

**WYDZIAŁ ZARZĄDZANIA**

Ekonometria Finansowa i Dynamiczna

Projekt zaliczeniowy

*Opracowanie dla spółek 11 bit Studios oraz CD Projekt*

Autorzy: *Dawid Ciochoń i Jakub Górowski*

Kierunek studiów: Informatyka i ekonometria, studia II stopnia

Kraków, 2021

Spis treści

[Wstęp 3](#_Toc69568968)

[Cel projektu 3](#_Toc69568969)

[Metodyka badań 3](#_Toc69568970)

[Opis i wizualizacja danych 4](#_Toc69568971)

[Analiza stacjonarności 6](#_Toc69568972)

[Logarytmy oraz stopy zwrotu 9](#_Toc69568973)

[Analiza autokorelacji 11](#_Toc69568974)

[Wybór modelu dla danych 11 bit Studios 13](#_Toc69568975)

[Auto arima 13](#_Toc69568976)

[Wybór i weryfikacja istotności parametrów modelu SARIMA 14](#_Toc69568977)

[Normalność rozkładu reszt oraz ich stacjonarność 15](#_Toc69568978)

[Występowanie efektu ARCH/GARCH 17](#_Toc69568979)

[Model GARCH 19](#_Toc69568980)

[Wyznaczenie parametrów modelu GARCH 20](#_Toc69568981)

[Prognoza na zbiorze uczącym (Zadanie 9) 41](#_Toc69568982)

[Zadanie 9 (GARCH) 42](#_Toc69568983)

[Prognoza na zbiorze testowym (Zadanie 11) 44](#_Toc69568984)

[Prognozowanie - SARIMA 45](#_Toc69568985)

[Prognozowanie - SARIMA-GARCH 47](#_Toc69568986)

[Budowa modelu VAR 48](#_Toc69568987)

[Wybór opóźnienia i weryfikacja założeń 49](#_Toc69568988)

[Przyczynowość Grangera 51](#_Toc69568989)

[Test Grangera 51](#_Toc69568990)

[Kointegracja 52](#_Toc69568991)

[Test Johansena 53](#_Toc69568992)

[Procedura Engle’a i Granger’a 54](#_Toc69568993)

# Wstęp

Choć Polska nie jest kolebką gamedevu, jak również swoich oddziałów nie posiadają w naszym kraju giganci branży, to jednak wiele milionów graczy z całego świata ma okazję spędzać czas w produkcjach autorstwa polskich przedsiębiorstw. Najbardziej znanym jest oczywiście CD Projekt, którego uznać można za lidera polskiego gamedevu. Jej początki sięgają dystrybucji i wydawania polskich edycji zagranicznych produkcji. Obecnie w skład grupy kapitałowej CD Projekt, oprócz podmiotów zajmujących się wymienionymi czynnościami wchodzi również GOG (sklep internetowy zajmujący się cyfrową dystrybucją gier) oraz CD Projekt RED - studio tworzące gry. Właśnie za sprawą tego ostatniego przedsiębiorstwo zdobyło międzynarodową rozpoznawalność. Wiedźmińska trylogia (a w szczególności trzecia część) została ciepło przyjęta przez recenzentów i graczy na całym świecie i sprawiła że studio to było uważane za jednego z bardziej godnych zaufania i docenianych twórców gier na świecie. Pod koniec ubiegłego roku miała jednak premierę kolejna gra Redów - Cyberpunk 2077 - który był produktem niedopracowanym i pozostawiającym wiele do życzenia. Z tego powodu analiza notowań tej spółki będzie niezmiernie ciekawa - zauważyć będzie można wpływ nieudanej premiery na ceny akcji na warszawskiej giełdzie. Drugą ze spółek która została wybrana do projektu jest 11 bit studios. Jest to przedsiębiorstwo mniejsze, lecz również mające na swoim koncie spory sukces - grę This War of Mine ukazującą okrócieństwa wojny z perspektywy grupy cywili. Została doceniona przez recenzentów i graczy, a także stała się pierwszą grą w Polsce, która trafiła do kanonu lektur (powodując tym samym liczne dyskusje i zmieniając nastawienie do gier komputerowych).

# Cel projektu

We wstępie przedstawiono wybrane do projektu przedsiębiorstwa i zarysowano tematykę notowań giełdowych. Precyzując, w projekcie poruszone zostaną zagadnienia:

* analizy notowań i stóp zwrotu przedsiębiorstw
* dopasowanie modelu szeregu czasowego wraz z weryfikacją
* badanie kointegracji i przyczynowości

Projekt ten nie tylko umożliwi matematyczny zapis szeregu czasowego opartego o dane historyczne dotyczące polskich przedsiębiorstw z sektora gamedevu, lecz również na sprawdzenie, czy spółki w tym sektorze mają na siebie wpływ (z punktu widzenia inwestorów)

# Metodyka badań

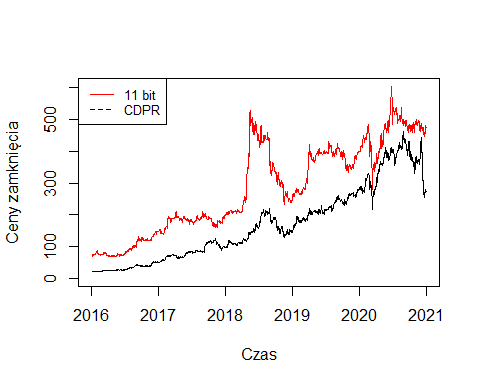
Projekt opiera się na analizie szeregów czasowych w postaci cen zamknięcia akcji dwóch spółek giełdowych, dlatego też wykorzystane zostaną techniki pozwalające na uchwycenie zależności między rozważanymi szeregami oraz budowę modelu procesu w oparciu o obserwowane zmiany w czasie cen zamknięcia, a także weryfikację wybranych modeli pod względem dopasowania. Bardzo waznym etapem analizy szeregów czasowych jest analiza stacjonarności. Jest to zasadnicza kwestia, gdyż wykorzystywane w dalszej części modele opierają się na konkretnej postacji szeregów czasowych. Jednym z takich modeli jest model (S)AR(I)MA. To właśnie na jego podstawie dobiera się odpowiedni model ARCH/GARCH wykorzystywany do dokonywania prognoz. Poza wspomnianymi modelami zostanie zbudowany również model VAR, a nastepnie zweryfikowana zostanie wzajemna zależność przyczynowa pomiędzy badanymi spółkami.

# Opis i wizualizacja danych

Dane do projektu zostały pobrane ze strony Giełdy Papierów Wartościowych jako dzienne ceny zamknięcia akcji dwóch spółek z branży gier - lidera rynku CD Project oraz jednej z największych firm na polskim rynku, czyli 11 bit Studios. Pochodzą one z okresu od 4.01.2016 do 14.01.2021. Liczność zbioru wynosi 1314 obserwacji. W celu zbadania sezonowości pierwotnie pobrane dane uzupełniono o obserwacje z pominiętych (ze względu na święta, przez co giełda była nieczynna) dni roboczych uzupełniając braki obserwacją z dnia poprzedniego. Tak przygotowane dane podzielono na zbiór treningowy liczący 1305 obserwacji oraz zbiór testowy z 9 obserwacjami. Zbiór uczący wykorzystany zostanie do budowy modelu oraz wykazania zależności przyczynowych pomiędzy analizowanymi szeregami, natomiast na podstawie pozostałej części danych wykonana zostanie prognoza i zbadane zostanie dopasowanie modelu.

Na wstępie warto zobrazować sobie na wykresie jak kształtowały się ceny zamknięcia akcji obu firm w zależności od czasu.

## Data Zamkniecie\_11bit Zamkniecie\_CDPR   
## Min. :2016-01-04 00:00:00 Min. : 68.0 Min. : 20.21   
## 1st Qu.:2017-04-06 12:00:00 1st Qu.:176.0 1st Qu.: 69.38   
## Median :2018-07-11 00:00:00 Median :271.0 Median :146.50   
## Mean :2018-07-10 23:58:54 Mean :285.5 Mean :168.84   
## 3rd Qu.:2019-10-14 12:00:00 3rd Qu.:409.0 3rd Qu.:248.00   
## Max. :2021-01-14 00:00:00 Max. :604.0 Max. :460.80



W rozpatrywanym okresie czyli od początku 2016 roku aż do końca 2020 roku ceny zamknięcia akcji 11 bit Studios były zawsze wyższe niż analogiczne ceny twórców Cyberpunka 2077. Najpewniej wpływ na taką sytuację ma liczba wyemitowanych akcji, która dla CD Projectu jest zdecydowanie większa. Nawiązując do głównego celu projektu, czyli badania zależności przyczynowych pomiędzy omawianymi przedsiębiorstwami na podstawie wykresu można stwierdzić, iż zależność występuje, gdyż ceny podążają w tym samym kierunku praktycznie przez cały okres.  
Wykres ten świetnie przedstawia reakcję giełdy na wydarzenia związane z sytuacją wskazanych przedsiębiorstw oraz ogólnoświatową. Dla 11bit najwyższą wartość akcje osiągnęły w okresie w którym miała miejsce konferencja inwestorska, gdzie przedstawiono m. in. plany na przyszłość. Spowodowało to, że pod koniec czerwca wartość cen zamknięcia przekroczyła barierę 600 PLN za akcję. Drugi ze szczytowych okresów związany był z premierą gry Frostpunk (24.04.2018). Już wtedy 11bit cieszyło się dobrą reputacją po niezwykle ciepłym przyjęciu This War of Mine (która zdobyła średnią 83% w serwisie Metacritic, gdzie brane pod uwagę są jedynie recenzje z uznanych portali branżowych). Premiera Frostpunk (gry ponownie zmuszającej gracza do trudnych wyborów) utwierdziła graczy w tym przekonaniu (tym razem 84% w serwisie Metacritic) i znacząco wpłynęła na ówczesne ceny akcji, podnosząc je z poziomu 200 PLN, aż do 500 PLN.

Premiera nowej gry nie zawsze ma jednak pozytywny wpływ na notowania giełdowe. Dobrym kontrprzykładem jest premiera Cyberpunk’a 2077. Po ogromnym sukcesie trzeciej części Wiedźmina gracze spodziewali się równie ambitnego i solidnie przygotowanego tytułu. Dodatkowo działania marketingowe jedynie utwierdzały ich w tym przekonaniu. Cyberpunk 2077 miał być grą rewolucyjną, prawdziwym opus magnum CD Projekt RED. Gra była trzykrotnie opóźniana, aby ostatecznie zadebiutować 10 grudnia 2020 roku. tuż przed tą datą wartość akcji CD Projektu osiągnęła drugą najwyższą wartość w historii (około 443 PLN), aby następnie spaść do poziomu 256 PLN w okresie przedświątecznym. Wpłynęły na to problemy ze stabilnością rozgrywki, ogromna ilość błędów oraz nieprzemyślane rozwiązania zastosowane w grze. Gra ta stała się wręcz precedensem - niektóre kanały jej dystrybucji zostały zamknięte (e względu na niezgodność towaru z umową), Metacritic zastosował oddzielne skale dla wersji na komputery oraz konsole (na niektórych z nich produkt był wręcz niegrywalny), a sam CD Projekt zobowiązał się do zwrotu pieniędzy dla osób które zakupiły produkcję przed premierą i były nią rozczarowane. Nad przedsiębiorstwem widniało również widmo pozwu zbiorowego, lecz kryzys ten został zażegnany. Wszystkie wymienione problemy przyczyniły się do ogromnego spatku cen akcji (oraz znacznego zwiększenia wolumenu obrotu) pod koniec 2020 roku. Lecz w rozpatrywanym okresie miały miejsce również wydarzenia sprzyjające wzrostowi cen akcji CD Projektu - była to np. premiera serialu Wiedźmin produkcji Netflixa, która sprawiła że kolejne osoby sięgały po komputerową adaptację przygód Białego Wilka (nawet po 4,5 roku po premierze). Spowodowana tym podwyższona sprzedaż pod koniec roku 2019 r. prawdopodobnie wpłynęła na postępujący wzrost cen akcji na początku roku 2020. Obydwa przedsiębiorstwa zostały natomiast dotknięte nerwowymi reakcjami inwestorów na rozwój pandemii koronawirusa COVID-19, które w Europie miały miejsce na początku marca 2020. Był to moment gdy poziom cen tych przedsiębiorstw był do siebie najbardziej zbliżony.

W celu lepszego poznania danych poniżej zamieszczono podstawowe statystyki opisowe:

## vars n mean sd median trimmed mad min max  
## Zamkniecie\_11bit 1 1305 284.07 142.43 268.00 283.00 186.07 68.00 604.0  
## Zamkniecie\_CDPR 2 1305 168.17 117.59 144.91 157.28 120.76 20.21 460.8  
## range skew kurtosis se  
## Zamkniecie\_11bit 536.00 0.08 -1.34 3.94  
## Zamkniecie\_CDPR 440.59 0.62 -0.66 3.26

Analizując statystyki opisowe obu szeregów czasowych można zauważyć, że charakteryzowały się one trendem wzrostowym, jednak w przypadku cen zamknięcia akcji 11 bit Studio zarówno wzrosty jak i spadki były większe o czym świadczy większa wariancja wahań niż dla cen zamknięcia akcji CD Project. Również średnia cena zamknięcia akcji jest zdecydowanie większa dla analizowanej spółki w porównaniu z liderem sektora gier. Ceny minimalne oraz maksymalne zdają się potwierdzać poprzednie stwierdzenie dotyczące liczby akcji, która może mieć główny wpływ na kształtowanie się cen akcji.

# Analiza stacjonarności

Stacjonarność może występować jednej z trzech form: jako ścisła stacjonarność, słaba stacjonarność oraz niestacjonarność. Chan[1] definiuje zbiór elementów {} jako ściśle stacjonarny, jeżeli dla każdego n, dla każdego i dla każdego h:

gdzie operator  oznacza równość rozkładów.

Jednak taka definicja jest bardzo restrykcyjna, ponieważ zakłada stałość w czasie zarówno wariancji, jak i średniej, dlatego też warunek ten, z reguły, jest trudny do zweryfikowania. W tym miejscu warto zwrócić uwagę na drugi rodzaj stacjonarności, czyli tzw. słabą stacjonarność, która jest już wystarczająca do dalszej analizy szeregów czasowych. Opiera się ona na dwóch założeniach: -

, czyli stałej wartości średniej oraz

które oznacza, że kowariancja zależy tylko i wyłącznie od opóźnienia w czasie.

Ostatnim rodzajem stacjonarności jest niestacjonarność, która dotyczy procesu stochastycznego, w którym co najmniej jedna z wartości takich jak średnia, wariancja czy funkcja autokorelacji ulega zmianie wraz z upływem czasu. W takich przypadkach wymagane jest podejście, które pozwoli przekształcić szereg niestacjonarny w stacjonarny. Do zbadania szeregów czasowych pod kątem stacjonarności użyte zostaną dwa testy – rozszerzony test Dickey’a-Fullera (test ADF) oraz test Kwiatkowskiego-Phillipsa-Schmidta-Shina (test KPSS).

Test KPSS różni się od tradycyjnych testów weryfikujących występowanie pierwiastka jednostkowego, ponieważ testowana jest tu niestacjonarność przeciwko hipotezie zerowej stanowiącej o stacjonarności szeregu. Rozważany jest szereg czasowy postaci [2]:

gdzie to ciąg niezależnych zmiennych losowych o rozkładzie normalnym , natomiast szeregjest błądzeniem losowym w którym również jest ciągiem niezależnych zmiennych losowych o rozkładzie normalnym . Hipotezy testu KPSS można zapisać następująco:

H\_0:   
H\_1:

Parametry szeregu są szacowane metodą najmniejszych kwadratów, a następnie po ustaleniu rzędu opóźnienia k obliczana jest suma reszt oraz wartość estymatora wariancji długookresowej reszt. Po wykonaniu wspomnianych czynności można przejść do obliczenia statystyki testowej zgodnej ze wzorem:

, gdzie:

– suma reszt szeregu ,  
 – estymator wariancji długookresowej reszt,  
N – liczba obserwacji,  
k – rząd opóźnienia.

Wartość krytyczna odczytywana jest z odpowiednich tablic. Gdy statystyka obliczona w teście jest większa niż ta odczytana z tablic, hipoteza zerowa jest odrzucana na poziomie istotności α, a szereg czasowy uznaje się za niestacjonarny. W przeciwnej sytuacji nie ma podstaw do odrzucenia hipotezy zerowej.

Drugim z wykorzystanych testów jest rozszerzony test Dickey’a-Fullera. Test ADF wykorzystuje pojęcie pierwiastka jednostkowego i jego związku z brakiem stacjonarności. Obecność pierwiastka jednostkowego wpływa na rekację szregu czasowego na wystąpienie *impulsu* (nagłego zaburzenia wartości jednego z jego wyrazów) - gdy występuje on to rekacja taka będzie stałą i nigdy nie wygaśnie. Natomaist w przypadku jego braku, reakcja stopniowo będzie podlegać wygaszeniu. Opisywane zjawisko dotyczy modelu:

Stał on się podstawą do opracowanie testu Dickeya-Fullera, którego rozszerzoną wersję stanowi test ADF. Wymaga on oszacowania następującego równania:

Taka postać równania umożliwia włączenie procesów ARMA dla błędów losowych [3] . W teście ADF hipoteza zerowa oznacza brak stacjonarności (szereg jest charakteryzowany przez process pierwiastka jednostkowego), natomiast hipoteza alternatywna - że szereg jest stacjonarny. Zapisać można je w postaci:

Statystyka testu ma postać:

Postać ta zbliżona jest do statystyki testu istotności t-Studenta (również dzielimy wartość parametru przez jego błąd), natomiast rozkład tej statystyki w znaczący sposób różni się od rozkładu t, przez co należy stosować specjalne tablice dotyczące testu ADF [4] . Wadą testu ADF jest mała moc, natomiast można stosować je razem z testami posiadajacymi odwrotne hipotezy (np. test KPSS) w ramach *analizy potwierdzającej*, aby uzyskać dodatkowe potwierdzenie prawdziwości otrzymywanych wyników. Natomiast należy odnotować, że badania dotyczące tego podejścia podważają jego skuteczność [5].

# Logarytmy oraz stopy zwrotu

##   
## Augmented Dickey-Fuller Test  
##   
## data: train[, 4]  
## Dickey-Fuller = -2.1599, Lag order = 10, p-value = 0.5106  
## alternative hypothesis: stationary

##   
## Augmented Dickey-Fuller Test  
##   
## data: train[, 5]  
## Dickey-Fuller = -1.5486, Lag order = 10, p-value = 0.7694  
## alternative hypothesis: stationary

## Warning in tseries::kpss.test(train[, 4]): p-value smaller than printed p-value

##   
## KPSS Test for Level Stationarity  
##   
## data: train[, 4]  
## KPSS Level = 14.093, Truncation lag parameter = 7, p-value = 0.01

## Warning in tseries::kpss.test(train[, 5]): p-value smaller than printed p-value

##   
## KPSS Test for Level Stationarity  
##   
## data: train[, 5]  
## KPSS Level = 15.457, Truncation lag parameter = 7, p-value = 0.01

Po przekształeceniu danych do postaci logarytmów oraz zastosowaniu wyżej wymienionych testów można wnioskować, iż oba szeregi czasowe są niestacjonarne, gdyż z testu ADF p-value jest zdecydowanie większe niż przyjęty poziom istotności wynoszący 5%, natomiast p-value z testu KPSS jest mniejsze niż 5%, więc wskazuje to na odrzucenie hipotezy zerowej stanowiącej o stacjonarności.

W takiej sytuacji literatura wskazuje na kilka możliwych podejść do przekształcenia danych do postaci stacjonarnej. Jednym z nich jest różnicowanie aż do momentu, gdy szeregi będą już stacjonarne. Innym popularnym sposobem radzenia sobie z niestacjonarnością jest transformacja Boxa-Coxa. W tym przypadku podjęto decyzję o zastosowaniu różnicowania.

## Warning in tseries::adf.test(dane6, alternative = "stationary"): p-value smaller  
## than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: dane6  
## Dickey-Fuller = -10.03, Lag order = 10, p-value = 0.01  
## alternative hypothesis: stationary

## Warning in tseries::adf.test(dane7, alternative = "stationary"): p-value smaller  
## than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: dane7  
## Dickey-Fuller = -11.208, Lag order = 10, p-value = 0.01  
## alternative hypothesis: stationary

## Warning in tseries::kpss.test(dane6): p-value greater than printed p-value

##   
## KPSS Test for Level Stationarity  
##   
## data: dane6  
## KPSS Level = 0.14963, Truncation lag parameter = 7, p-value = 0.1

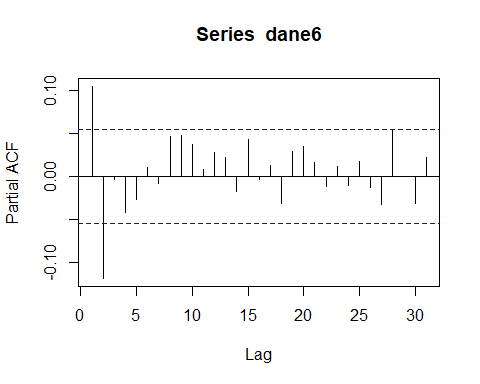
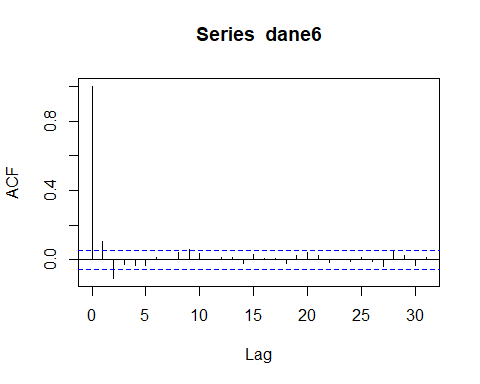
## Warning in tseries::kpss.test(dane7): p-value greater than printed p-value

##   
## KPSS Test for Level Stationarity  
##   
## data: dane7  
## KPSS Level = 0.30553, Truncation lag parameter = 7, p-value = 0.1

Zastosowanie pierwszych różnic pomogło otrzymać stacjonarne szeregi, ponieważ wartości p-value z testu ADF wskazują na odrzucenie hipotezy zerowej mówiącej o niestacjonarności, a w przypadku testu KPSS wartości p-value były większe niż przyjęty poziom istotności (5%), więc nie było podstaw do odrzucenia hipotezy zerowej. Tak przygotowane dane będą mogły posłużyć do budowy modelu.

# Analiza autokorelacji

Analiza autokorelacji jest jednym z głównych kroków prowadzących do budowy odpowiedniego modelu, ponieważ dostarcza informacje na temat parametrów przyszłego modelu. Analizując funkcję autokorelacji można wyciągnąć wnioski co do części MA, czyli średniej ruchomej. Analogiczne dane dostarcza funkcja PACF dla części AR modelu. Pamiętając, iż analizowane dane okazały się stacjonarne dopiero po pierwszym różnicowaniu, można założyć, że jednym z rozważanych modeli będzie model ARIMA lub, gdy wykryta zostanie sezonowość, model SARIMA.



# Wybór modelu dla danych 11 bit Studios

Z wykresu funkcji ACF można odczytać, że istotne są współczynniki autokorelacji dla opóźnień równych 0, 1, 2 oraz dla opóźnienia równego 9. Stąd też parametr *q* przyjmie najpewniej jedną z wymienionych wartości. Przechodząc do parametru *p*, jego wartość wyniesie 1 lub 2, gdyż na podstawie wykresu tylko te współczynniki autokorelacji cząstkowej są istotne.

## Auto arima

##   
## Fitting models using approximations to speed things up...  
##   
## ARIMA(2,1,2) with drift : -5711.825  
## ARIMA(0,1,0) with drift : -5684.029  
## ARIMA(1,1,0) with drift : -5698.856  
## ARIMA(0,1,1) with drift : -5700.292  
## ARIMA(0,1,0) : -5682.262  
## ARIMA(1,1,2) with drift : -5714.255  
## ARIMA(0,1,2) with drift : -5712.593  
## ARIMA(1,1,1) with drift : -5708.553  
## ARIMA(1,1,3) with drift : -5712.349  
## ARIMA(0,1,3) with drift : -5711.775  
## ARIMA(2,1,1) with drift : -5712.476  
## ARIMA(2,1,3) with drift : -5709.813  
## ARIMA(1,1,2) : -5712.416  
##   
## Now re-fitting the best model(s) without approximations...  
##   
## ARIMA(1,1,2) with drift : -5719.598  
##   
## Best model: ARIMA(1,1,2) with drift

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.3257902 0.2007121 1.6232 0.10455   
## ma1 -0.2097506 0.1991021 -1.0535 0.29212   
## ma2 -0.1487518 0.0316524 -4.6995 2.607e-06 \*\*\*  
## drift 0.0014628 0.0007093 2.0623 0.03918 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] -5719.644

## Model ARIMA

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 0.116605 0.027771 4.1988 2.683e-05 \*\*\*  
## ma2 -0.104193 0.028061 -3.7130 0.0002048 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] -5718.085

##   
## Shapiro-Wilk normality test  
##   
## data: funkcja2$residuals  
## W = 0.91941, p-value < 2.2e-16

Funkcja auto.arima() dostępna w języku R służy do wybrania najlepszego modelu analizując kryteria informacyjne: AIC oraz BIC (najlepszy model to ten z najmniejszymi wartościami kryteriów). Na podstawie powyższych wyników wspomnianej funkcji model najlepszy to model ARIMA(1,1,2). Jednak nie została zweryfikowana potencjalna sezonowość, dlatego też zostanie dodana część odpowiadająca za możliwą sezonowość, a następnie wybrane modele zostaną poddane testowi istotności parametrów. Jeżeli wszystkie parametry będą istotne statystycznie ostateczny model zostanie wybrany na podstawie kryteriów informacyjnych. W modelu ARIMA(1,1,2) nie wszystkie zmienne są istotne, przez co należy wybrać inny model, w którym każda zmienna pozytywnie przejdzie test istotności. Rozważany będzie model ARIMA(0,1,2) - następny pod względem kryteriów informacyjnych.

## Wybór i weryfikacja istotności parametrów modelu SARIMA

## Wybór odpowiedniej postaci Sarimy

## AIC

## Warning in log(s2): wyprodukowano wartości NaN

## Warning in arima(train[, 4], order = c(0, 1, 2), seasonal = list(order = c(p, :  
## possible convergence problem: optim gave code = 1

## Q=0 Q=1 Q=2 Q=3 Q=4  
## P=0 -4757.267 -5667.774 -5666.944 -5667.136 -5667.487  
## P=1 -5190.744 -5667.039 -5665.701 -5668.043 -5667.337  
## P=2 -5357.728 -5667.031 -5667.553 -5666.303 -5664.216  
## P=3 -5444.973 -5667.336 -5666.530 -5663.753 -5684.906  
## P=4 -5487.957 -5667.353 -5665.450 -5662.987 -5660.978

## BIC

## Warning in log(s2): wyprodukowano wartości NaN  
  
## Warning in log(s2): possible convergence problem: optim gave code = 1

## Q=0 Q=1 Q=2 Q=3 Q=4  
## P=0 -4741.759 -5647.096 -5641.097 -5636.120 -5631.302  
## P=1 -5170.067 -5641.193 -5634.684 -5631.857 -5625.982  
## P=2 -5331.881 -5636.015 -5631.368 -5624.948 -5617.692  
## P=3 -5413.957 -5631.150 -5625.175 -5617.229 -5633.212  
## P=4 -5451.772 -5625.998 -5618.926 -5611.294 -5604.115

## Model SARIMA

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 0.1135645 0.0278427 4.0788 4.527e-05 \*\*\*  
## ma2 -0.1067125 0.0282048 -3.7835 0.0001546 \*\*\*  
## sma1 -0.9999996 0.0079639 -125.5672 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## [1] -5667.774

##   
## Shapiro-Wilk normality test  
##   
## data: funkcja3$residuals  
## W = 0.91997, p-value < 2.2e-16

Spośród modeli SARIMA najlepszym był model (0,1,2)(3,1,4)[5] natomiast R napotkał problemy z jego całkowitym opisem (być może przez przeparametryzowanie). Z tego powodu postanowiono wybrać model SARIMA(0,1,2)(0,1,1)[5], który również posiadał odpowiednie wartości kryteriów informacyjnych i co ważniejsze - wszystkie jego parametry były istotne.

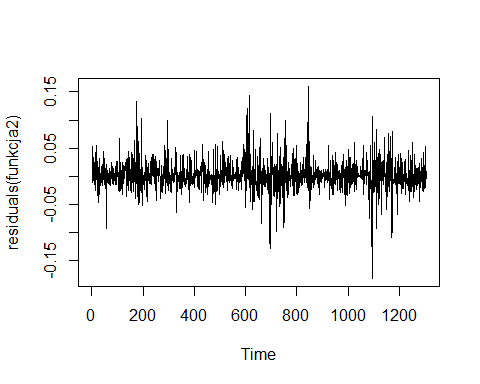
## Normalność rozkładu reszt oraz ich stacjonarność

Wybranymi modelami są zatem ARIMA(0,1,2) oraz SARIMA(0,1,2)(0,1,1)[5]. Choć przeprowadzone powyżej testy normalności Shapiro-Wilka wykazują brak normalności rozkładu reszt, to jednak powołując się na Centralne Twierdzenie Graniczne przyjąć można, że posiadają one asymptotyczny rozkład normalny. Są też stacjonarne, co potwierdza poniższy wykres oraz wynik testu KPSS. Test istotności rownież nie wykazał, by którakolwiek ze zmiennych była nieistotna, co oznacza, że do dalszej analizy zostały wybrane wspomniane dwa modele.

## ARIMA

## Warning in tseries::kpss.test(funkcja2$residuals): p-value greater than printed  
## p-value

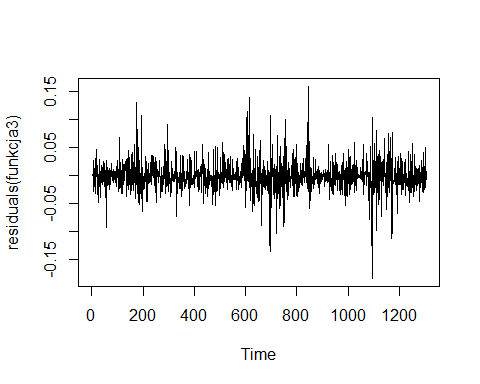
##   
## KPSS Test for Level Stationarity  
##   
## data: funkcja2$residuals  
## KPSS Level = 0.15425, Truncation lag parameter = 7, p-value = 0.1



## SARIMA

## Warning in tseries::kpss.test(funkcja3$residuals): p-value greater than printed  
## p-value

##   
## KPSS Test for Level Stationarity  
##   
## data: funkcja3$residuals  
## KPSS Level = 0.081718, Truncation lag parameter = 7, p-value = 0.1



W dalszej kolejności postanowiono sprawdzić występowanie efektu ARCH.

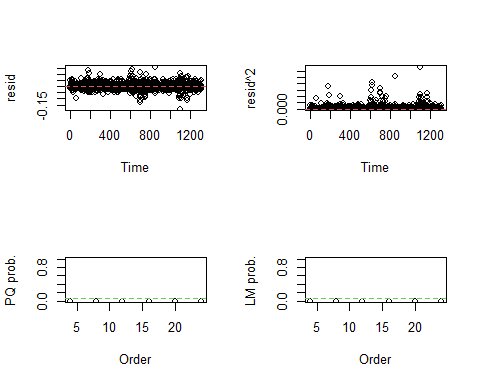
# Występowanie efektu ARCH/GARCH

Badania Engle’a [6] wykazały, że dla pewnych danych może zachodzić zjawisko heteroskedastyczności w wariancji składnika losowego. Najczęściej ma to miejsce w przypadku analizy danych dotyczących zmienności inflacji, stóp zwrotu z inwestycji czy rynków walutowych [7]. Engle zaproponował więc test pozwalający wykryć niejednorodność w wariancji składnika losowego. Z uwagi na podatność danych finansowych na możliwość występowania efektu ARCH test ten zostanie przeprowadzony na resztach otrzymanych z utworzonego modelu ARIMA oraz SARIMA. Hipotezy testu ARCH-LM:

H0: brak efektu ARCH  
H1: występuje efekt ARCH

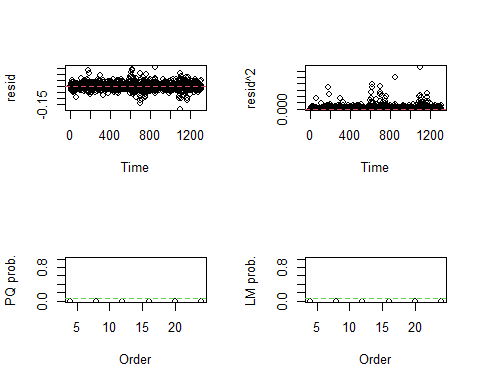
## ARIMA

## ARCH heteroscedasticity test for residuals   
## alternative: heteroscedastic   
##   
## Portmanteau-Q test:   
## order PQ p.value  
## [1,] 4 164 0  
## [2,] 8 179 0  
## [3,] 12 220 0  
## [4,] 16 225 0  
## [5,] 20 232 0  
## [6,] 24 234 0  
## Lagrange-Multiplier test:   
## order LM p.value  
## [1,] 4 1153 0  
## [2,] 8 560 0  
## [3,] 12 339 0  
## [4,] 16 250 0  
## [5,] 20 194 0  
## [6,] 24 160 0



## SARIMA

## ARCH heteroscedasticity test for residuals   
## alternative: heteroscedastic   
##   
## Portmanteau-Q test:   
## order PQ p.value  
## [1,] 4 161 0  
## [2,] 8 174 0  
## [3,] 12 208 0  
## [4,] 16 213 0  
## [5,] 20 221 0  
## [6,] 24 223 0  
## Lagrange-Multiplier test:   
## order LM p.value  
## [1,] 4 1167 0  
## [2,] 8 570 0  
## [3,] 12 346 0  
## [4,] 16 256 0  
## [5,] 20 197 0  
## [6,] 24 162 0



Jak widać efekt ARCH występuje dla opóźnień każdego opóźnienia w obu modelach. Z uwagi na fakt, iż efekt ARCH jest obecny w modelach, warto spróbować zbudować model, który dopuszcza występowanie tego efektu, czyli model GARCH.

## Model GARCH

GARCH jest to uogólniona forma modelu ARCH stworzona przez Bollersleva. Model ten jest lepszy od ARCH, ponieważ:

* lepiej dostosowuje się on do opisu danych w których na końcach rozkładów występuje znaczna gęstość prawdopodobieństwa (grube ogony)
* nie powoduje otrzymania ujemnej wariancji warunkowej (co mógł powodować ARCH)

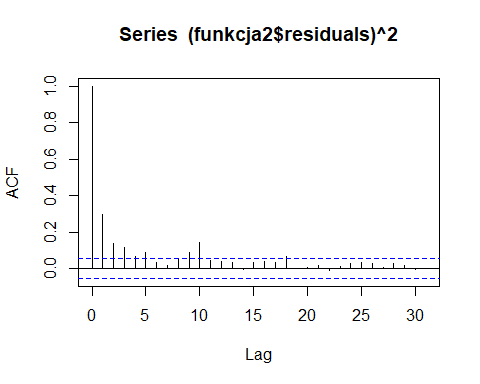
Model ten wyrazić można równaniami:

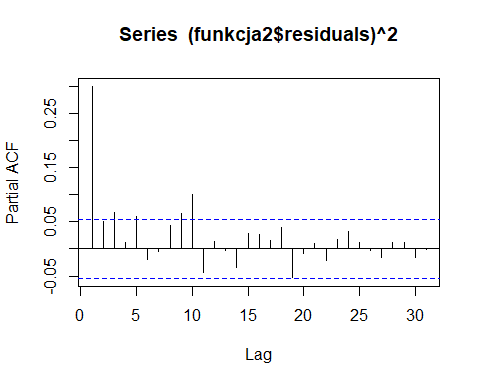
Gdzie to Gaussowski biały szum, a >0,≥0,≥0. W procesie przedstawionym jako GARCH(p,q) zmienność teoretycznie jest kształtowana przez nieskończoną liczbę kwadratów opóźnień zmiennej y. Warto również zaznaczyć, że gdy w GARCH(p,q) p=0, to jest to model ARCH(q).

### Wyznaczenie parametrów modelu GARCH

Aby wyznaczyć odpowiednie wartości parametrów p i q w modelu GARCH posłużyć można się wykresami ACF oraz PACF dla kwadratów reszt z modelu ARIMA.

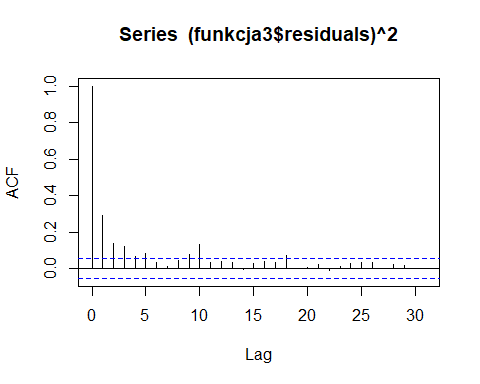
## ARIMA

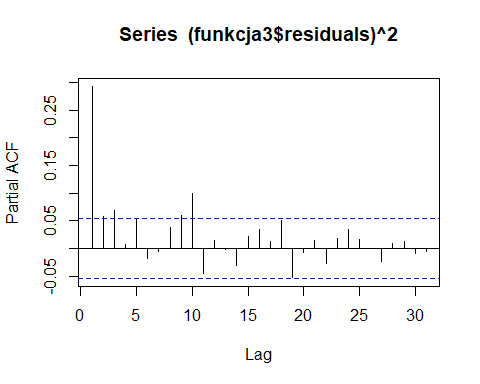




Wykres funkcji PACF dla modelu ARIMA(0,1,2) wskazuje, iż opóźnienia równe 1, 3 oraz 5 są istotne, czyli parametr P w modelu GARCH powinien przyjąć jedną z tych wartości. Natomiast z wykresu fukcji ACF można odczytać, że opóźnienia od 1 do 5 są istotne.

## SARIMA





Dla modelu SARIMA jest podobnie - możliwe wartości parametru Q zawierają się pomiędzy 1 a 5, a dla parametru P rozważane będą wartości 1, 2, 3 oraz 5.

#### ARIMA eGARCH

## Kryterium AIC

## Q=1 Q=2 Q=3  
## P=1 -4.569815 -4.578693 -4.578886  
## P=3 -4.579655 -4.578766 -4.578484  
## P=5 -4.580869 -4.579821 -4.590086

## Kryterium BIC

## Q=1 Q=2 Q=3  
## P=1 -4.542045 -4.546956 -4.543181  
## P=3 -4.543950 -4.539095 -4.534845  
## P=5 -4.537230 -4.532215 -4.538512

##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : eGARCH(3,3)  
## Mean Model : ARFIMA(1,0,1)  
## Distribution : norm   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.000645 0.000496 1.30065 0.193379  
## ar1 -0.514627 0.044359 -11.60144 0.000000  
## ma1 0.572658 0.042237 13.55824 0.000000  
## omega -1.223667 0.345734 -3.53933 0.000401  
## alpha1 0.008541 0.031463 0.27146 0.786041  
## alpha2 -0.114571 0.032510 -3.52416 0.000425  
## alpha3 0.079183 0.036339 2.17904 0.029328  
## beta1 0.999999 0.063713 15.69528 0.000000  
## beta2 -0.936163 0.071752 -13.04717 0.000000  
## beta3 0.765276 0.047771 16.01962 0.000000  
## gamma1 0.407991 0.045116 9.04309 0.000000  
## gamma2 -0.149977 0.053939 -2.78050 0.005428  
## gamma3 0.307164 0.051747 5.93585 0.000000  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.000645 0.000474 1.36272 0.172971  
## ar1 -0.514627 0.030089 -17.10372 0.000000  
## ma1 0.572658 0.025685 22.29568 0.000000  
## omega -1.223667 0.707664 -1.72916 0.083780  
## alpha1 0.008541 0.055082 0.15505 0.876779  
## alpha2 -0.114571 0.051574 -2.22147 0.026319  
## alpha3 0.079183 0.058281 1.35865 0.174258  
## beta1 0.999999 0.150235 6.65621 0.000000  
## beta2 -0.936163 0.134759 -6.94693 0.000000  
## beta3 0.765276 0.071868 10.64831 0.000000  
## gamma1 0.407991 0.073854 5.52431 0.000000  
## gamma2 -0.149977 0.088362 -1.69730 0.089640  
## gamma3 0.307164 0.099382 3.09074 0.001997  
##   
## LogLikelihood : 3005.736   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike -4.5901  
## Bayes -4.5385  
## Shibata -4.5903  
## Hannan-Quinn -4.5707  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 0.7555 0.3847  
## Lag[2\*(p+q)+(p+q)-1][5] 2.9822 0.4832  
## Lag[4\*(p+q)+(p+q)-1][9] 5.3767 0.3719  
## d.o.f=2  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 1.923 0.1655  
## Lag[2\*(p+q)+(p+q)-1][17] 6.210 0.7765  
## Lag[4\*(p+q)+(p+q)-1][29] 9.526 0.9032  
## d.o.f=6  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[7] 0.001091 0.500 2.000 0.9737  
## ARCH Lag[9] 1.063245 1.485 1.796 0.7519  
## ARCH Lag[11] 2.620808 2.440 1.677 0.6648  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: 1.8965  
## Individual Statistics:   
## mu 0.19289  
## ar1 0.17163  
## ma1 0.17076  
## omega 0.06876  
## alpha1 0.15976  
## alpha2 0.17162  
## alpha3 0.09821  
## beta1 0.06687  
## beta2 0.06142  
## beta3 0.06064  
## gamma1 0.04549  
## gamma2 0.05794  
## gamma3 0.10904  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Joint Statistic: 2.89 3.15 3.69  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 1.9087 0.05653 \*  
## Negative Sign Bias 1.9548 0.05082 \*  
## Positive Sign Bias 0.6413 0.52145   
## Joint Effect 5.1319 0.16239   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 171.9 1.195e-26  
## 2 30 197.3 6.012e-27  
## 3 40 223.5 9.697e-28  
## 4 50 216.9 5.496e-23  
##   
##   
## Elapsed time : 0.696188

##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : eGARCH(1,1)  
## Mean Model : ARFIMA(1,0,1)  
## Distribution : norm   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001370 0.000407 3.36814 0.000757  
## ar1 -0.464522 0.032549 -14.27143 0.000000  
## ma1 0.518910 0.029819 17.40219 0.000000  
## omega -0.547826 0.136687 -4.00790 0.000061  
## alpha1 -0.018711 0.020893 -0.89559 0.370474  
## beta1 0.922272 0.018632 49.49905 0.000000  
## gamma1 0.317341 0.037646 8.42952 0.000000  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001370 0.000297 4.62013 0.000004  
## ar1 -0.464522 0.012489 -37.19429 0.000000  
## ma1 0.518910 0.014976 34.65027 0.000000  
## omega -0.547826 0.227530 -2.40770 0.016053  
## alpha1 -0.018711 0.036717 -0.50962 0.610319  
## beta1 0.922272 0.031136 29.62084 0.000000  
## gamma1 0.317341 0.057318 5.53652 0.000000  
##   
## LogLikelihood : 2986.519   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike -4.5698  
## Bayes -4.5420  
## Shibata -4.5699  
## Hannan-Quinn -4.5594  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 1.461 0.22677  
## Lag[2\*(p+q)+(p+q)-1][5] 4.510 0.01606  
## Lag[4\*(p+q)+(p+q)-1][9] 7.485 0.08799  
## d.o.f=2  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 3.498 0.06145  
## Lag[2\*(p+q)+(p+q)-1][5] 4.986 0.15416  
## Lag[4\*(p+q)+(p+q)-1][9] 6.358 0.25915  
## d.o.f=2  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[3] 0.3675 0.500 2.000 0.5444  
## ARCH Lag[5] 0.4022 1.440 1.667 0.9120  
## ARCH Lag[7] 1.0630 2.315 1.543 0.9031  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: 0.766  
## Individual Statistics:   
## mu 0.11620  
## ar1 0.24273  
## ma1 0.24645  
## omega 0.06739  
## alpha1 0.05793  
## beta1 0.06437  
## gamma1 0.07604  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Joint Statistic: 1.69 1.9 2.35  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 1.3733 0.1699   
## Negative Sign Bias 1.5133 0.1304   
## Positive Sign Bias 0.3156 0.7523   
## Joint Effect 3.7931 0.2847   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 186.4 1.669e-29  
## 2 30 197.8 4.830e-27  
## 3 40 215.6 2.622e-26  
## 4 50 246.8 3.441e-28  
##   
##   
## Elapsed time : 0.219378

#### ARIMA sGARCH

## Kryterium AIC

## Q=1 Q=2 Q=3  
## P=1 -4.561598 -4.571730 -4.570507  
## P=2 -4.560867 -4.570197 -4.568973  
## P=3 -4.559081 -4.568455 -4.567440

## Kryterium BIC

## Q=1 Q=2 Q=3  
## P=1 -4.537795 -4.543960 -4.538770  
## P=2 -4.533097 -4.538459 -4.533269  
## P=3 -4.527344 -4.532750 -4.527768

##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : sGARCH(1,1)  
## Mean Model : ARFIMA(0,0,1)  
## Distribution : norm   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001553 0.000640 2.4259 0.015271  
## ma1 0.047882 0.034408 1.3916 0.164047  
## omega 0.000045 0.000011 4.0981 0.000042  
## alpha1 0.144071 0.024350 5.9166 0.000000  
## beta1 0.799949 0.029730 26.9070 0.000000  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001553 0.000572 2.7162 0.006603  
## ma1 0.047882 0.035299 1.3565 0.174949  
## omega 0.000045 0.000019 2.4496 0.014301  
## alpha1 0.144071 0.037551 3.8367 0.000125  
## beta1 0.799949 0.048122 16.6232 0.000000  
##   
## LogLikelihood : 2979.264   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike -4.5618  
## Bayes -4.5419  
## Shibata -4.5618  
## Hannan-Quinn -4.5543  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 1.838 0.175175  
## Lag[2\*(p+q)+(p+q)-1][2] 4.112 0.003358  
## Lag[4\*(p+q)+(p+q)-1][5] 6.632 0.026824  
## d.o.f=1  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 2.291 0.1301  
## Lag[2\*(p+q)+(p+q)-1][5] 3.322 0.3514  
## Lag[4\*(p+q)+(p+q)-1][9] 4.646 0.4833  
## d.o.f=2  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[3] 0.2956 0.500 2.000 0.5867  
## ARCH Lag[5] 0.4142 1.440 1.667 0.9086  
## ARCH Lag[7] 1.2345 2.315 1.543 0.8727  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: 0.7607  
## Individual Statistics:   
## mu 0.21216  
## ma1 0.28782  
## omega 0.04170  
## alpha1 0.07572  
## beta1 0.04466  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Joint Statistic: 1.28 1.47 1.88  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 1.3028 0.1929   
## Negative Sign Bias 1.6297 0.1034   
## Positive Sign Bias 0.2199 0.8260   
## Joint Effect 3.6833 0.2978   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 240.3 2.880e-40  
## 2 30 248.8 8.620e-37  
## 3 40 272.5 8.550e-37  
## 4 50 287.7 1.625e-35  
##   
##   
## Elapsed time : 0.153599

##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : sGARCH(2,1)  
## Mean Model : ARFIMA(1,0,1)  
## Distribution : norm   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001530 0.000649 2.3581 0.018371  
## ar1 -0.432805 0.269550 -1.6057 0.108349  
## ma1 0.489015 0.260115 1.8800 0.060109  
## omega 0.000045 0.000012 3.8053 0.000142  
## alpha1 0.142431 0.031616 4.5051 0.000007  
## alpha2 0.000000 0.043522 0.0000 1.000000  
## beta1 0.802296 0.037823 21.2118 0.000000  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001530 0.000650 2.3553 0.018507  
## ar1 -0.432805 0.197648 -2.1898 0.028540  
## ma1 0.489015 0.191052 2.5596 0.010479  
## omega 0.000045 0.000019 2.4147 0.015750  
## alpha1 0.142431 0.042619 3.3420 0.000832  
## alpha2 0.000000 0.071518 0.0000 1.000000  
## beta1 0.802296 0.064052 12.5258 0.000000  
##   
## LogLikelihood : 2980.685   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike -4.5609  
## Bayes -4.5331  
## Shibata -4.5609  
## Hannan-Quinn -4.5504  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 1.308 0.25275  
## Lag[2\*(p+q)+(p+q)-1][5] 4.370 0.02433  
## Lag[4\*(p+q)+(p+q)-1][9] 7.436 0.09150  
## d.o.f=2  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 2.209 0.1372  
## Lag[2\*(p+q)+(p+q)-1][8] 4.205 0.4722  
## Lag[4\*(p+q)+(p+q)-1][14] 6.376 0.5882  
## d.o.f=3  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[4] 0.1072 0.500 2.000 0.7434  
## ARCH Lag[6] 0.8928 1.461 1.711 0.7787  
## ARCH Lag[8] 2.0567 2.368 1.583 0.7301  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: 2.8391  
## Individual Statistics:   
## mu 0.20708  
## ar1 0.33435  
## ma1 0.33358  
## omega 0.04121  
## alpha1 0.07055  
## alpha2 0.17561  
## beta1 0.04330  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Joint Statistic: 1.69 1.9 2.35  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 1.4353 0.15143   
## Negative Sign Bias 1.7405 0.08201 \*  
## Positive Sign Bias 0.1602 0.87273   
## Joint Effect 4.1280 0.24797   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 200.7 2.478e-32  
## 2 30 195.8 1.136e-26  
## 3 40 229.1 9.587e-29  
## 4 50 245.6 5.660e-28  
##   
##   
## Elapsed time : 0.1935139

Dalsza analiza w postaci budowy modelu ARIMA-GARCH nie przyniosła oczekiwanego skutku, gdyż pomimo dobierania różnych wartości parametrów P i Q oraz różnych postaci modelu GARCH nie udało się zbudowac modelu, w którym wszystkie zmienne byłyby istotne. Postanowiono więc porzucić model ARMA-GARCH i skupić się na modelu SARIMA-GARCH, który umożiwiłby uchwycenie potencjalnej sezonowości.

#### SARIMA - Przygotowanie external reggressors

Pierwszym krokiem w kierunku budowy modelu SARIMA-GARCH było przygotowanie zmiennych zero-jedynkowych obrazujących poszczególne dni tygodnia:

* 1,0,0,0,0 - wtorek
* 0,1,0,0,0 - środa
* 0,0,1,0,0 - czwartek
* 0,0,0,1,0 - piątek
* 0,0,0,0,0 - poniedziałek

Zaczynają się one od wtorku, gdyż pierwsza obserwacja dotczyła poniedziałku (5.01.2016), natomaist po jednokrotnym różnicowaniu pierwszy rekord musiał zostać usunięty, stąd początkiem szeregu czasowego jest wtorek, 6 stycznia 2016. Tak przygotowana macierz zostanie przekazana jako argument external regressors dla funkcji ugarchspec(). Jest to konieczne, gdyż żaden z pakietów R nie umożliwia tworzenia modeli SARIMA-GARCH w formie zbliżonej do ARIMA-GARCH (poprzez podanie parametrów P, D i Q). Przekazanie macierzy dummies umożliwi zaznaczenie występującej w danych sezonowości[8]

Następnie, podobnie jak w przypadku modelu ARMA-GARCH, dopasowywano różne postaci modelu GARCH oraz różne opóźnienia (przy wyborze opóźnienia kierowano się głównie kryterium AIC, niekiedy również BIC). Należy pamiętać, że wyniki w wierszu P=4 nie powinny być traktowane jako miarodajne, gdyż wykres autokorelacji PACF wykazał, że nie powinno rozpatrywać się tego efektu dla P=4 (wiersz ten natomiast )

#### SARIMA - eGARCH

## Kryterium AIC

## Kryterium AIC

## Q=1 Q=2 Q=3 Q=4 Q=5  
## P=1 -4.573219 -4.582678 -4.583133 -4.582741 -4.582056  
## P=2 -4.584007 -4.583299 -4.583004 -4.582145 -4.580966  
## P=3 -4.584832 -4.583776 -4.591859 -4.598402 -4.596298  
## P=5 -4.585872 -4.584826 -4.618349 -4.619403 -4.620663

## Kryterium BIC

## Q=1 Q=2 Q=3 Q=4 Q=5  
## P=1 -4.529580 -4.535072 -4.531560 -4.527200 -4.522548  
## P=2 -4.532433 -4.527759 -4.523497 -4.518670 -4.513524  
## P=3 -4.525325 -4.520301 -4.524417 -4.526993 -4.520922  
## P=5 -4.510495 -4.505482 -4.535039 -4.532125 -4.529418

##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : eGARCH(5,5)  
## Mean Model : ARFIMA(0,0,2)  
## Distribution : norm   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.002201 0.000028 78.250 0  
## ma1 -0.021568 0.000102 -211.259 0  
## ma2 -0.098398 0.004013 -24.523 0  
## mxreg1 -0.002903 0.000028 -103.682 0  
## mxreg2 -0.004245 0.000172 -24.727 0  
## mxreg3 -0.000809 0.000065 -12.373 0  
## mxreg4 0.001118 0.000029 38.172 0  
## omega -0.393055 0.000131 -2994.241 0  
## alpha1 -0.039314 0.000196 -200.370 0  
## alpha2 -0.091478 0.000326 -280.636 0  
## alpha3 0.105783 0.000125 845.490 0  
## alpha4 -0.058935 0.000109 -541.363 0  
## alpha5 0.107019 0.000379 282.072 0  
## beta1 1.000000 0.000048 20684.677 0  
## beta2 -0.651062 0.000019 -33724.272 0  
## beta3 0.636256 0.000024 26033.540 0  
## beta4 0.323917 0.000001 436726.666 0  
## beta5 -0.363794 0.000373 -976.311 0  
## gamma1 0.485547 0.000031 15471.912 0  
## gamma2 -0.206571 0.000082 -2524.330 0  
## gamma3 0.301778 0.000011 26334.676 0  
## gamma4 -0.265613 0.000075 -3528.062 0  
## gamma5 -0.086489 0.000102 -848.201 0  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.002201 0.000144 15.3092 0.000000  
## ma1 -0.021568 0.000459 -46.9523 0.000000  
## ma2 -0.098398 0.020282 -4.8516 0.000001  
## mxreg1 -0.002903 0.000143 -20.2625 0.000000  
## mxreg2 -0.004245 0.000873 -4.8604 0.000001  
## mxreg3 -0.000809 0.000335 -2.4175 0.015626  
## mxreg4 0.001118 0.000153 7.3176 0.000000  
## omega -0.393055 0.000615 -639.1560 0.000000  
## alpha1 -0.039314 0.000994 -39.5700 0.000000  
## alpha2 -0.091478 0.001623 -56.3590 0.000000  
## alpha3 0.105783 0.000635 166.5716 0.000000  
## alpha4 -0.058935 0.000595 -99.0719 0.000000  
## alpha5 0.107019 0.001910 56.0318 0.000000  
## beta1 1.000000 0.000251 3990.3508 0.000000  
## beta2 -0.651062 0.000099 -6585.9074 0.000000  
## beta3 0.636256 0.000297 2145.7518 0.000000  
## beta4 0.323917 0.000002 136324.2299 0.000000  
## beta5 -0.363794 0.001885 -193.0300 0.000000  
## gamma1 0.485547 0.000160 3026.0001 0.000000  
## gamma2 -0.206571 0.000420 -492.3856 0.000000  
## gamma3 0.301778 0.000058 5168.4318 0.000000  
## gamma4 -0.265613 0.000380 -698.1138 0.000000  
## gamma5 -0.086489 0.000478 -181.1100 0.000000  
##   
## LogLikelihood : 3035.672   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike -4.6207  
## Bayes -4.5294  
## Shibata -4.6213  
## Hannan-Quinn -4.5864  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 8.22 4.144e-03  
## Lag[2\*(p+q)+(p+q)-1][5] 10.08 2.146e-13  
## Lag[4\*(p+q)+(p+q)-1][9] 13.51 2.171e-04  
## d.o.f=2  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 0.07886 0.7788  
## Lag[2\*(p+q)+(p+q)-1][29] 7.83934 0.9673  
## Lag[4\*(p+q)+(p+q)-1][49] 19.76396 0.8147  
## d.o.f=10  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[11] 0.2049 0.500 2.000 0.6508  
## ARCH Lag[13] 0.2714 1.493 1.855 0.9607  
## ARCH Lag[15] 1.0705 2.471 1.754 0.9401  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: no.parameters>20 (not available)  
## Individual Statistics:   
## mu 0.02567  
## ma1 0.03537  
## ma2 0.01394  
## mxreg1 0.02765  
## mxreg2 0.01763  
## mxreg3 0.01399  
## mxreg4 0.02533  
## omega 0.02666  
## alpha1 0.02078  
## alpha2 0.02438  
## alpha3 0.02237  
## alpha4 0.02642  
## alpha5 0.02212  
## beta1 0.03335  
## beta2 0.01447  
## beta3 0.02442  
## beta4 0.03477  
## beta5 0.03282  
## gamma1 0.02124  
## gamma2 0.03016  
## gamma3 0.02147  
## gamma4 0.02641  
## gamma5 0.02270  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 1.712 0.08715 \*  
## Negative Sign Bias 1.062 0.28826   
## Positive Sign Bias 0.463 0.64346   
## Joint Effect 3.131 0.37188   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 117.1 3.900e-16  
## 2 30 137.6 4.526e-16  
## 3 40 157.2 4.046e-16  
## 4 50 156.3 3.973e-13  
##   
##   
## Elapsed time : 4.662502

Najlepszym modelem z rodziny eGARCH okazał się model o parametrach P=5 oraz Q=5 (pod względem kryterium informacyjnego AIC). Choć według kryteriów BIC nie jest ona najlepszy, to jednak również może być traktowany jako jeden z lepszych wyborów. Szczególnie, że Model ten posiada wszystkie parametry istotne.

#### SARIMA - sGARCH

## Kryterium AIC

## Q=1 Q=2 Q=3 Q=4 Q=5  
## P=1 -4.563266 -4.572916 -4.571666 -4.573499 -4.571695  
## P=2 -4.561733 -4.571382 -4.570132 -4.572128 -4.570328  
## P=3 -4.559952 -4.569641 -4.568598 -4.572850 -4.570894  
## P=5 -4.556960 -4.566639 -4.566941 -4.571573 -4.570134

##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : sGARCH(1,2)  
## Mean Model : ARFIMA(0,0,2)  
## Distribution : norm   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001963 0.001322 1.48440 0.137704  
## ma1 0.008722 0.034458 0.25313 0.800169  
## ma2 -0.072794 0.030672 -2.37330 0.017630  
## mxreg1 -0.001175 0.001872 -0.62792 0.530056  
## mxreg2 -0.003092 0.001933 -1.59953 0.109704  
## mxreg3 -0.000585 0.001968 -0.29747 0.766106  
## mxreg4 0.001238 0.001916 0.64626 0.518111  
## omega 0.000075 0.000018 4.20677 0.000026  
## alpha1 0.227513 0.037368 6.08841 0.000000  
## beta1 0.113072 0.073769 1.53278 0.125330  
## beta2 0.564655 0.068095 8.29219 0.000000  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001963 0.001430 1.37303 0.169742  
## ma1 0.008722 0.036873 0.23655 0.813010  
## ma2 -0.072794 0.031192 -2.33372 0.019611  
## mxreg1 -0.001175 0.002023 -0.58103 0.561221  
## mxreg2 -0.003092 0.002293 -1.34810 0.177625  
## mxreg3 -0.000585 0.001990 -0.29421 0.768599  
## mxreg4 0.001238 0.001950 0.63501 0.525423  
## omega 0.000075 0.000032 2.37150 0.017716  
## alpha1 0.227513 0.065937 3.45044 0.000560  
## beta1 0.113072 0.118412 0.95491 0.339624  
## beta2 0.564655 0.106577 5.29810 0.000000  
##   
## LogLikelihood : 2992.541   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike -4.5729  
## Bayes -4.5293  
## Shibata -4.5731  
## Hannan-Quinn -4.5565  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 4.960 2.594e-02  
## Lag[2\*(p+q)+(p+q)-1][5] 6.742 3.352e-06  
## Lag[4\*(p+q)+(p+q)-1][9] 10.182 7.703e-03  
## d.o.f=2  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 0.2626 0.6084  
## Lag[2\*(p+q)+(p+q)-1][8] 1.7120 0.9005  
## Lag[4\*(p+q)+(p+q)-1][14] 3.9881 0.8814  
## d.o.f=3  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[4] 0.08423 0.500 2.000 0.7716  
## ARCH Lag[6] 0.70380 1.461 1.711 0.8337  
## ARCH Lag[8] 1.92601 2.368 1.583 0.7565  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: 1.2844  
## Individual Statistics:   
## mu 0.17037  
## ma1 0.35138  
## ma2 0.02987  
## mxreg1 0.28968  
## mxreg2 0.04666  
## mxreg3 0.07860  
## mxreg4 0.02602  
## omega 0.04444  
## alpha1 0.06767  
## beta1 0.03893  
## beta2 0.04386  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Joint Statistic: 2.49 2.75 3.27  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 1.5054 0.1325   
## Negative Sign Bias 1.1961 0.2319   
## Positive Sign Bias 0.4085 0.6830   
## Joint Effect 2.6433 0.4500   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 162.2 9.613e-25  
## 2 30 167.4 2.057e-21  
## 3 40 189.8 1.033e-21  
## 4 50 203.7 9.090e-21  
##   
##   
## Elapsed time : 0.5196438

Spośród rozpatrywanych modeli sGARCH najlepsze wartości kryterium informacyjnego AIC posiadały modele (1,4) oraz (1,2). Pierwszy z modeli nie posiadał natomiast wszystkich parametrów istotnych - mogło być to spowodowane przeparametryzowaniem. Z tego powodu model o niejszej wartości Q powinien spełnić wymóg istotności parametrów - tak jednak się nie stało i model ten również posiadał parametry nieistotne - na przykład ma1, co wykluczało jego wykorzystanie w dalszej analizie.

#### SARIMA - gjrGARCH

## Kryterium AIC

## Q=1 Q=2 Q=3 Q=4 Q=5  
## P=1 -4.562316 -4.571524 -4.570233 -4.572173 -4.570354  
## P=2 -4.561173 -4.568478 -4.567215 -4.570015 -4.568189  
## P=3 -4.557975 -4.565342 -4.564182 -4.574971 -4.573007

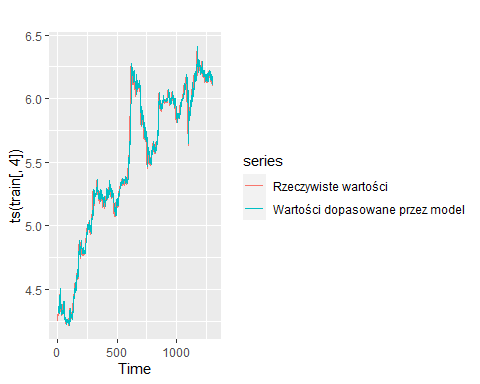
##   
## \*---------------------------------\*  
## \* GARCH Model Fit \*  
## \*---------------------------------\*  
##   
## Conditional Variance Dynamics   
## -----------------------------------  
## GARCH Model : gjrGARCH(3,5)  
## Mean Model : ARFIMA(0,0,2)  
## Distribution : norm   
##   
## Optimal Parameters  
## ------------------------------------  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001560 0.001346 1.158717 0.246571  
## ma1 0.021943 0.035146 0.624350 0.532398  
## ma2 -0.051624 0.032074 -1.609547 0.107497  
## mxreg1 -0.000926 0.001890 -0.490128 0.624043  
## mxreg2 -0.003023 0.001925 -1.570642 0.116266  
## mxreg3 -0.000074 0.001958 -0.037849 0.969808  
## mxreg4 0.001538 0.001919 0.801855 0.422637  
## omega 0.000122 0.000068 1.797786 0.072211  
## alpha1 0.328590 0.060155 5.462400 0.000000  
## alpha2 0.001911 0.267754 0.007137 0.994306  
## alpha3 0.024028 0.046336 0.518546 0.604077  
## beta1 0.000000 0.795332 0.000000 1.000000  
## beta2 0.000000 0.064123 0.000000 1.000000  
## beta3 0.128939 0.066564 1.937073 0.052736  
## beta4 0.353760 0.049323 7.172292 0.000000  
## beta5 0.000000 0.352582 0.000000 1.000000  
## gamma1 -0.128113 0.072988 -1.755263 0.079214  
## gamma2 0.051091 0.166639 0.306600 0.759148  
## gamma3 0.100052 0.133278 0.750706 0.452830  
##   
## Robust Standard Errors:  
## Estimate Std. Error t value Pr(>|t|)  
## mu 0.001560 0.001636 0.953650 0.340261  
## ma1 0.021943 0.048773 0.449900 0.652783  
## ma2 -0.051624 0.056302 -0.916919 0.359185  
## mxreg1 -0.000926 0.002052 -0.451377 0.651718  
## mxreg2 -0.003023 0.002285 -1.323067 0.185813  
## mxreg3 -0.000074 0.002218 -0.033411 0.973347  
## mxreg4 0.001538 0.002605 0.590547 0.554824  
## omega 0.000122 0.000246 0.493785 0.621458  
## alpha1 0.328590 0.098169 3.347188 0.000816  
## alpha2 0.001911 1.004326 0.001903 0.998482  
## alpha3 0.024028 0.191909 0.125203 0.900363  
## beta1 0.000000 2.996221 0.000000 1.000000  
## beta2 0.000000 0.331431 0.000000 1.000000  
## beta3 0.128939 0.159569 0.808047 0.419064  
## beta4 0.353760 0.162870 2.172036 0.029853  
## beta5 0.000000 1.329368 0.000000 1.000000  
## gamma1 -0.128113 0.144345 -0.887546 0.374785  
## gamma2 0.051091 0.615717 0.082979 0.933868  
## gamma3 0.100052 0.509468 0.196386 0.844308  
##   
## LogLikelihood : 3000.6   
##   
## Information Criteria  
## ------------------------------------  
##   
## Akaike -4.5730  
## Bayes -4.4976  
## Shibata -4.5734  
## Hannan-Quinn -4.5447  
##   
## Weighted Ljung-Box Test on Standardized Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 2.926 0.08719  
## Lag[2\*(p+q)+(p+q)-1][5] 4.961 0.00378  
## Lag[4\*(p+q)+(p+q)-1][9] 8.286 0.04519  
## d.o.f=2  
## H0 : No serial correlation  
##   
## Weighted Ljung-Box Test on Standardized Squared Residuals  
## ------------------------------------  
## statistic p-value  
## Lag[1] 0.06582 0.7975  
## Lag[2\*(p+q)+(p+q)-1][23] 6.38211 0.9427  
## Lag[4\*(p+q)+(p+q)-1][39] 13.68939 0.9003  
## d.o.f=8  
##   
## Weighted ARCH LM Tests  
## ------------------------------------  
## Statistic Shape Scale P-Value  
## ARCH Lag[9] 0.6637 0.500 2.00 0.4153  
## ARCH Lag[11] 3.1488 1.490 1.83 0.3254  
## ARCH Lag[13] 3.3474 2.459 1.72 0.5539  
##   
## Nyblom stability test  
## ------------------------------------  
## Joint Statistic: 5.3821  
## Individual Statistics:   
## mu 0.17874  
## ma1 0.21337  
## ma2 0.02777  
## mxreg1 0.29010  
## mxreg2 0.03780  
## mxreg3 0.06702  
## mxreg4 0.02809  
## omega 0.07406  
## alpha1 0.04003  
## alpha2 0.08645  
## alpha3 0.06393  
## beta1 0.06113  
## beta2 0.05740  
## beta3 0.04366  
## beta4 0.05232  
## beta5 0.05311  
## gamma1 0.05527  
## gamma2 0.06576  
## gamma3 0.17487  
##   
## Asymptotic Critical Values (10% 5% 1%)  
## Joint Statistic: 4.03 4.33 4.92  
## Individual Statistic: 0.35 0.47 0.75  
##   
## Sign Bias Test  
## ------------------------------------  
## t-value prob sig  
## Sign Bias 1.260 0.2080   
## Negative Sign Bias 1.316 0.1883   
## Positive Sign Bias 1.012 0.3115   
## Joint Effect 2.810 0.4219   
##   
##   
## Adjusted Pearson Goodness-of-Fit Test:  
## ------------------------------------  
## group statistic p-value(g-1)  
## 1 20 170.8 1.967e-26  
## 2 30 175.6 6.473e-23  
## 3 40 183.5 1.280e-20  
## 4 50 201.9 1.848e-20  
##   
##   
## Elapsed time : 2.679833

Podczas projektu dotyczącego spółki z sektora farmaceutycznego to właśnie model gjr-GARCH najlepiej przedstawił badany szereg czasowy. Dla danych dotyczących polskiego gamedevu wyniki były jednak równie rozczarowujące co w przypadku modeli sGARCH. Żaden z trzech najlepszych modeli, pomimo różnej liczby parametrów nie posiadał wszystkich wartości istotnych. Były to modele: (3,5)/(3,4)/(1,2)

Zatem ostateczną postacią modelu okazał się model SARIMA-eGARCH(0,1,1)(5,1,5)[5]. Jest to jedyny model, w którym wszystkie zmienne są istotne.

# Prognoza na zbiorze uczącym (Zadanie 9)

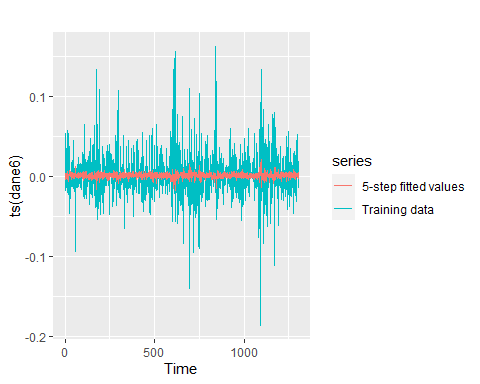
Aby uzyskać wartości dopasowane przez model SARIMA(0,1,2)(0,1,1)[5] posłużono się funkcją fitted(). Parametr h oznaczał w niej przesunięcie - wybrano wartości odleglejsze o 5, od liczby dni roboczych w tygodniu, wyznaczających zarazem odległość pomiędzy analogicznymi sezonami. Następnie stworzono wykres porównujący wartości rzeczywiste z wartościami dopasowanymi przez model (funkcja autoplot()). W tym celu niezbędne było potraktowanie danych jako typ time series (funkcja ts()). Jak można zauważyć, wartości dopasowane przez model odpowiadają wartością rzeczywistym - zatem na danych treningowych model sprawdził się doskonale.



## Zadanie 9 (GARCH)

Postanowiono sprawdzić dodatkowo również dokładność dopasowania danych z modelu GARCH. Wykres otrzymany za pomocą funkcji autoplot() znacznie odbiega od wykresu dla modelu SARIMA. Postać wykresu oczywiście jest inna ze względu na inny typ danych (SARIMA - logarytmy z notowań; SARIMA-GARCH - logarytmiczne stopy zwrotu). Lecz wartości dopasowane przez model SARIMA-eGARCH(0,1,2)(5,1,5)[5] nie posiadają aż tak znaczących odchyleń od 0, jak rzeczywiste dane - natomiast kierunek ich zmian jest podobny. Jest to natomiast poprawna zależność - analogiczne rezultaty uzyskali S. Chand, S. Kamal oraz I. Ali w artykule *Modeling and volatility analysis of share prices using ARCH and GARCH models*[9] , utrzymując przy tym poprawność dopasowania modelu. Przedstawiono również tablicę pomyłek (confusion matrix) przedstawiającą dokładność przewidywań modelu. Przy określeniu progu na poziomie .85 percentyla (wartości powyżej to wystąpienie “szoku giełdowego”, określane jako TRUE, a poniżej to “norma” - FALSE), większość wskazań modelu zakwalifikowano jako TN - poprawne wskazanie stanu FALSE. Część wskazań to FN - błędne wskazanie stanu FALSE, podczas gdy w rzeczywistości był to TRUE. Natomiast nie wystąpiły zarówno wartości TP, jak i FP - co wynika z niewielkich wahań przyjmowanych przez wartości dopasowane przez model. Takie podejście do weryfikacji poprawności dopasowania modelu z częścią GARCH zaprezentował Bohdan Pavlyshenko[10].  
Postanowiono wykonać również badanie poprawności prognoz pod kątem znaku. Gdy stopa zwrotu była liczbą dodatnią, to była klasyfikowana jako 1. W każdym innym przypadku traktowana była ona jako 0. Rezultaty porównania tego sposobu klasyfikacji przez model z danymi rzeczywistymi przedstawia druga macierz pomyłek. Dla tego sposobu klasyfikacji obliczono również miary takie jak: \* dokładność (59,8%) - taka część zbioru została poprawnie zaprognozowana \* czułość (63,3%) - określa procent pokrycia klasy pozytywnej przez pozytywne przewidywanie \* specyficzność (29,2%) - określa procent pokrycia klasy negatywnej przez prognozę negatywną

Stworzono również krzywą ROC, która jest graficznym przedstawieniem efektywności modelu - im jest on lepszy, tym bardziej krzywa oddala się od krzywej y=x. Otrzymane AUC wynosi 0,53, co jest niezbyt zadowalającym wynikiem (dla AUC-0,5 klasyfikator jest losowy, a dla równego 1 - idealny)[11].



## confusion matrix, przy progu 0.85 (0.85 percentyl)

## Actual-positive Actual-negative  
## Predicted-positive 0 0  
## Predicted-negative 196 1108

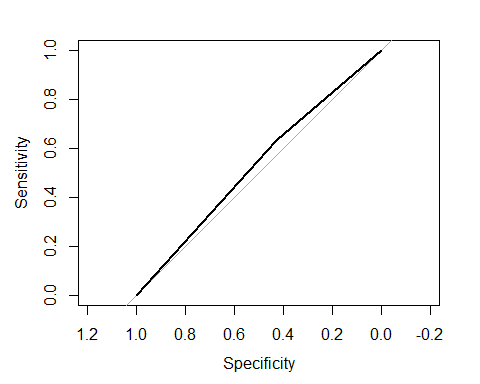
## Type 'citation("pROC")' for a citation.

##   
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases



## Pole pod krzywą ROC (AUC) wynosi 0.531485

## confusion matrix, przy poprawności znaku (+ to 1, wartość 0 lub - to 0)

## Actual-positive Actual-negative  
## Predicted-positive 366 312  
## Predicted-negative 212 414

## Dokładność 0.5981595

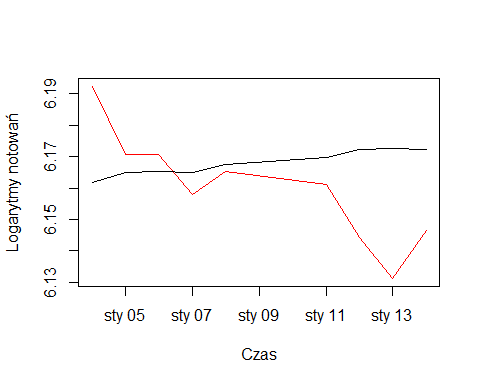
## Czułość 0.633218

## Specyficzność 0.292011

# Prognoza na zbiorze testowym (Zadanie 11)

Po wyestymowaniu zarówno modelu SARIMA, jak i GARCH oraz weryfikacji ich na zbiorze treningowym, postanowiono sprawdzić rezultaty ich prognoz dotyczących zbioru testowego. Satysfakcjonujące rezultaty dotyczące zbioru treningowego nie muszą bowiem oznaczać idealnego modelu, który potrafiłby skutecznie przewidywać przyszłość. Modele budowane są w oparciu o przeszłe dane i to w oparciu o nie estymowane są ich parametry, zatem często da zupełnie nowych danych ich wyniki nie spełniają oczekiwań. Szerzej w uczeniu maszynowym takie zjawisko określa się mianem przetrenowania modelu. Taki model tak mocno dostosował się do znanych danych i charakteryzuącej je struktury, że dla kolejnych obserwacji odbiegających od przyjętego schematu staje się wręcz bezradny.  
Weryfikacji poddano zarówno model SARIMA jak i SARIMA-GARCH:

## Prognozowanie - SARIMA



## Prognoza Wartosc realna y-yP |y-yp| |(y-yp)/y|  
## [1,] 6.1619 6.1924 0.0305 0.0305 0.0049  
## [2,] 6.1649 6.1707 0.0057 0.0057 0.0009  
## [3,] 6.1654 6.1707 0.0053 0.0053 0.0009  
## [4,] 6.1649 6.1580 -0.0068 0.0068 0.0011  
## [5,] 6.1676 6.1654 -0.0022 0.0022 0.0004  
## [6,] 6.1698 6.1612 -0.0086 0.0086 0.0014  
## [7,] 6.1723 6.1442 -0.0281 0.0281 0.0046  
## [8,] 6.1727 6.1312 -0.0415 0.0415 0.0068

## ME: -0.0057

## MAE: 0.0161

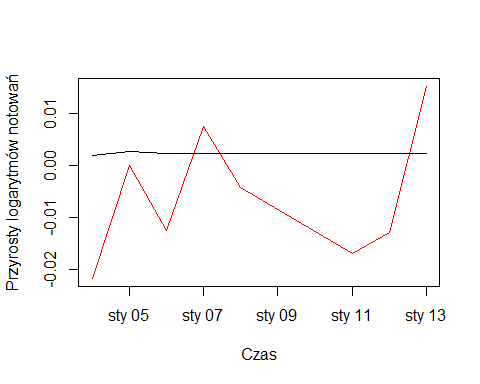
## MSE: 5e-04

## MAPE: 0.2614 %

Predykcji dokonano z wykorzystaniem funkcji predict(). Powyższy wykres prezentuje wartości rzeczywiste oraz prognozowane przez model. Zauważyć można znaczne różnice, natomiast model dąży do symulacji wahań. Powyższa tabela przedstawia konkretne wartości prognoz, wartości rzeczywistych oraz ich różnic, służących do obliczenia błędów predykcji ex-post (ME,MAE,MSE oraz MAPE).

* ME (-0.0057) - średni błąd predykcji (ang. mean error - ME) ex post . Stosowanie tego błędu jest niewskazane, gdy wartości różnic pomiędzy wartością rzeczywistą a prognozowaną są zarówno dodatnie jak i ujemne. ME jest wówczas sztucznie zaniżany, gdyż wartości te niwelują się nawzajem - po przeanalizowaniu tabeli z prognozami zauważyć można, że właśnie takie zjawisko miało miejsce.
* MAE (MAE: 0.0161) - średni bezwzględny błąd predykcji (ang. mean absolute error - MAE) ex post. Różnica wartości MAE i ME wskazywać może na wahania zmiennej objaśnianej w czasie. Jej wartość zmniejszała się i zwiększała, przez co różnica jej wartości od wartości prognozowanej była dla części obserwacji ujemna, a dla części - dodatnia. W projekcie MAE jest około trzy razy większy niż ME, co potwierdza, że nie powinno stosować się zwykłego średniego błędu predykcji do oceny jakości prognoz
* MSE (0,0005) - średniokwadratowy błąd predykcji (ang. mean square error - MSE) ex post. Błąd ten wykorzystuje kwadrat różnicy pomiędzy wartością rzeczywistą zmiennej objaśnianej a jej prognozą. Pozwala to sprawdzić, czy prognoza zawiera znaczne błędy - zostaną one wówczas podkreślone w dużej wartości MSE. Zatem błąd ten ma zastosowanie w ocenie poprawności wykonania prognozy, gdyż wskaże istotne różnice mogące wynikać z błędu w formule lub danych (np. występowanie outlierów). MSE może posłużyć również do sprawdzenia użyteczności prognoz, ponieważ jego duża wartość wskazuje na wysokie ryzyko otrzymania wyniku znacznie odbiegającego od rzeczywistości. Natomiast gdy głównym celem jest zaprezentowanie różnic pomiędzy prognozami dokonywanymi z użyciem stworzonego modelu, a rzeczywistością lepiej sprawdzi się MAE. Błąd ten nie tylko jest bardziej miarodajny niż ME, lecz posiada ten sam rząd wielkości co dane - zatem można interpretować go w porównaniu do wartości prognoz. Uzyskana wielkość MSE podkreśla kolejną z jego wad - dla danych w których różnice pomiędzy prognozami, a wartościami rzeczywistymi wyrazić można w częściach dziesiątych czy też setnych błąd średniokwadratowy przybiera bardzo niewielkie wartości (co wynika z podnoszenia do kwadratu). Dla rozpatrywanych danych nie jest to bardzo negatywny aspekt (który jak podkreślono powyżej, potwierdza brak znaczących błędów), natomiast dla obliczeń w których precyzja jest niezmiernie istotna, a rząd wielkości niewielki (np. stężenia niebezpiecznych substancji w powietrzu) MSE zupełnie się nie sprawdzi.
* MAPE (0,2614%) - średni bezwzględny procentowy błąd predykcji (ang. mean absolute percentage error - MAPE) pozwala na jeszcze bardziej intuicyjną ocenę prognoz, gdyż wyraża się go jako procentowa część wartości prognozy. W posługiwaniu się nim nie trzeba znać zatem wartości prognoz, aby określić ich jakość. Błąd ten nie powinien być jednak stosowany w każdym przypadku - ze względu na fakt, że w mianowniku równania znajduje się wartość zmiennej objaśnianej, dla danych w których zmienne te są bliskie 0 wartości MAPE będą skrajnie duże (a dla y=0 wystąpi nawet błąd kalkulacji). Dla omawianych danych, w których wartości są znacznie większe niż 0, błąd ten sprawdza się i pozwala na intuicyjną interpretację - model SARIMA w prognozach myli się średnio o 0,26% ich wartości.

## Prognozowanie - SARIMA-GARCH



## Prognoza Wartosc realna y-yP |y-yp| |(y-yp)/y|  
## [1,] 0.0019 -0.0217 -0.0237 0.0237 1.0898  
## [2,] 0.0026 0.0000 -0.0026 0.0026 Inf  
## [3,] 0.0022 -0.0126 -0.0148 0.0148 1.1744  
## [4,] 0.0022 0.0074 0.0052 0.0052 0.7018  
## [5,] 0.0022 -0.0042 -0.0064 0.0064 1.5226  
## [6,] 0.0022 -0.0170 -0.0192 0.0192 1.1293  
## [7,] 0.0022 -0.0130 -0.0152 0.0152 1.1698  
## [8,] 0.0022 0.0151 0.0129 0.0129 0.8543

## ME: -0.008

## MAE: 0.0125

## MSE: 2e-04

## MAPE: 95.525 %

Predykcji dokonano z wykorzystaniem funkcji ugarchforecast(). Powyższy wykres prezentuje wartości rzeczywiste oraz prognozowane przez model. Zauważyć można znaczne różnice, natomiast model (w przeciwieństwie do SARIMY) przyjmuje wartości, które nie różnią się zbytnio od siebie (ich różnice są o rząd mniejsze niż rzeczywistych wahań). Powyższa tabela dokładnie przedstawia wartości prognoz, wartości rzeczywistych oraz ich różnic, służących do obliczenia błędów predykcji ex-post (ME,MAE,MSE oraz MAPE).:

* ME (-0,008)
* MAE (0,0125)
* MSE (0,0002)
* MAPE (95,525%)

Wartości pierwszych trzech błędów łączy podobna relacja co w przypadku SARIMY. Należy pamiętać jednak o znacznej różnicy w rzędzie wielkości danych (jedności i części setne). W tym kontekście błędy prognoz drugiego z rozpatrywanych modeli są znaczące, a prognoza wydaje się bardzo niedokładna. Warto zwrócić uwagę na mniejsze “znoszenie się” wartości dodatnich i ujemnych w różnicach (MAE jest tylko 64% większe niż ME) oraz na znaczną wartość błędu MAPE. Błąd ten musiał zostać jednak obliczony dla mniejszej liczby danych, gdyż wśród wartości realnych wystąpiło 0, co spwodowało wartość błędu równą nieskończoność. Po poprawie błąd ten wynosi niemal 100%, co jest zupełną skrajnością dla wartości 0,21% dotyczącej modelu SARIMA. Ocena prognozy modelu SARIMA-GARCH ukazuje szczególnie zarówno zalety (kontekst błędu) jak i wady (ryzyko otrzymania niepoprawnych wyników dla danych zbliżonych do 0) błędu MAPE.

# Budowa modelu VAR

Model VAR jest jednym z etapów przed zastosowaniem testu przyczynowości Grangera. Model wektorowo-autoregresyjny (VAR) jest to model wielorównaniowy, wykorzystywany do analizy wielu szeregów czasowych, w którym jako zmienne egzogeniczne występują zarówno opóźnienia zmiennej objaśnianej jak i opóźnienia pozostałych zmiennych. Uogólniona postać modelu VAR wygląda następująco:

Gdzie:  
- to wektor obserwacji bieżących wartości wszystkich n zmiennych modelu,

- to macierze parametrów przy opóźnionych zmiennych, które nie zawierają elementów zerowych,

- to wektor procesów resztkowych, w którym poszczególne składowe powinny być ze sobą skorelowane, lecz nie zawierają autokorelacji, - k to rząd modelu VAR.

- to wektor deterministycznych składowych równań,

- to macierz parametrów przy zmiennych wektora

Do estymacji parametrów modelu VAR można posłużyć się metodą najmniejszych kwadratów.

## Wybór opóźnienia i weryfikacja założeń

Przed budową modelu należy wybrać opóźnienie. Wybór ten zostanie dokonany za pomocą funkcji VARselect() dostępnej w języku R. Wspomniana funkcja dostarcza informacji, dla jakich opóźnień wartości kryteriów informacyjnych będą najmniejsze, co znacznie zawęża zbior dostępnych opóźnień.

## $selection  
## AIC(n) HQ(n) SC(n) FPE(n)   
## 2 2 1 2   
##   
## $criteria  
## 1 2 3 4 5  
## AIC(n) -1.464578e+01 -1.465925e+01 -1.465740e+01 -1.465535e+01 -1.465311e+01  
## HQ(n) -1.463667e+01 -1.464407e+01 -1.463615e+01 -1.462803e+01 -1.461971e+01  
## SC(n) -1.462152e+01 -1.461882e+01 -1.460081e+01 -1.458259e+01 -1.456418e+01  
## FPE(n) 4.359339e-07 4.301004e-07 4.308948e-07 4.317782e-07 4.327466e-07  
## 6 7 8 9 10  
## AIC(n) -1.464900e+01 -1.464278e+01 -1.464077e+01 -1.463770e+01 -1.463725e+01  
## HQ(n) -1.460953e+01 -1.459723e+01 -1.458915e+01 -1.458001e+01 -1.457348e+01  
## SC(n) -1.454390e+01 -1.452151e+01 -1.450333e+01 -1.448410e+01 -1.446747e+01  
## FPE(n) 4.345315e-07 4.372435e-07 4.381229e-07 4.394685e-07 4.396709e-07  
## 11 12 13 14 15  
## AIC(n) -1.463245e+01 -1.462865e+01 -1.462684e+01 -1.462343e+01 -1.462182e+01  
## HQ(n) -1.456262e+01 -1.455274e+01 -1.454486e+01 -1.453538e+01 -1.452769e+01  
## SC(n) -1.444651e+01 -1.442653e+01 -1.440855e+01 -1.438898e+01 -1.437119e+01  
## FPE(n) 4.417835e-07 4.434686e-07 4.442729e-07 4.457903e-07 4.465126e-07  
## 16 17 18 19 20  
## AIC(n) -1.461915e+01 -1.461888e+01 -1.461562e+01 -1.461166e+01 -1.460633e+01  
## HQ(n) -1.451895e+01 -1.451260e+01 -1.450328e+01 -1.449324e+01 -1.448183e+01  
## SC(n) -1.435236e+01 -1.433591e+01 -1.431649e+01 -1.429636e+01 -1.427486e+01  
## FPE(n) 4.477050e-07 4.478314e-07 4.492928e-07 4.510801e-07 4.534948e-07  
## 21 22 23 24 25  
## AIC(n) -1.460171e+01 -1.459909e+01 -1.459324e+01 -1.459063e+01 -1.459002e+01  
## HQ(n) -1.447114e+01 -1.446246e+01 -1.445053e+01 -1.444184e+01 -1.443517e+01  
## SC(n) -1.425407e+01 -1.423529e+01 -1.421326e+01 -1.419448e+01 -1.417771e+01  
## FPE(n) 4.555983e-07 4.567928e-07 4.594799e-07 4.606846e-07 4.609673e-07  
## 26 27 28 29 30  
## AIC(n) -1.458703e+01 -1.458354e+01 -1.458636e+01 -1.458160e+01 -1.457807e+01  
## HQ(n) -1.442610e+01 -1.441654e+01 -1.441328e+01 -1.440245e+01 -1.439285e+01  
## SC(n) -1.415855e+01 -1.413889e+01 -1.412553e+01 -1.410461e+01 -1.408491e+01  
## FPE(n) 4.623538e-07 4.639746e-07 4.626756e-07 4.648887e-07 4.665391e-07

W analizowanym przypadku trzy kryteria wskazały na opóźnienie równe 2, a jedno na opóźnienie równe 1. DLatego też najpierw został zbudowany model VAR(1), w którym następnie zweryfikowano jedno z najważniejszych założeń, które wskazuje na poprawność modelu, czyli występowanie autokorelacji w resztach. Wykorzystano do tego test Ljung-Box’a, w którym hipotezy mają postać:

H0: reszty rozkładają się niezależnie (brak autokorelacji)  
H1: ~ H0

Wnioskowanie w teście Ljung-Boxa jest identyczne jak w pozostałych testach - gdy wartośc statystyki testowej jest większa niż wartość krytyczna, odrzucamy hipotezę zerową o braku autokorelacji. Decyzję można podjąć również w oparciu o wartość p-value, gdy jest ona mniejsza niż przyjęty poziom istotności odrzucamy hipotezę zerową.

##   
## Box-Ljung test  
##   
## data: model\_var1$varresult$dane6$residuals  
## X-squared = 0.16455, df = 1, p-value = 0.685

##   
## Box-Ljung test  
##   
## data: model\_var1$varresult$dane7$residuals  
## X-squared = 0.00018557, df = 1, p-value = 0.9891

Wartości p-value z testu Ljung-Boxa są zdecydowanie większe niż poziom istotności równy 5%, więc nie ma podstaw do odrzucenia hipotezy zerowej, czyli można przyjąć, że w modelu nie występuje autokorelacja reszt.

# Przyczynowość Grangera

Po zbudowaniu modelu pora na najważniejszą część, czyli przeprowadzenie testu przyczynowości Grangera, który pozwoli wykazać siłę oraz kierunek zależności pomiędzy rozważanymi szeregami czasowymi.

Test przyczynowości Grangera oraz samo pojęcie przyczynowości, na którym oparty jest ten test stanowi centralny punkt rozważań na temat zależności przyczynowych pomiędzy dwoma analizowanymi szeregami czasowymi oraz w szerszym kontekście pomiędzy firmami 11 bit Studios oraz CD Projekt.

Definicja przyczynowości, jaką wprowadził Granger polega na założeniu, że skutek nie może poprzedzać przyczyny. Dlatego więc, mając dwie zmienne np. X i Y mówimy, że X jest przyczyną w sensie Grangera dla Y, jeśli przeszłe wartości zmiennej X pozwalają lepiej prognozować obecne wartości Y przy niezmienionych pozostałych informacjach [12].

## Test Grangera

W tesćie Grangera utworzyć trzeba dwa modele - jeden z pełną informacją, a drugi z restrykcjami:

hipotezy w teście Grangera mają następujący układ:

dla każdego - brak przyczynowości  
dla każdego - występuje przyczynowość

## Granger causality test  
##   
## Model 1: dane7 ~ Lags(dane7, 1:1) + Lags(dane6, 1:1)  
## Model 2: dane7 ~ Lags(dane7, 1:1)  
## Res.Df Df F Pr(>F)  
## 1 1300   
## 2 1301 -1 0.2937 0.588

## Granger causality test  
##   
## Model 1: dane6 ~ Lags(dane6, 1:1) + Lags(dane7, 1:1)  
## Model 2: dane6 ~ Lags(dane6, 1:1)  
## Res.Df Df F Pr(>F)  
## 1 1300   
## 2 1301 -1 1.499 0.221

Jak wykazał test Grangera zależność przyczynowa pomiędzy 11 bit Studios a CD Project nie występuje w żadnym kierunku, gdyż wartości p-value są zdecydowanie większe niż 0,05 (przyjęty poziom istotności). Jest to zaskakujący wynik, ponieważ wykresy cen zamknięcia podążały praktycznie przez cały czas w tym samym kierunku, co pozwalało przypuszczać, iż zależność będzie istniała.

# Kointegracja

Jeżeli szereg jest stacjonarny, to nie jest on zintegrowany () Jeżeli przyrost szeregu czasowego jest stacjonarny to szereg ten jest zintegrowany w stopniu 1 (). Na podstawie poprzedniej części sprawozdania powiedzieć można, że zarówno szereg czasowy reprezentujący logarytmy z notowań 11 bit, jak i CDPR są zintegrowane w stopniu 1.  
Natomiast jeżeli mielibyśmy 2 niestacjonarne szeregi y oraz x, to byłyby one skointegrowane, gdyby istniało Beta takie że:

Czyli kombinacja liniowa tych dwóch szeregów byłaby stacjonarnym szeregiem czasowym. Wówczas można zapisać równanie:

Jego interpretacja jest następująca: z czasem te szeregi nie oddalają się od siebie w sposób znaczący, więc występuje między nimi relacja długookresowej równowagi.  
Posiadamy 2 szeregi czasowe, które są niestacjonarne - aby sprawdzić czy są skointegrowane wykonamy test Johansena oraz procedurę Engle’a i Granger’a.

## Test Johansena

Po sprawdzeniu, że rozpatrywane szeregi czasowe są niestacjonarne, można wykonać test na wykrycie rzędu kointegracji. Søren Johansen zaproponował dwa rodzaje tego testu:

* test śladu - ze statystykami testowymi:
* test największej wartości własnej - ze statystykami testowymi:

W sprawozdaniu wykonana zostanie wersja testu śladu (za pomocą funkcji ca.jo() z parametrem type=“trace”). Jej statystyka ma postać:

Gdzie:  
 - liczba rozpatrywanych szeregów  
 - liczba obserwacji  
 - wartości własne  
 - testowany rząd kointegracji

Test ten jest wykonywany automatycznie dla szeregu hipotez, a ich liczba zależy od liczby rozpatrywanych szeregów. W projekcie są 2 zmienne, zatem sprawdzone zostaną 2 pary hipotez (o istnieniu 0 lub 1 wektora kointegrującego). Hipoteza alternatywna zawsze jest postaci, że istnieje istotnie więcej relacji kointegrujących niż przyjęto w hipotezie głównej. Zestawienie wyników przy różnych poziomach istotności pozwala podjąć decyzję o rzędzie kointegracji w modelu. Gdy statystyka testowa będzie większa niż odpowiednia wartość krytyczna, hipotezę główną odrzuca się na rzecz hipotezy alternatywnej. Jednym z argumentów, jakie należy podać do funkcji ca.jo() jest K. Jest to kolejność opóźnień szeregów (poziomów) w modelu VAR (według dokumentacji R). Przyjęto K=2, gdyż jest to wymiar macierzy i zarazem liczba rozpatrywanych szeregów (11 bit i CDPR).

##   
## ######################   
## # Johansen-Procedure #   
## ######################   
##   
## Test type: maximal eigenvalue statistic (lambda max) , with linear trend   
##   
## Eigenvalues (lambda):  
## [1] 0.009366761 0.002843583  
##   
## Values of teststatistic and critical values of test:  
##   
## test 10pct 5pct 1pct  
## r <= 1 | 3.74 6.50 8.18 11.65  
## r = 0 | 12.36 12.91 14.90 19.19  
##   
## Eigenvectors, normalised to first column:  
## (These are the cointegration relations)  
##   
## V4.l2 V5.l2  
## V4.l2 1.00000 1.00000000  
## V5.l2 -0.71927 0.08066684  
##   
## Weights W:  
## (This is the loading matrix)  
##   
## V4.l2 V5.l2  
## V4.d -0.009779923 -0.001724097  
## V5.d 0.008742354 -0.001710456

Dla przyjętego poziomu istotności nie ma podstaw do odrzucenia hipotezy głównej postaci . Na tej podstawie przyjąć można że rozpatrywane szeregi czasowe nie są skointegrowane.

## Procedura Engle’a i Granger’a

Alternatywą dla testu Johansena jest metoda przedstawiona przez Engle’a i Granger’a. Posiada ona mniejszą moc i gorsze właściwości statystyczne, niemniej jednak została ona zaprezentowana w formie ciekawostki. Jej pierwszym etapem jest estymacja z użyciem MNK. Następnie następuje wykonanie testu ADF dla reszt otrzymanej regrsji. Gdy odrzuci się hipotezę główną o braku stacjonarności oznaczać to będzie, że oszacowany wektor MNK jest wektorem kointegrującym. W rozpatrywanym przypadku procedura zostanie wykonana 2 razy - najpierw w MNK za y zostaną przyjęte logarytmy notowań 11 bit, a objaśniać będą je logarytmy notowań CDPR Następnie nastąpi zamiana. W pakiecie R istnieją również testy dotyczące tej procedury. Porównano autorską metodę (polegającą na wyestymowaniu modelu funkcją lm(), a następnie weryfikacji stacjonarności reszt tego modelu testem ADF) z funkcją coint.test() z pakietu aTSA.

## MNK w którym y=11bit, x=CDPR

##   
## Augmented Dickey-Fuller Test  
##   
## data: modelMNK$residuals  
## Dickey-Fuller = -3.003, Lag order = 10, p-value = 0.1537  
## alternative hypothesis: stationary

## Response: dane[, 4]   
## Input: dane[, 5]   
## Number of inputs: 1   
## Model: y ~ X + 1   
## -------------------------------   
## Engle-Granger Cointegration Test   
## alternative: cointegrated   
##   
## Type 1: no trend   
## lag EG p.value   
## 1.0000 -3.1075 0.0329   
## -----   
## Type 2: linear trend   
## lag EG p.value   
## 1.000 0.145 0.100   
## -----   
## Type 3: quadratic trend   
## lag EG p.value   
## 1.000 0.341 0.100   
## -----------   
## Note: p.value = 0.01 means p.value <= 0.01   
## : p.value = 0.10 means p.value >= 0.10

## MNK w którym y=CDPR, x=11bit

##   
## Augmented Dickey-Fuller Test  
##   
## data: modelMNK$residuals  
## Dickey-Fuller = -3.0462, Lag order = 10, p-value = 0.1355  
## alternative hypothesis: stationary

## Response: dane[, 5]   
## Input: dane[, 4]   
## Number of inputs: 1   
## Model: y ~ X + 1   
## -------------------------------   
## Engle-Granger Cointegration Test   
## alternative: cointegrated   
##   
## Type 1: no trend   
## lag EG p.value   
## 1.0000 -3.0657 0.0359   
## -----   
## Type 2: linear trend   
## lag EG p.value   
## 1.000 0.297 0.100   
## -----   
## Type 3: quadratic trend   
## lag EG p.value   
## 1.000 -0.509 0.100   
## -----------   
## Note: p.value = 0.01 means p.value <= 0.01   
## : p.value = 0.10 means p.value >= 0.10

Bazując na ręcznie wykonywanym teście, na poziomie istotności równym 0,05 brak podstaw do odrzucenia hipotezy głównej o braku stacjonarności, stąd można założyć że to wektor MNK nie jest wektorem kointegrującym. Natomiast funkcja coint.test dla wersji no trend zarówno w pierwszej jak i drugiej relacji wskazuje na istnienie kointegracji (wartości p-value są mniejsze niż przyjęty poziom istotności, co wskazuje na konieczność odrzucenia hipotezy głównej na rzecz hipotezy alternatywnej). Jak wskazano jednak powyżej, procedura Engle’a i Granger’a jest gorszym testem na wykrywanie kointegracji, zatem postanowiono przyjąć wyniki lepszego testu Johansena.

# Literatura

[1] Chan H. N., Time series. Applications to Finance with R and S-Plus, John Wiley & Sons, Hoboken 2010, s. 16

[2] Kozłowski E., Analiza i identyfikacja szeregów czasowych, Politechnika Lubelska, Lublin 2015, s. 192

[3] Maddala G. S., Ekonometria, s.612-614

[4] Niezbędnik SGH, Stacjonarność i niestacjonarność szeregów czasowych (<https://www.e-sgh.pl/niezbednik/plik.php?id=27273993&pid=3472>)

[5] Maddala G. S., op. cit., s.615-619

[6] Engle Robert F., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica, vol. 50, no. 4, 1982, pp. 987–1007. JSTOR,

[7] Greene W. H., Econometric Analysis, Pearson, 2012, s. 970.

[8] Hardy R. [How to fit a SARIMA + GARCH in R?](https://quant.stackexchange.com/questions/17129/how-to-fit-a-sarima-garch-in-r) (odpowiedź w serwisie StackExchange)

[9] Chand S., Kamal S., Ali I., [Modeling and volatility analysis of share prices using ARCH and GARCH models](https://www.researchgate.net/figure/Actual-fitted-and-residuals-under-ARIMA1-1-0-with-GARCH1-1-model_fig2_236270682)

[10] Pavlyshenko B., [Stock volatility prediction using GARCH models and machine learning approach](https://www.linkedin.com/pulse/stock-volatility-prediction-using-garch-models-bohdan-pavlyshenko)

[11] Wolak J., Uczenie statystyczne w R (wykład), s. 19-29

[12] Granger C. W. J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, Econometrica, vol. 37, no. 3, 1969, pp. 424–438. JSTOR, www.jstor.org/stable/1912791