

Budowa modelu SUR

Jakub Augustynek

12/10/2019

Celem tego projektu jest zbudowanie modelu SUR, czyli układu pozornie niepowiązanych równań regresji. Model zbudowany będzie dla stóp zwrotu z indeksów giełdowych sektorów takich jak:

- banki
- budownictwo
- chemia
- energia
- info
- media
- nieruchomości
- paliwa
- spożywczy
- telkom

Wszystkie z indeksów bazują na polskiej giełdzie, więc czynniki losowe równań regresji mogą być skorelowane. Stopy zwrotu z indeksów sektorowych zostaną objaśnione za pomocą stóp zwrotu dla indeksu WIG. W celu sprawdzenia istotności wyestymowanych parametrów obliczona zostanie statystyka GRS.

Dane pobrane zostały ze strony (<https://stooq.pl>). Pobrano miesięczne dane historyczne indeksów od stycznia 2010 roku do grudnia 2019. Wśród indeksów jest WIG oraz 10 indeksów sektorowych.

```
#wczytywanie danych
library(readr)
library(kableExtra)
#indeks wig
wig <- read_csv("data/wig_m 2.csv")

#indeksy sektorowe
wig_banki <- read_csv("data/wig_banki.csv")
wig_budow <- read_csv("data/wig_budow.csv")
wig_chemia <- read_csv("data/wig_chemia.csv")
wig_energ <- read_csv("data/wig_energ.csv")
wig_info <- read_csv("data/wig_info.csv")
wig_info <- wig_info[(226:344),]
wig_media <- read_csv("data/wig_media.csv")
wig_nrchom <- read_csv("data/wig_nrchom.csv")
wig_paliwa <- read_csv("data/wig_paliwa.csv")
wig_spozyw <- read_csv("data/wig_spozyw.csv")
wig_telkom <- read_csv("data/wig_telkom.csv")

kable(head(wig))
```

Data	Otwarcie	Najwyzszy	Najnizszy	Zamkniecie	Wolumen
2010-01-31	40156.1	41725.2	39634.3	40058.3	1545940868
2010-02-28	39828.6	40226.3	37038.9	38708.6	1600201708
2010-03-31	38937.6	42731.4	38702.2	42446.5	2918507628
2010-04-30	42644.8	44302.5	42131.9	43295.2	1667119190
2010-05-31	43527.9	43533.8	38900.2	41530.1	1884074256
2010-06-30	41336.0	41425.1	39269.6	39392.5	1583407086

Dla wartości zamknięcia obliczono logarytmiczne stopy zwrotu.

```
stopy<-matrix(NA, (nrow(wig)-1), 11)

for(i in 1:(nrow(wig)-1))
{
  stopy[i,1]=log(wig$Zamkniecie[i+1]/wig$Zamkniecie[i])
  stopy[i,2]=log(wig_banki$Zamkniecie[i+1]/wig_banki$Zamkniecie[i])
  stopy[i,3]=log(wig_budow$Zamkniecie[i+1]/wig_budow$Zamkniecie[i])
  stopy[i,4]=log(wig_chemia$Zamkniecie[i+1]/wig_chemia$Zamkniecie[i])
  stopy[i,5]=log(wig_energ$Zamkniecie[i+1]/wig_energ$Zamkniecie[i])
  stopy[i,6]=log(wig_info$Zamkniecie[i+1]/wig_info$Zamkniecie[i])
  stopy[i,7]=log(wig_media$Zamkniecie[i+1]/wig_media$Zamkniecie[i])
  stopy[i,8]=log(wig_nrchom$Zamkniecie[i+1]/wig_nrchom$Zamkniecie[i])
  stopy[i,9]=log(wig_paliwa$Zamkniecie[i+1]/wig_paliwa$Zamkniecie[i])
  stopy[i,10]=log(wig_spozyw$Zamkniecie[i+1]/wig_spozyw$Zamkniecie[i])
  stopy[i,11]=log(wig_telkom$Zamkniecie[i+1]/wig_telkom$Zamkniecie[i])
}

colnames(stopy)<-c("wig", "banki", "budow", "chemia", "energ", "info", "media", "nrchom", "paliwa", "spoz", "telkom")
```

Następnie stopy zwrotu skorygowano - stopa wolna od ryzyka (3%).

```
for(i in 1:(nrow(wig)-1))
{
  stopy[i,1]=stopy[i,1]-((0.03)/12)
  stopy[i,2]=stopy[i,2]-((0.03)/12)
  stopy[i,3]=stopy[i,3]-((0.03)/12)
  stopy[i,4]=stopy[i,4]-((0.03)/12)
  stopy[i,5]=stopy[i,5]-((0.03)/12)
  stopy[i,6]=stopy[i,6]-((0.03)/12)
  stopy[i,7]=stopy[i,7]-((0.03)/12)
  stopy[i,8]=stopy[i,8]-((0.03)/12)
  stopy[i,9]=stopy[i,9]-((0.03)/12)
  stopy[i,10]=stopy[i,10]-((0.03)/12)
  stopy[i,11]=stopy[i,11]-((0.03)/12)
}

head(stopy) %>% kable()
```

wig	banki	budow	chemia	energ	info	media	nrchom	paliwa	spoz	telkom
-0.0367741	-0.0673888	-0.0230061	0.0129041	-0.0630758	-0.0607573	0.0639266	-0.0570885	-0.0585458	0.0129041	0.0129041
0.0896827	0.0862630	0.1267781	0.1389166	0.0454283	0.0514012	0.1024205	0.1332905	0.1305096	0.1389166	0.1389166
0.0172973	0.0414225	0.0117321	-0.0374329	-0.0053205	-0.0104859	0.0120774	-0.0015822	0.0132704	-0.0374329	-0.0374329
-0.0441233	-0.0223866	-0.0650704	-0.0419230	-0.0329863	-0.0194948	-0.0388926	-0.0448005	-0.0256537	-0.0419230	-0.0419230
-0.0553430	-0.0887844	-0.0424300	-0.0618733	-0.0121797	-0.0508488	-0.0665988	-0.0547180	-0.0713108	-0.0618733	-0.0618733
0.0725977	0.0542138	0.0618713	0.1006164	0.0233241	0.0230889	0.0887543	0.0625596	0.0825190	0.1006164	0.1006164

Następnym etapem jest utworzenie modeli MNK, w których stopy zwrotu indeksów sektorowych są wyjaśniane przez stopy zwrotu indeksu głównego (WIG).

```
#estymowanie rownanie po rownaniu mnk
m1<-lm(stopy[, "banki"] ~ stopy[, "wig"])
m2<-lm(stopy[, "budow"] ~ stopy[, "wig"])
m3<-lm(stopy[, "chemia"] ~ stopy[, "wig"])
m4<-lm(stopy[, "energ"] ~ stopy[, "wig"])
m5<-lm(stopy[, "info"] ~ stopy[, "wig"])
m6<-lm(stopy[, "media"] ~ stopy[, "wig"])
m7<-lm(stopy[, "nrchom"] ~ stopy[, "wig"])
m8<-lm(stopy[, "paliwa"] ~ stopy[, "wig"])
m9<-lm(stopy[, "spozyw"] ~ stopy[, "wig"])
m10<-lm(stopy[, "telkom"] ~ stopy[, "wig"])
```

Dla wyestymowanych modeli obliczona będzie statystyka GRS. Aby obliczyć jej wartość, potrzebna będzie:

- macierz sigma - macierz wariancji-kowariancji,
- wektor współczynników alfa, a także
- średnia i wariancja wartości stóp zwrotu indeksu WIG.

Wartości reszt, będących różnicą pomiędzy wartością empiryczną, a teoretyczną zmiennej objaśnianej, a także wartości współczynników alfa zapisano w macierzy. Obliczono średnią oraz odwrotność wariancji dla stóp zwrotu indeksu wig.

```
reszty<-matrix(0, (nrow(wig)-1), 10)
for(i in 1:(nrow(wig)-1))
{
  reszty[i,1]<-m1$residuals[i]
  reszty[i,2]<-m2$residuals[i]
  reszty[i,3]<-m3$residuals[i]
  reszty[i,4]<-m4$residuals[i]
  reszty[i,5]<-m5$residuals[i]
  reszty[i,6]<-m6$residuals[i]
  reszty[i,7]<-m7$residuals[i]
  reszty[i,8]<-m8$residuals[i]
  reszty[i,9]<-m9$residuals[i]
  reszty[i,10]<-m10$residuals[i]
}
colnames(reszty)<-c("banki", "budow", "chemia", "energ", "info", "media", "nrchom", "paliwa", "spozyw", "telkom")

#zapisanie wartosci wsp alfa wyestymowanych modeli
alfa <-matrix(data = 0, nrow = 10, ncol = 1)
alfa[1,1] <- m1$coefficients[1]
alfa[2,1] <- m2$coefficients[1]
alfa[3,1] <- m3$coefficients[1]
alfa[4,1] <- m4$coefficients[1]
alfa[5,1] <- m5$coefficients[1]
alfa[6,1] <- m6$coefficients[1]
alfa[7,1] <- m7$coefficients[1]
alfa[8,1] <- m8$coefficients[1]
alfa[9,1] <- m9$coefficients[1]
alfa[10,1] <- m10$coefficients[1]
rownames(alfa)<-c("banki", "budow", "chemia", "energ", "info", "media", "nrchom", "paliwa", "spozyw", "telkom")

sr_wig=(mean(stopy[,1]))
```

```
wariancja<-1/var(stopy[,1])
```

Za pomocą testu ADF sprawdzono, czy reszty modeli są stacjonarne. Hipoteza zerowa tego testu oznacza niestacjonarność.

```
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
```

```
##   method      from
```

```
##   as.zoo.data.frame zoo
```

```
adf.test(reszty[,1],alternative = c("stationary", "explosive"))
```

```
##
```

```
##   Augmented Dickey-Fuller Test
```

```
##
```

```
## data:  reszty[, 1]
```

```
## Dickey-Fuller = -4.8985, Lag order = 4, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

```
adf.test(reszty[,2],alternative = c("stationary", "explosive"))
```

```
##
```

```
##   Augmented Dickey-Fuller Test
```

```
##
```

```
## data:  reszty[, 2]
```

```
## Dickey-Fuller = -3.6113, Lag order = 4, p-value = 0.03527
```

```
## alternative hypothesis: stationary
```

```
adf.test(reszty[,3],alternative = c("stationary", "explosive"))
```

```
##
```

```
##   Augmented Dickey-Fuller Test
```

```
##
```

```
## data:  reszty[, 3]
```

```
## Dickey-Fuller = -5.5314, Lag order = 4, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

```
adf.test(reszty[,4],alternative = c("stationary", "explosive"))
```

```
##
```

```
##   Augmented Dickey-Fuller Test
```

```
##
```

```
## data:  reszty[, 4]
```

```
## Dickey-Fuller = -5.3162, Lag order = 4, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

```
adf.test(reszty[,5],alternative = c("stationary", "explosive"))
```

```
##
```

```
##   Augmented Dickey-Fuller Test
```

```
##
```

```
## data:  reszty[, 5]
```

```
## Dickey-Fuller = -4.1941, Lag order = 4, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

```
adf.test(reszty[,6],alternative = c("stationary", "explosive"))
```

```
##
```

```
## Augmented Dickey-Fuller Test
##
## data: reszty[, 6]
## Dickey-Fuller = -4.1072, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
adf.test(reszty[,7],alternative = c("stationary", "explosive"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: reszty[, 7]
## Dickey-Fuller = -4.7918, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
adf.test(reszty[,8],alternative = c("stationary", "explosive"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: reszty[, 8]
## Dickey-Fuller = -3.8551, Lag order = 4, p-value = 0.01871
## alternative hypothesis: stationary
adf.test(reszty[,9],alternative = c("stationary", "explosive"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: reszty[, 9]
## Dickey-Fuller = -3.9101, Lag order = 4, p-value = 0.01603
## alternative hypothesis: stationary
adf.test(reszty[,10],alternative = c("stationary", "explosive"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: reszty[, 10]
## Dickey-Fuller = -4.0585, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

Przyjęto poziom istotności 0,05. Wartość p dla wykonanych testów jest mniejsza od poziomu istotności, oznacza to, że hipotezę zerową, mówiącą o niestacjonarności należy odrzucić, reszty modelu są stacjonarne.

Obliczono macierz sigma - macierz wariancji - kowariancji.

```
#macierz sigma
dl<-nrow(wig)-1
sigma<-matrix(NA, 10, 10)
reszty<-as.matrix(reszty)

sigma<-t(reszty)%*%reszty/dl
sigma
```

```
##          banki          budow          chemia          energ          info
## banki    7.076323e-04 -2.175556e-04 -5.068279e-04 -0.0002756955 -0.0001419834
## budow   -2.175556e-04  3.872831e-03  1.395289e-04 -0.0005076671  0.0008974966
## chemia  -5.068279e-04  1.395289e-04  3.872447e-03  0.0003018506 -0.0001364417
```

```
## energ -2.756955e-04 -5.076671e-04 3.018506e-04 0.0020201993 -0.0001550784
## info -1.419834e-04 8.974966e-04 -1.364417e-04 -0.0001550784 0.0013825784
## media -1.426039e-05 9.568831e-04 -9.308179e-06 -0.0002739873 0.0002502530
## nrchom -1.257074e-04 1.135035e-03 6.831951e-06 -0.0002748829 0.0001989325
## paliwa -2.763905e-04 -7.677986e-04 1.893245e-05 0.0001124451 -0.0002191321
## spozyw -2.318364e-04 3.536281e-04 2.500484e-04 -0.0003053458 0.0002862954
## telkom -2.378461e-04 5.278192e-05 3.013395e-04 0.0006038904 0.0002497739
##          media          nrchom          paliwa          spozyw          telkom
## banki -1.426039e-05 -1.257074e-04 -2.763905e-04 -0.0002318364 -2.378461e-04
## budow 9.568831e-04 1.135035e-03 -7.677986e-04 0.0003536281 5.278192e-05
## chemia -9.308179e-06 6.831951e-06 1.893245e-05 0.0002500484 3.013395e-04
## energ -2.739873e-04 -2.748829e-04 1.124451e-04 -0.0003053458 6.038904e-04
## info 2.502530e-04 1.989325e-04 -2.191321e-04 0.0002862954 2.497739e-04
## media 1.485369e-03 4.136919e-04 -7.285857e-04 0.0002856930 1.329213e-04
## nrchom 4.136919e-04 1.449258e-03 -1.505102e-04 0.0001971614 -1.282630e-04
## paliwa -7.285857e-04 -1.505102e-04 2.122901e-03 -0.0003753232 -1.023526e-03
## spozyw 2.856930e-04 1.971614e-04 -3.753232e-04 0.0021637382 -2.378322e-04
## telkom 1.329213e-04 -1.282630e-04 -1.023526e-03 -0.0002378322 4.246667e-03
```

Kowariancja nie przyjmuje wartości zerowych, oznacza to, że składniki losowe modeli są ze sobą powiązane. Obliczenie statystyki GRS da odpowiedź na pytanie, czy indeks WIG jest jedyną zmienną wpływającą na zmianę wartości indeksów sektorowych. Hipoteza zerowa statystyki GRS mówi o tym, że wartości współczynników alfa są równe 0.

```
T=(nrow(wig)-1)
N=10
L=1
# wektor wsp alfa
talfa<-t(alfa)
# odwrocona macierz sigma
odwr<-solve(sigma)

#statystyka GRS
GRS<-T/N * ((T-N-L)/(T-L-1)) * ((talfa%*%odwr%*%alfa)/(1+sr_wig*wariancja*sr_wig))
GRS

##          [,1]
## [1,] 2.081158

#wartość p
pf(GRS, N, T-N-L, lower.tail = FALSE)

##          [,1]
## [1,] 0.03210869
```

Wartość p dla przeprowadzonego testu jest mniejsza od poziomu istotności, hipotezę zerową mówiącą, że wartość współczynników alfa jest równa 0 należy więc odrzucić. Świadczy to o tym, że sektory nie są powiązane tylko z indeksem WIG.