1 cinf1

-0.0573560 0.2356683 -0.4579704

dhr1 ew1 hig1

0.1823505 -0.1174990 -0.2126511

isrg1 mdt1 mmc1

0.0221482 -0.1777310 0.4927278

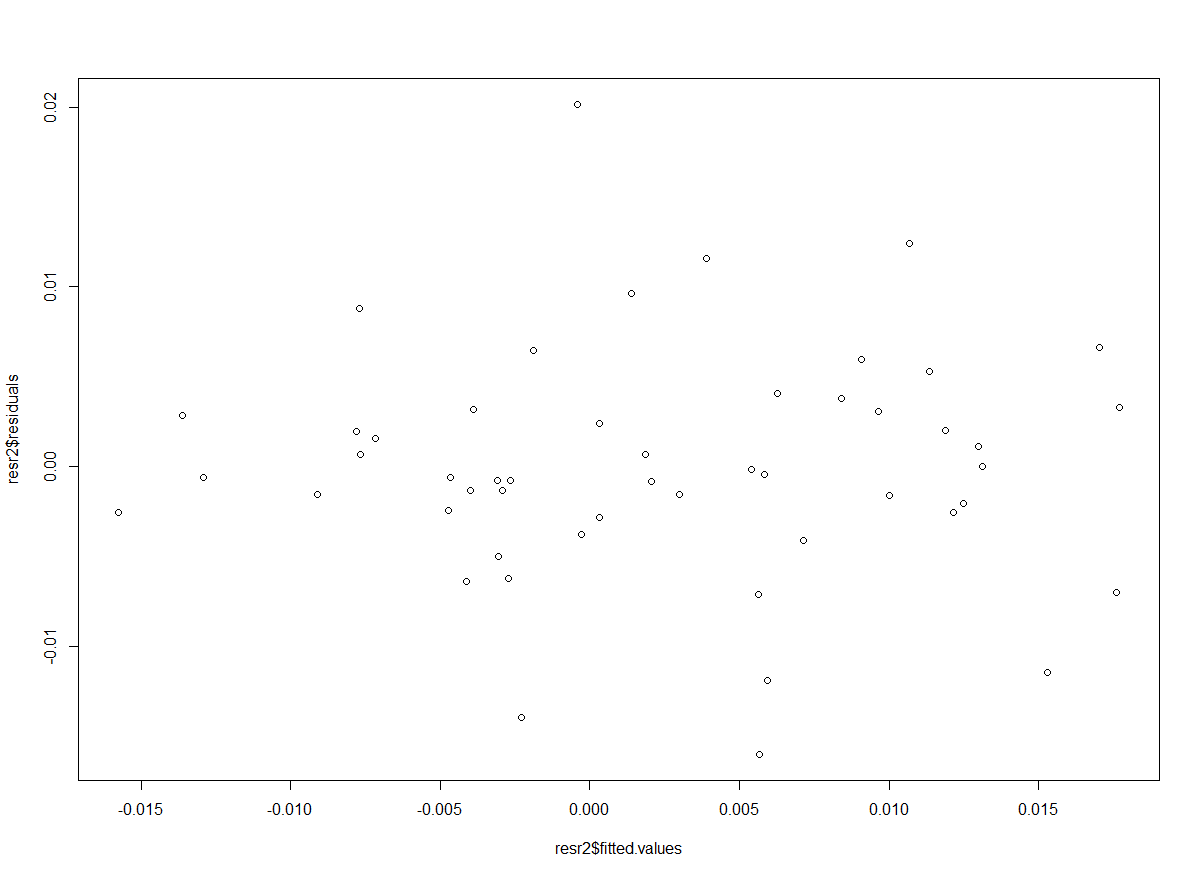
pbct1 pfg1 pru1

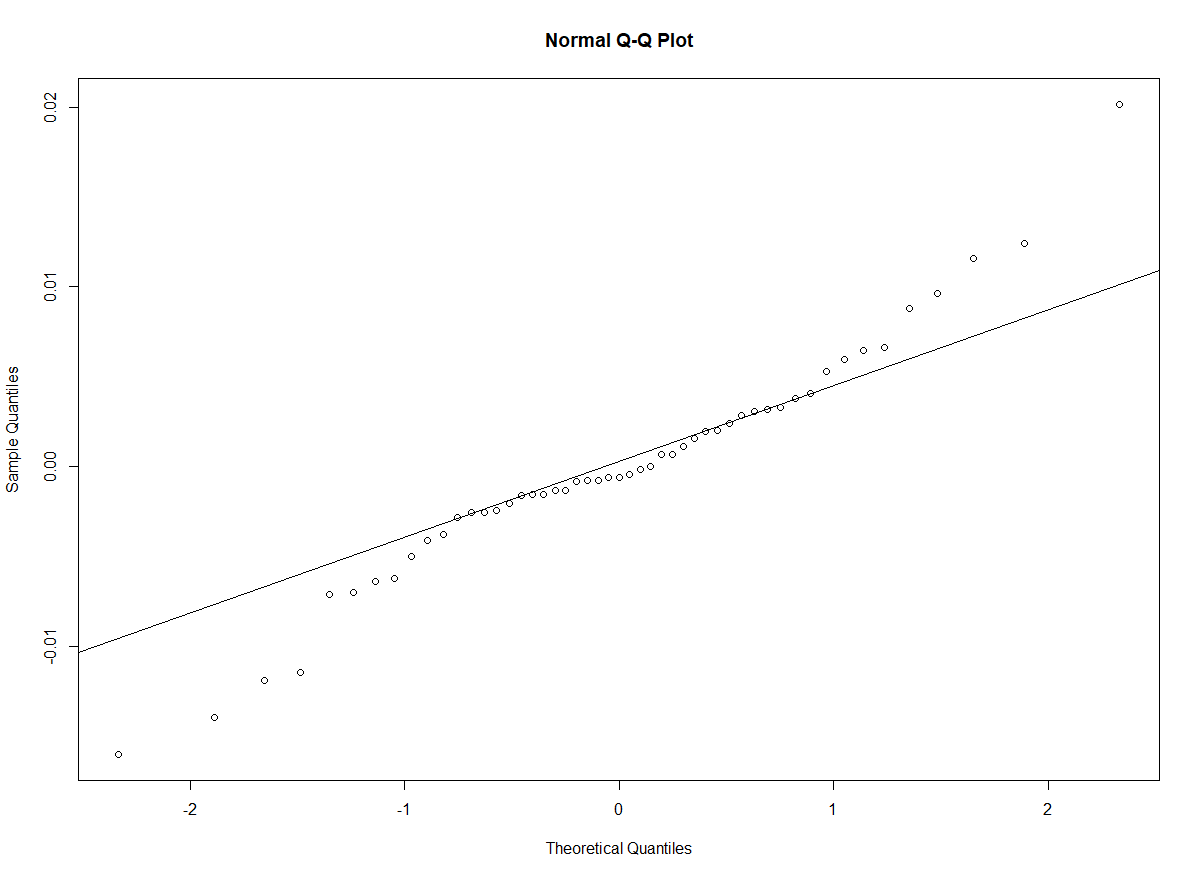
-0.0206812 0.3404977 -0.0707477

tmk1 wlp1

0.5612484 0.1529316

plot(resr2$fitted.values,resr2$residuals)





> influencePlot(model2)

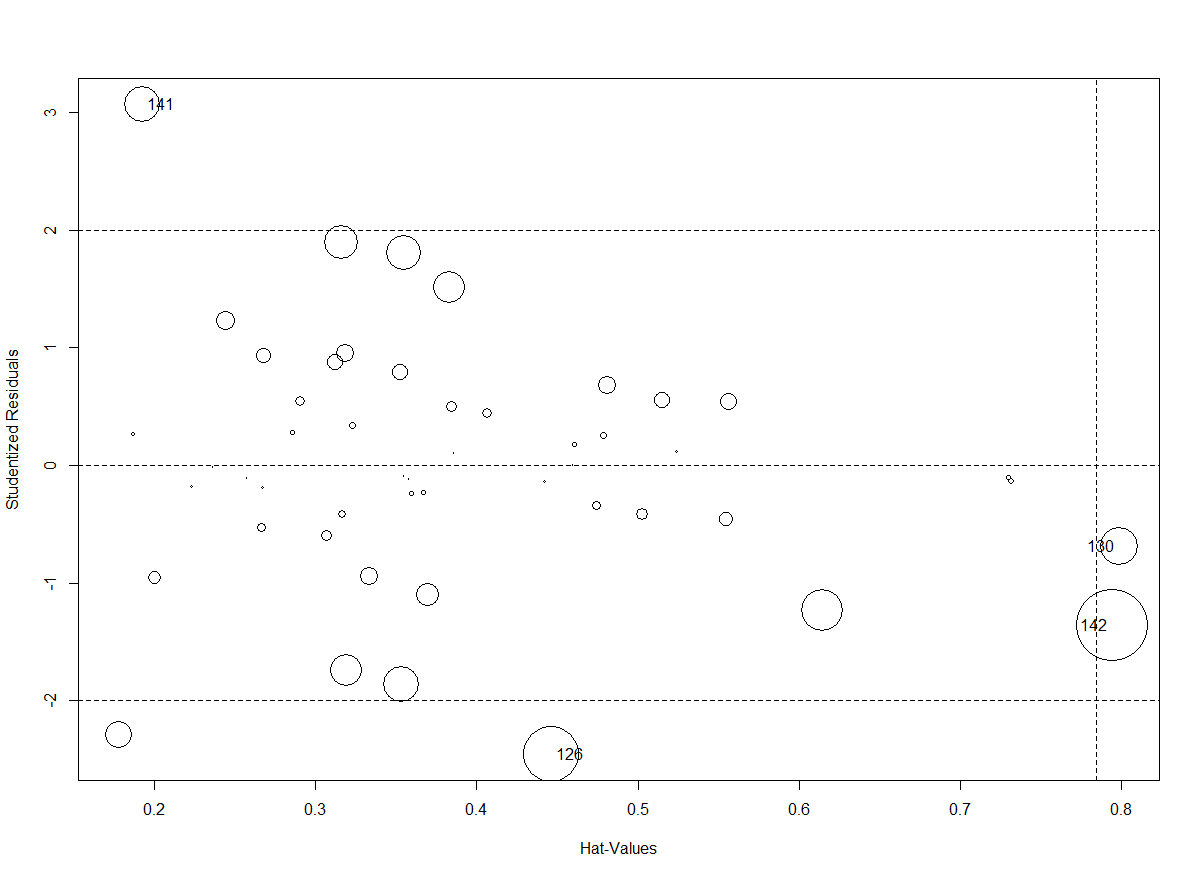
StudRes Hat CookD

126 -2.4533328 0.4466375 0.20905490

130 -0.6849839 0.7986771 0.09469115

141 3.0659483 0.1929531 0.08841332

142 -1.3625481 0.7942711 0.34874685



Проаналізуємо нашу модель, вилучив викиди:

> model22<-lm(afl1~.-afl1,data=data2[-c(14,18,49,50),])

> summary(model22)

Call:

lm(formula = afl1 ~ . - afl1, data = data2[-c(14, 18, 49, 50),

])

Residuals:

Min 1Q Median 3Q

-0.0167707 -0.0026691 0.0000127 0.0037518

Max

0.0101301

Coefficients:

Estimate Std. Error t value

(Intercept) 0.0004792 0.0013227 0.362

abbv1 -0.1003057 0.0955142 -1.050

abt1 0.0638770 0.1144503 0.558

aig1 0.3154028 0.1705502 1.849

aon1 -0.3028525 0.2200316 -1.376

bax1 0.1622083 0.1985109 0.817

bsx1 -0.0593668 0.0644662 -0.921

cb1 0.1069230 0.2804564 0.381

cinf1 -0.2742820 0.2427833 -1.130

dhr1 0.1293208 0.1767352 0.732

ew1 -0.0916383 0.0898139 -1.020

hig1 -0.1169722 0.1878646 -0.623

isrg1 0.0276741 0.0468216 0.591

mdt1 -0.1180287 0.1651523 -0.715

mmc1 0.4153818 0.3134393 1.325

pbct1 0.0189876 0.1574116 0.121

pfg1 0.2776001 0.2147478 1.293

pru1 -0.0001018 0.2030425 -0.001

tmk1 0.4707586 0.3856584 1.221

wlp1 0.0198917 0.1616862 0.123

Pr(>|t|)

(Intercept) 0.7199

abbv1 0.3030

abt1 0.5814

aig1 0.0754 .

aon1 0.1800

bax1 0.4210

bsx1 0.3653

cb1 0.7060

cinf1 0.2685

dhr1 0.4706

ew1 0.3166

hig1 0.5387

isrg1 0.5594

mdt1 0.4809

mmc1 0.1962

pbct1 0.9049

pfg1 0.2071

pru1 0.9996

tmk1 0.2328

wlp1 0.9030

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.00762 on 27 degrees of freedom

Multiple R-squared: 0.6956, Adjusted R-squared: 0.4813

F-statistic: 3.247 on 19 and 27 DF, p-value: 0.002624

Коєфіцієнт детермінації збільшився, але не покращив наш аналіз.

**Таким чином з log-returns даними ми не можемо нічого спрогнузувати.**

**Зробимо аналіз без log-returns**

1) по всім даних, крім останніх 10

> model111<-lm(afl~.-afl,data=Data[1:(nn-10),1:20])

> summary(model111)

Call:

lm(formula = afl ~ . - afl, data = Data[1:(nn - 10), 1:20])

Residuals:

Min 1Q Median 3Q Max

-1.72835 -0.47516 -0.04421 0.40844 1.65056

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 35.070836 7.652318 4.583 1.12e-05 \*\*\*

abbv 0.217595 0.076088 2.860 0.004988 \*\*

abt -0.180233 0.133256 -1.353 0.178705

aig 0.079365 0.102433 0.775 0.439953

aon 0.147061 0.072369 2.032 0.044314 \*

bax 0.437464 0.085943 5.090 1.32e-06 \*\*\*

bsx -0.548092 0.318257 -1.722 0.087573 .

cb -0.118100 0.116096 -1.017 0.311038

cinf -0.496089 0.185057 -2.681 0.008362 \*\*

dhr 0.083271 0.095882 0.868 0.386839

ew -0.048819 0.019949 -2.447 0.015823 \*

hig -0.491455 0.133738 -3.675 0.000355 \*\*\*

isrg 0.001371 0.003254 0.421 0.674278

mdt -0.214500 0.137708 -1.558 0.121908

mmc -0.130154 0.233074 -0.558 0.577579

pbct -1.493740 0.333015 -4.486 1.66e-05 \*\*\*

pfg 0.088184 0.174849 0.504 0.614928

pru 0.429918 0.086217 4.986 2.06e-06 \*\*\*

tmk 0.258889 0.179276 1.444 0.151279

wlp 0.153368 0.057551 2.665 0.008743 \*\*

---

Signif. codes:

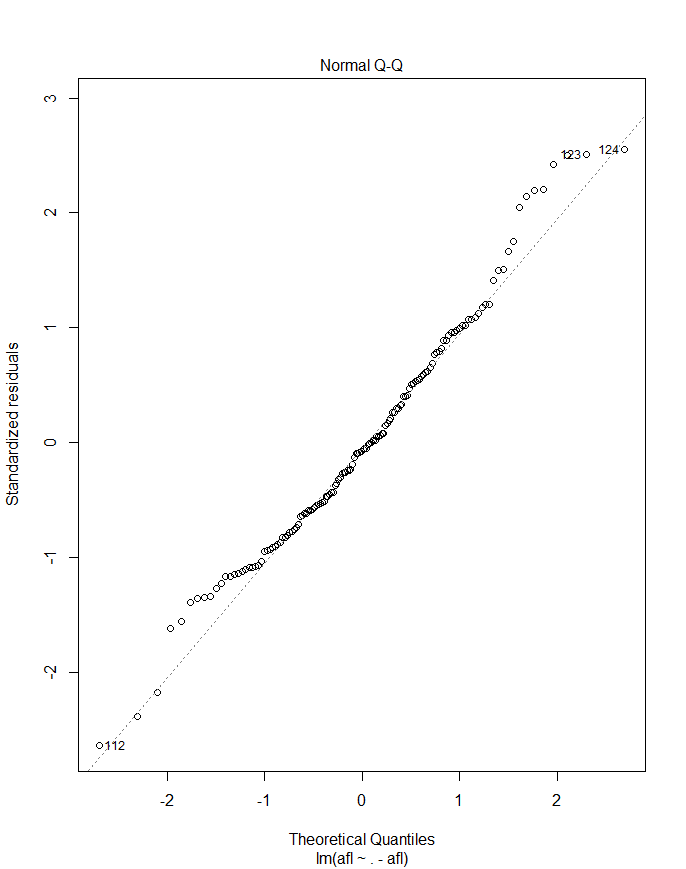
0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

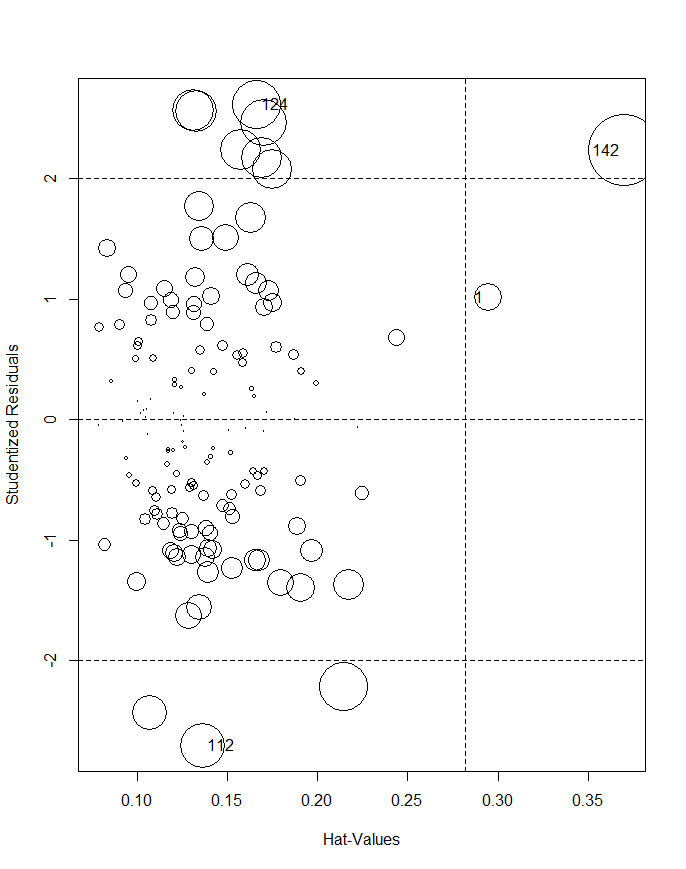
Residual standard error: 0.7054 on 122 degrees of freedom

Multiple R-squared: 0.962, Adjusted R-squared: 0.9561

F-statistic: 162.5 on 19 and 122 DF, p-value: < 2.2e-16

Коефіціент детермінації значно кращий, він є більшим за 0,8, а рівень значущості прямую до нуля, тому ця модель дає нам змогу зробити висновки про певну залежність.





2) по 50 даним

> model<-lm(afl~.-afl,data=Data[(nn-60):(nn-10),1:20])

> summary(model)

Call:

lm(formula = afl ~ . - afl, data = Data[(nn - 60):(nn - 10),

1:20])

Residuals:

Min 1Q Median 3Q Max

-0.69893 -0.22119 -0.02222 0.21122 0.78662

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 21.489497 10.274986 2.091 0.0448 \*

abbv -0.073206 0.130581 -0.561 0.5791

abt -0.039617 0.185408 -0.214 0.8322

aig 0.193173 0.251441 0.768 0.4481

aon -0.139569 0.170150 -0.820 0.4183

bax 0.001828 0.151664 0.012 0.9905

bsx 0.202912 0.418928 0.484 0.6315

cb -0.265464 0.170771 -1.555 0.1302

cinf -0.240018 0.244014 -0.984 0.3329

dhr 0.329534 0.138844 2.373 0.0240 \*

ew 0.048838 0.071592 0.682 0.5002

hig 0.146615 0.339830 0.431 0.6691

isrg 0.001354 0.006120 0.221 0.8264

mdt 0.089480 0.210583 0.425 0.6738

mmc 0.199899 0.364909 0.548 0.5877

pbct 0.170991 0.448797 0.381 0.7058

pfg 0.773035 0.281749 2.744 0.0100 \*

pru -0.044356 0.138119 -0.321 0.7503

tmk -0.036829 0.352813 -0.104 0.9175

wlp 0.053950 0.132273 0.408 0.6862

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4182 on 31 degrees of freedom

Multiple R-squared: 0.9627, Adjusted R-squared: 0.9398

F-statistic: 42.09 on 19 and 31 DF, p-value: < 2.2e-16

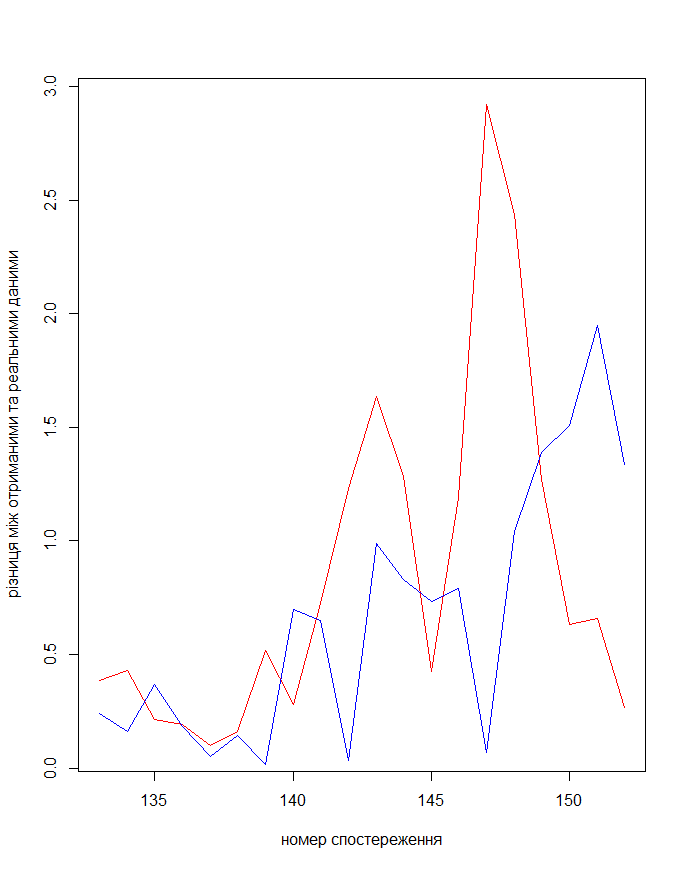
Коефіцієнт детермінації ще вижче ніж у попередній моделі, а рівень значущості прямує до 0.

Тепер подивимося как наші моделі працюють на останніх 20 даних, оцінимо теоритичні та практичні отримані дані.

|  |  |  |
| --- | --- | --- |
| Реальні дані з таблиці | Model 1 | Model 2 |
| 59.11 | 59.49414 | 58.87113 |
| 58.86 | 59.28865 | 59.02138 |
| 59.17 | 59.38505 | 58.80033 |
| 59.08 | 58.88644 | 59.26433 |
| 59.26 | 59.16061 | 59.31156 |
| 59.42 | 59.57990 | 59.5611 |
| 60.15 | 59.63347 | 60.16508 |
| 59.72 | 59.99914 | 60.41893 |
| 60.91 | 60.18367 | 60.25828 |
| 60.42 | 59.19013 | 60.45134 |
| 61.09 | 59.45372 | 62.08055 |
| 60.98 | 59.69397 | 61.8123 |
| 60.85 | 60.42354 | 61.58365 |
| 61.61 | 60.41834 | 62.39901 |
| 63.58 | 60.65712 | 63.64654 |
| 62.28 | 59.84380 | 63.32067 |
| 62.07 | 60.80110 | 63.45984 |
| 61.68 | 61.04724 | 63.1885 |
| 61.30 | 60.64356 | 63.24891 |
| 61.92 | 61.65553 | 63.25773 |

Як бачимо за даними модель 1 є більш точніше к практичним даним, ніж друга.

Побудуємо різницю між теоретичними та практичними даними на одному малюнку, щоби краще бачити.



Таким чином висновок нашої роботи, що дані без log-returns є більш кращі для аналізу впливу, де ми можемо зробити висновок, що модель 1 є краще ніж друга з високим коєфіціентом детермінації та рівнем значущості, який прямує до 0.