

Analiza ekonometryczna spółki Amazon

01/01/2018-01/03/2022

CZERWIEC 5

Authored by: Daniel Pochopień



Logo
Name

Cele pracy

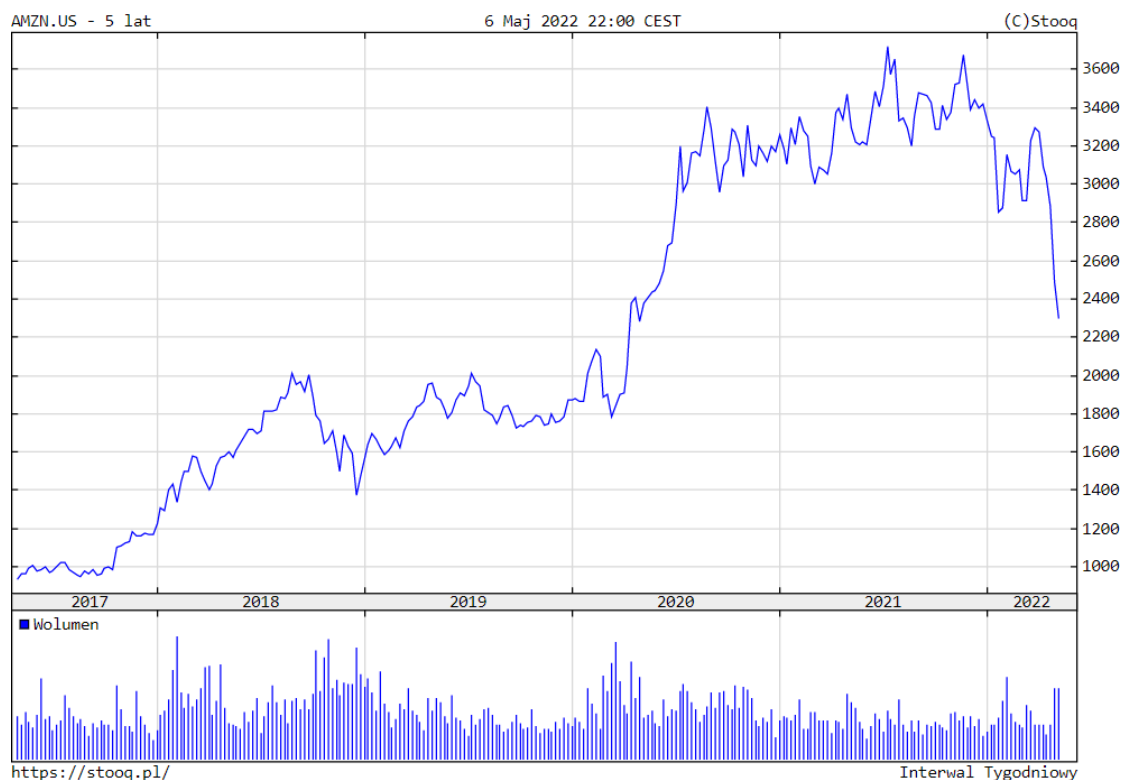
Celem projektu jest analiza ekonometryczna spółki Amazon, analiza szeregu czasowego, badanie stacjonarności, budowa modelu ARMA/ARIMA, badanie efektu ARCH w powstałym modelu, dopasowanie modelu GARCH w przypadku występowania efektu ARCH, prognoza log. stóp procentowych, logarytmów cen i cen oraz późniejsze porównanie z rzeczywistymi wynikami.

Podstawowe informacje o spółce

Amazon.com, Inc. – amerykańskie przedsiębiorstwo handlowe, spółka akcyjna założona w 1994 w Seattle. Zajmuje się handlem elektronicznym (e-handlem) B2C i prowadzi największy na świecie sklep internetowy.

Wstępna analiza szeregu czasowego

Do przeprowadzenia analizy i budowy modelu wybrano dane historyczne z przedziału czasowego - 01/01/2018 - 01/03/2022. Do badania wykorzystano logarymiczne dzienne stopy zwrotu w konfiguracji **Zamknięcie - Zamknięcie**.



Badanie stacjonarności

Pierwszym krokiem było zbadanie stacjonarności szeregu czasowego. Wykorzystano do tego test ADF (Augmented Dickey-Fuller). Za poziom istotności przyjęto $\alpha = 0,05$.

Test ADF:

H0: Proces jest niestacjonarny

H1: Proces jest stacjonarny

Augmented Dickey-Fuller Test

```
data: ama_d$logstopy
Dickey-Fuller = -10.848, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

Osiągnięta wartość p-value < 0,05 oznacza, że hipoteza zerowa może zostać odrzucona - proces jest niestacjonarny.

Budowa modelu ARMA/ARIMA

Kolejnym krokiem projektu było dopasowanie konkretnego modelu ARMA/ARIMA. Tworzymy go poprzez ustalenie odpowiedniego rzędu opóźnień dla procesu autoregresyjnego i dla procesu średniej ruchomej oraz ewentualnego usunięcia trendu za pomocą różnicowania. Model z najmniejszą wartością dla danego kryterium informacyjnego (w naszym przypadku AIC) jest modelem najlepiej dopasowanym.

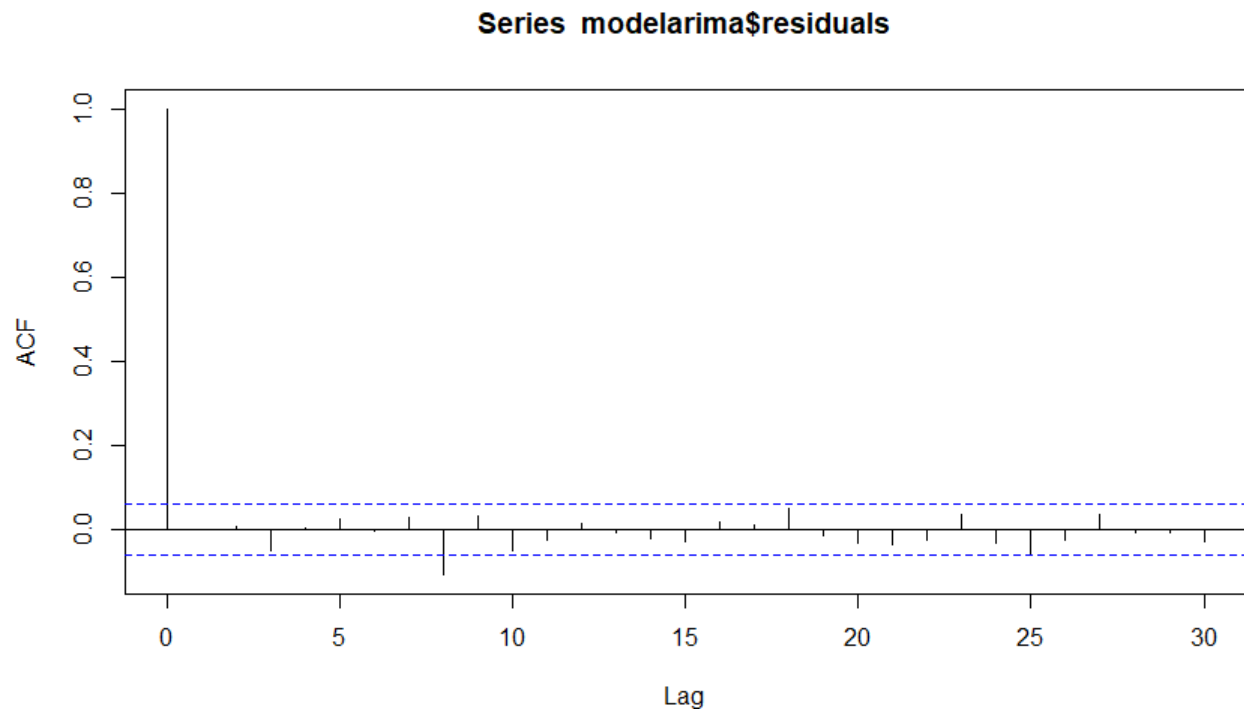
```
#Budowa modelu ARMA/ARIMA
modelarima <- auto.arima(ama_d$logstopy) # ARIMA(0,0,1)
```

Funkcja auto.arima wyliczyła najlepszy model z najmniejszą wartością dla kryterium informacyjnego aic - jest to model ARIMA(0,0,1).

```
#model ARIMA z najmniejszym aic to model ARIMA(0,0,1) - aic = -5198,657
arima1 <- arima(ama_d$logstopy, order = c(0,0,1))
```

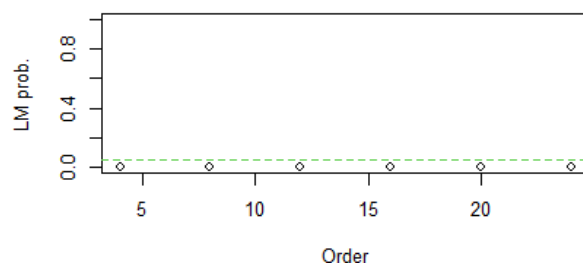
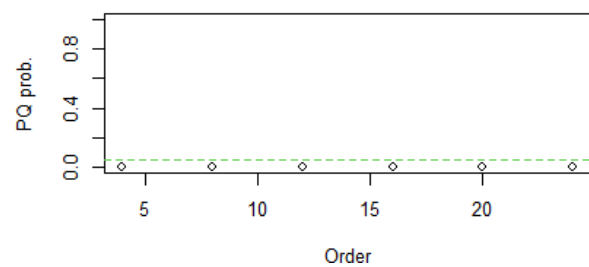
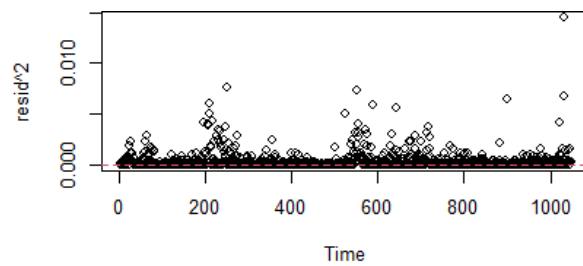
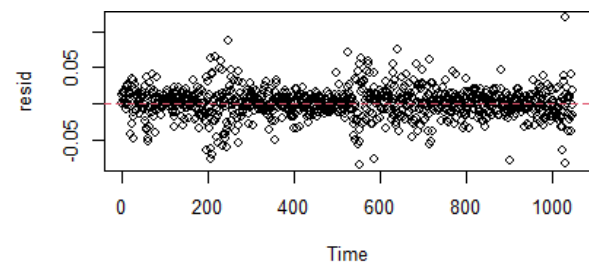
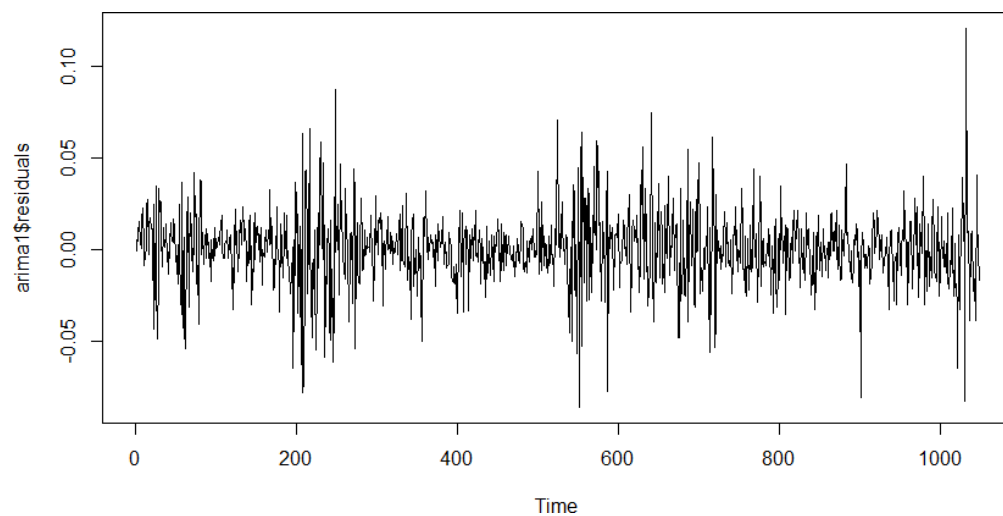
Badanie autokorelacji reszt modelu

Do zbadania autokorelacji reszt modelu użyto funkcji ACF. Pozwala ona określić, kiedy wartość obecna jest zależna od wartości poprzednich. Rezultaty przedstawiono na poniższym wykresie. Dane pokazują wartości współczynnika autokorelacji w zależności od danego opóźnienia.



Reszty dla ponad 90% procent opóźnień nie są ze sobą skorelowane, możemy założyć brak autokorelacji.

Rozkład reszt modelu ARIMA:



Badanie efektu ARCH w składnikach losowych

Kolejnym krokiem było sprawdzenie, czy w składnikach losowych modelu występuje efekt ARCH (*Auto-Regressive Conditional Heteroskedasticity*). Wykorzystano do tego test ARCH, który za hipotezę zerową przyjmuje, że reszty modelu ARIMA są homoskedastyczne. Składa się on z testów Portmanteau Q oraz Mnożnika Lagrange'a. Pierwszy z nich sprawdza, czy reszty podniesione do kwadratu są sekwencją białego szumu, drugi dopasowuje modele regresji liniowej dla reszt podniesionych do kwadratu i sprawdza, czy dopasowany model jest istotny.

Test ARCH

H0: Reszty są homoskedastyczne

H1: Reszty są heteroskedastyczne

```
ARCH heteroscedasticity test for residuals
alternative: heteroscedastic
```

```
Portmanteau-Q test:
```

	order	PQ	p.value
[1,]	4	133	0
[2,]	8	166	0
[3,]	12	224	0
[4,]	16	247	0
[5,]	20	264	0
[6,]	24	281	0

```
Lagrange-Multiplier test:
```

	order	LM	p.value
[1,]	4	696	0.00e+00
[2,]	8	337	0.00e+00
[3,]	12	202	0.00e+00
[4,]	16	144	0.00e+00
[5,]	20	110	9.77e-15
[6,]	24	90	7.24e-10

Wyniki obu testów wynoszą 0 lub są to wartości bardzo do niego zbliżone. Na tej podstawie można odrzucić hipotezę zerową, reszty modelu są heteroskedastyczne.

Na podstawie testu ARCH wykryto heteroskedastyczność reszt modelu, co oznacza występowanie efektu ARCH.

Dopasowanie modelu GARCH

Następnie dla reszt modelu stworzony został model GARCH (1,1) (Generalized Autoregressive Conditional Heteroskedascity). Poniżej przedstawione zostały uzyskane rezultaty.

```
garchmodel <- garch(arima1$residuals, order = c(1, 1))
```

***** ESTIMATION WITH ANALYTICAL GRADIENT *****

I	INITIAL X(I)	D(I)
1	3.675823e-04	1.000e+00
2	5.000000e-02	1.000e+00
3	5.000000e-02	1.000e+00

IT	NF	F	RELDF	PRELDF	RELDX	STPPAR	D*STEP	NPRELDF
0	1	-3.585e+03						
1	7	-3.586e+03	3.53e-04	5.51e-04	2.1e-04	4.4e+09	2.1e-05	1.21e+06
2	8	-3.586e+03	2.19e-05	2.56e-05	2.0e-04	2.0e+00	2.1e-05	1.56e+01
3	15	-3.600e+03	3.80e-03	5.37e-03	3.7e-01	2.0e+00	5.8e-02	1.55e+01
4	19	-3.634e+03	9.48e-03	9.91e-03	8.4e-01	1.8e+00	5.8e-01	5.31e-01
5	30	-3.654e+03	5.28e-03	1.47e-02	2.5e-05	4.0e+00	3.1e-05	6.84e-02
6	40	-3.658e+03	1.34e-03	1.26e-03	3.2e-02	1.9e+00	4.3e-02	9.92e-02
7	46	-3.660e+03	2.90e-04	4.77e-04	4.3e-06	6.5e+00	5.8e-06	9.05e-02
8	47	-3.660e+03	1.11e-06	1.55e-06	4.2e-06	2.0e+00	5.8e-06	4.58e-02
9	55	-3.664e+03	1.16e-03	2.71e-03	6.5e-02	1.8e+00	9.4e-02	4.52e-02
10	56	-3.664e+03	9.96e-05	6.29e-04	1.6e-02	0.0e+00	2.5e-02	6.29e-04
11	57	-3.665e+03	2.18e-04	1.50e-04	7.6e-03	0.0e+00	1.2e-02	1.50e-04
12	58	-3.665e+03	5.93e-05	6.71e-05	7.6e-03	8.7e-01	1.2e-02	1.13e-04
13	59	-3.665e+03	3.54e-05	3.40e-05	6.4e-03	4.7e-01	1.2e-02	3.96e-05
14	60	-3.665e+03	2.31e-05	2.19e-05	6.0e-03	4.4e-01	1.2e-02	2.44e-05
15	61	-3.665e+03	8.22e-06	8.52e-06	4.3e-03	0.0e+00	8.9e-03	8.52e-06
16	62	-3.665e+03	1.97e-07	2.63e-07	3.6e-04	0.0e+00	6.2e-04	2.63e-07
17	63	-3.665e+03	9.06e-08	5.42e-08	2.0e-04	0.0e+00	3.2e-04	5.42e-08
18	66	-3.665e+03	1.67e-10	6.18e-12	2.9e-07	2.0e+00	5.6e-07	2.05e-10
19	69	-3.665e+03	6.09e-12	2.29e-13	1.6e-08	2.0e+00	3.1e-08	2.00e-10
20	71	-3.665e+03	3.62e-13	4.51e-14	3.2e-09	2.0e+00	6.3e-09	1.99e-10
21	78	-3.665e+03	-1.07e-14	7.24e-19	7.8e-15	9.6e+01	1.3e-14	1.98e-10

***** FALSE CONVERGENCE *****

FUNCTION	-3.665420e+03	RELDX	7.830e-15
FUNC. EVALS	78	GRAD. EVALS	21
PRELDF	7.241e-19	NPRELDF	1.984e-10

I	FINAL X(I)	D(I)	G(I)
1	2.356193e-05	1.000e+00	2.098e-01
2	1.519021e-01	1.000e+00	8.502e-04
3	7.941994e-01	1.000e+00	-3.145e-02

Model GARCH jest poprawny, jeżeli reszty modelu oraz ich kwadraty nie są ze sobą skorelowane i mają zadany wcześniej rozkład.

Test Shapiro na normalność reszt modelu GARCH

```
shapiro.test(garchmodel$residuals)
```

```
      shapiro-wilk normality test

data:  garchmodel$residuals
W = 0.97899, p-value = 3.641e-11
```

Wartość parametru p-value < 0.05, co pozwala wnioskować, iż reszty modelu nie mają rozkładu normalnego.

Dla warunkowego rozkładu normalnego autokorelacja (I rzędu) nie występuje, jednak reszty nie mają rozkładu normalnego. Dlatego model ten jest niepoprawny i należy znaleźć inny rozkład brzegowy.

Rozkład warunkowy normalny

```
#warunkowy rozkład normalny
garch_1 <- garchFit(ama_d$logstopy, formula=garch(1,1), cond.dist = "norm")
garch_1$residuals
sd <- sd(garch_1$residuals)
std_residuals <- garch_1$residuals/sd
```

Test normalności:

```
shapiro.test(fGarch::residuals(garch_1))
```

```
      shapiro-wilk normality test

data:  fGarch::residuals(garch_1)
W = 0.95262, p-value < 2.2e-16
```

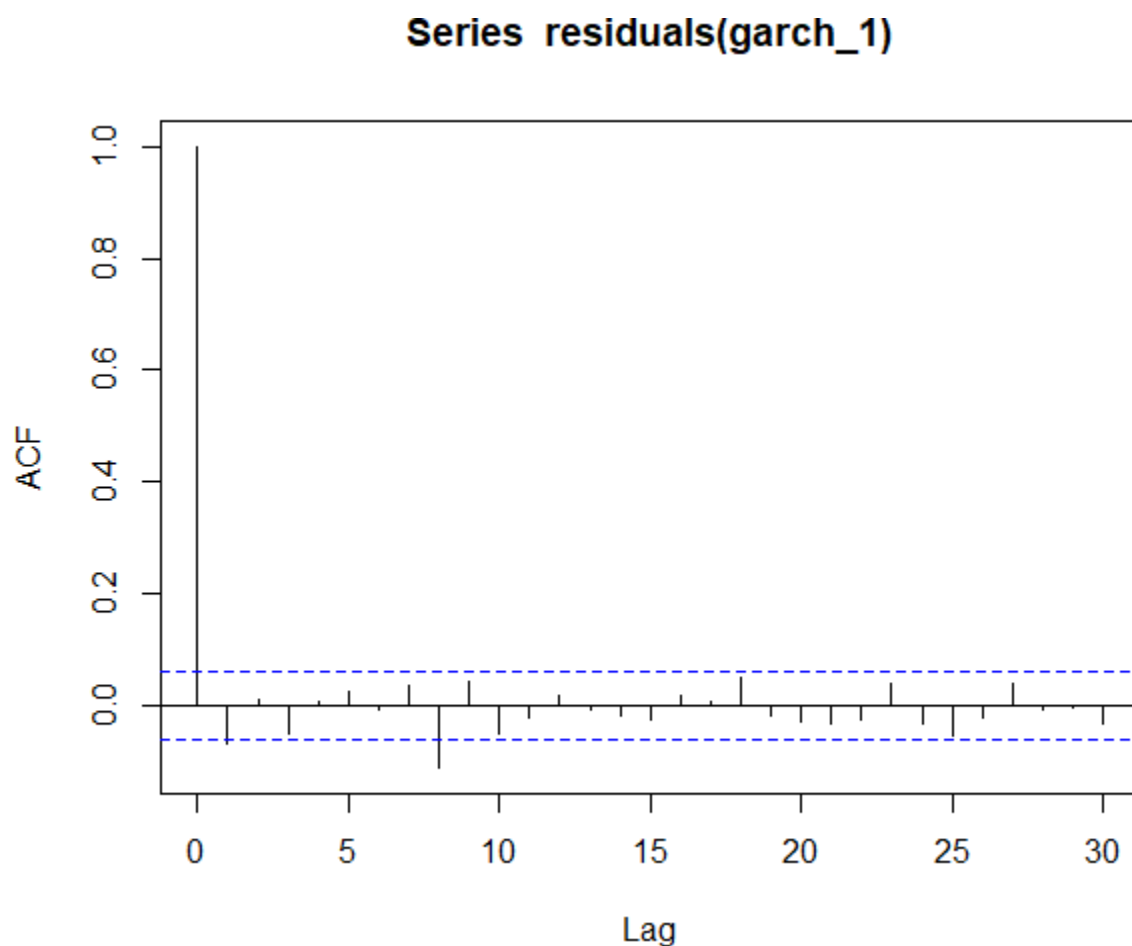

BP test na niezależność szeregu czasowego.

lag	p-value
1	0.0283
2	0.08428
3	0.05905
4	0.112
5	0.1492

lag	p-value
6	0.2272
7	0.2216
8	0.004471
9	0.00375
10	0.002783

Nie ma podstaw do odrzucenia H_0 - szereg czasowy jest niezależny.

Wykres autokorelacji w rozkładzie reszt modelu ARIMA-GARCH dla warunkowego rozkładu normalnego



Dla warunkowego rozkładu normalnego autokorelacja (I rzędu) nie występuje, jednak reszty nie mają rozkładu normalnego. Dlatego model ten jest niepoprawny i należy znaleźć inny rozkład brzegowy.

Rozkład warunkowy T-Studenta

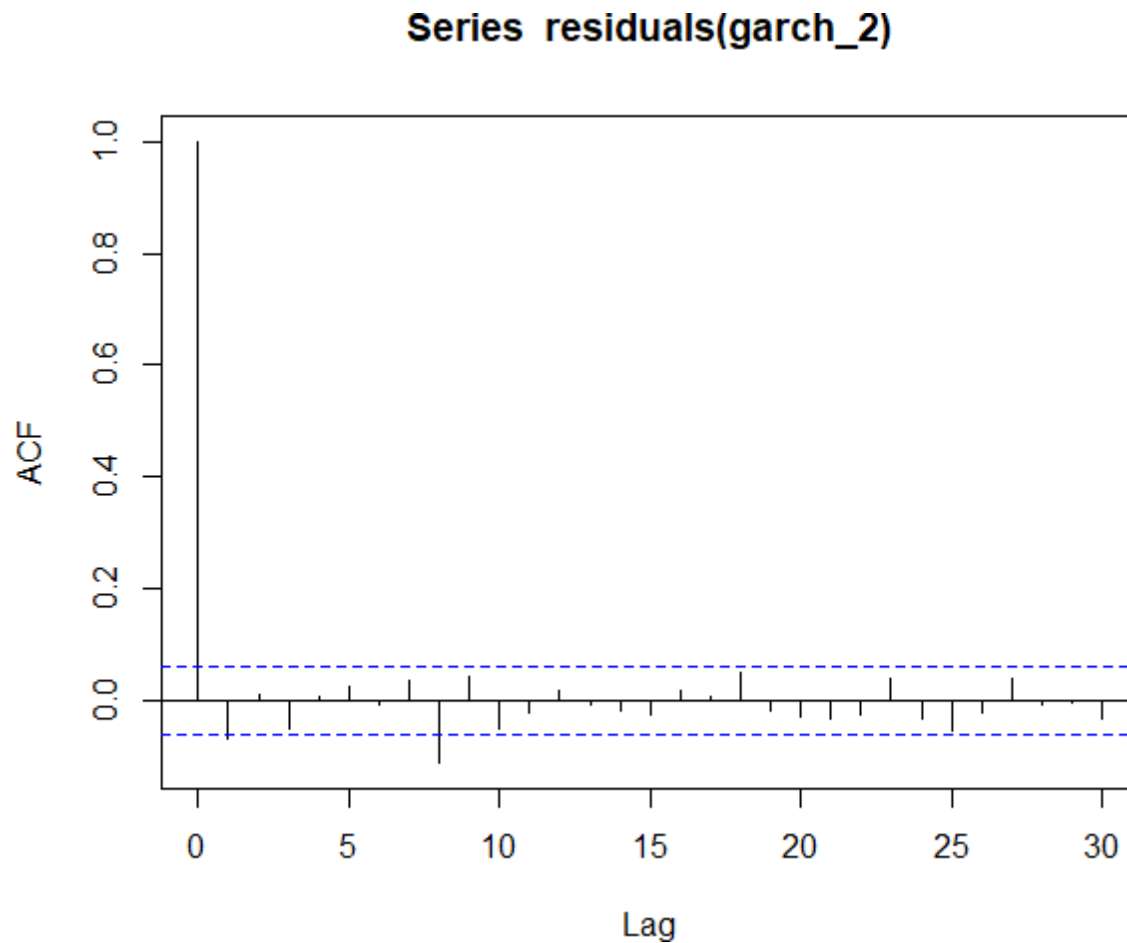
```
garch_cond <- garchFit(dane$stopy, formula=~garch(1,1), cond.dist ="sstd")
```

Test Boxa-Pierce'a

lag	p-value
1	0.0283
2	0.08428
3	0.05905
4	0.112
5	0.1492

lag	p-value
6	0.2272
7	0.2216
8	0.004471
9	0.00375
10	0.002783

Wykres autokorelacji



Test normalności:

```
shapiro-wilk normality test  
data:  fGarch::residuals(garch_2)  
W = 0.95262, p-value < 2.2e-16
```

Dla warunkowego rozkładu t-Studenta nie występuje autokorelacja jednak reszty nie mają warunkowego rozkładu normalnego, ten model również nie jest poprawny.

Prognozy

Prognozy modelu ARIMA dla 5 kolejnych sesji dla cen, zlogarytmowanych cen i logarytmicznych stóp zwrotu wraz z 95% dolnym i górnym przedziałem ufności.

```
#prognozy
arimaforecast <- ATSA::forecast(arima1, lead = 5)

logrates <- c(ama_d$logstopy[1048], arimaforecast[6:10])
logratesupper95 <- c(0, arimaforecast[21:25])
lograteslower95 <- c(0, arimaforecast[16:20])
logprices <- c(log(ama_d$zamkniecie[1048]), 0, 0, 0, 0, 0)
logpricesupper95 <- c(0, 0, 0, 0, 0, 0)
logpriceslower95 <- c(0, 0, 0, 0, 0, 0)
prices <- c(ama_d$zamkniecie[1048], 0, 0, 0, 0, 0)
pricesupper95 <- c(0, 0, 0, 0, 0, 0)
priceslower95 <- c(0, 0, 0, 0, 0, 0)

forecasts <- data.frame(logrates, logratesupper95, lograteslower95, logprices, logpricesupper95, logpriceslower95, prices, pricesupper95, priceslower95)

for (j in 2:6)
{
  forecasts$prices[j] <- exp(forecasts$logrates[j]) * forecasts$prices[j-1]
  forecasts$pricesupper95[j] <- exp(forecasts$logratesupper95[j]) * forecasts$prices[j-1]
  forecasts$priceslower95[j] <- exp(forecasts$lograteslower95[j]) * forecasts$prices[j-1]
  forecasts$logprices[j] <- log(forecasts$prices[j])
  forecasts$logpricesupper95[j] <- log(forecasts$pricesupper95[j])
  forecasts$logpriceslower95[j] <- log(forecasts$priceslower95[j])
}
```

	logrates	logratesupper95	lograteslower95	logprices	logpricesupper95	logpriceslower95	prices	pricesupper95	priceslower95
1	-0.0158911140	0.00000000	0.00000000	8.013952	0.000000	0.000000	3022.840	0.000	0.000
2	0.0020239947	0.04161509	-0.03756710	8.015976	8.055567	7.976385	3028.964	3151.290	2911.387
3	0.0008918805	0.04057211	-0.03878835	8.016868	8.056548	7.977188	3031.667	3154.383	2913.725
4	0.0008918805	0.04057211	-0.03878835	8.017760	8.057440	7.978080	3034.372	3157.198	2916.325
5	0.0008918805	0.04057211	-0.03878835	8.018652	8.058332	7.978971	3037.080	3160.015	2918.927
6	0.0008918805	0.04057211	-0.03878835	8.019544	8.059224	7.979863	3039.790	3162.834	2921.532

Prognozy logarytmicznych stóp zwrotu

n-ta kolejna sesja	Wartość	95% przedział ufności - dolna wartość	95% przedział ufności - górna wartość
1	0,00202	-0,03756	0,04162
2	0,00089	-0,03879	0,04057
3	0,00089	-0,03879	0,04057
4	0,00089	-0,03879	0,04057
5	0,00089	-0,03879	0,04057

Prognozy logarytmów cen

n-ta kolejna sesja	Wartość	95% przedział ufności - dolna wartość	95% przedział ufności - górna wartość
1	8,01598	7,97639	8,05557
2	8,01687	7,97719	8,05655
3	8,01776	7,97808	8,05744
4	8,01865	7,97897	8,05833
5	8,01954	7,97786	8,05922

Prognozy cen

n-ta kolejna sesja	Wartość	95% przedział ufności - dolna wartość	95% przedział ufności - górna wartość
1	3028,964	2911,387	3151,29
2	3031,667	2913,725	3154,383
3	3034,372	2916,325	3157,198
4	3037,080	2918,927	3160,015
5	3039,79	2921,532	3162,834

Porównanie prognoz z wartościami rzeczywistymi

Data	Prognozy	Wartości rzeczywiste	Błąd prognoz
3/6/2022	3039,79	2720,29	0,117450713
3/5/2022	3037,08	2749,06	0,104770358
3/4/2022	3034,372	2912,82	0,041730007
3/3/2022	3031,667	2957,97	0,024914722
3/2/2022	3028,964	3041,05	0,003974285
3/1/2022	3022,84	3022,84	0

Na podstawie rezultatów przedstawionych w powyższej tabeli można zaobserwować, że błąd między prognozą, a wartością rzeczywistą w przeciągu 3 dni od rozpoczętej obserwacji wynosi mniej niż 5%, jednak z każdym kolejnym dniem zwiększa się i po 5 dniach wynosi już prawie 12%. Wynika to ze znaczącego spadku wyceny akcji spółki, którego prognoza nie przewidziała. Spadek ten mógł być spowodowany rozpoczęciem agresji rosyjskiej na Ukrainę, co zakwalifikować można jako niespodziewane wydarzenie jednorazowe.