

Zadanie 1 - PWC

Aleksander Mackiewicz-Kubiak

2024-11-12

Pakiety

```
library(tidyr)
library(gamlss)
library(dplyr)
library(fitdistrplus)
library(usefun)
library(quantmod)
library(scales)
```

Dane

Preprocessing danych

Dane, które wybieram to ceny zamknięcia indeksu giełdowego S&P 500 od 1 listopada 2021 roku do 30 października 2024 roku:

```
getSymbols("^GSPC", src = "yahoo", from = "2021-11-01", to = "2024-10-30")
```

```
## [1] "GSPC"
```

Tworzę z nich szereg czasowy:

```
## Time-Series [1:753] from 2021 to 2024: 4614 4631 4661 4680 4698 ...
```

Oraz sprawdzam ich podstawowe statystyki:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3577   4090   4415   4531   4925   5865
```

```
##
##      [,1]
## średnia  4531.3259
## odchylenie  572.1666
## moda    4137.6401
```

Ponieważ jest to szereg czasowy indeksów, ze sprecyzowaną ilością dni handlowych w roku, to jest znikoma szansa na wystąpienie jakiś braków w danych, natomiast dla pewności upewniam się komendą anyNA:

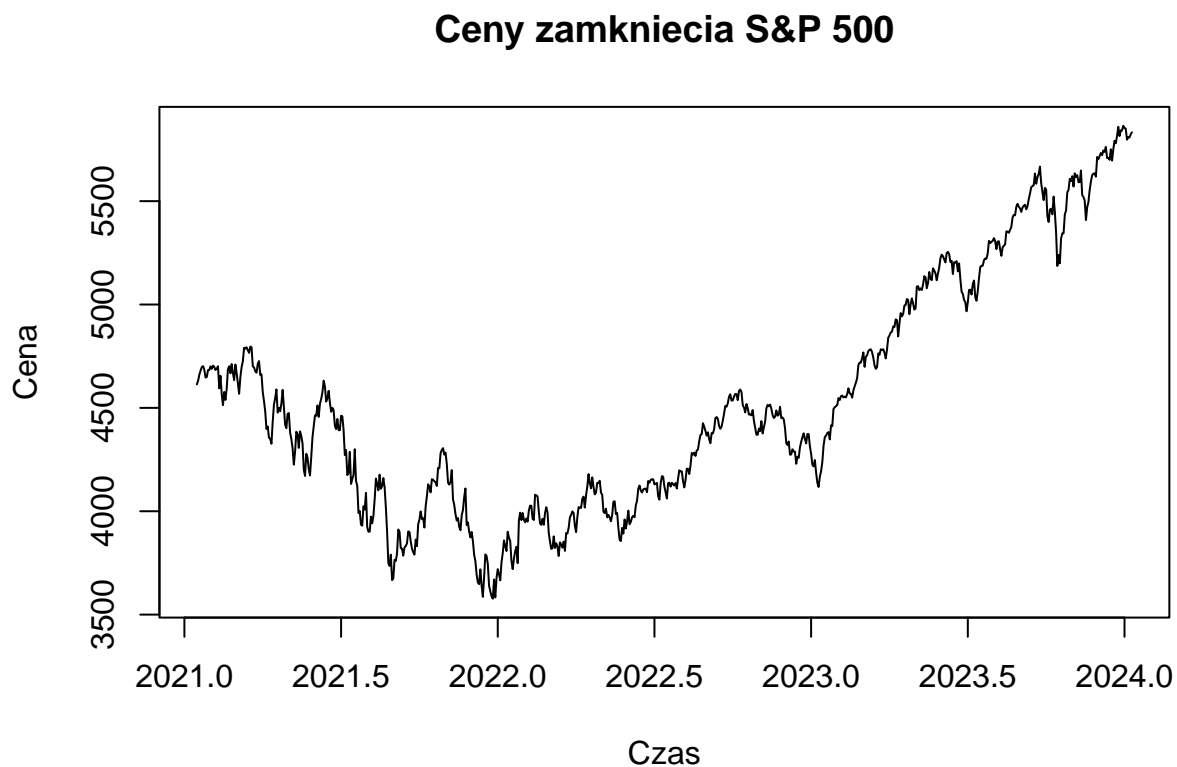
```
anyNA(dane)
```

```
## [1] FALSE
```

Wartość FALSE oznacza, że w moich danych nie ma żadnych braków i są one kompletne.

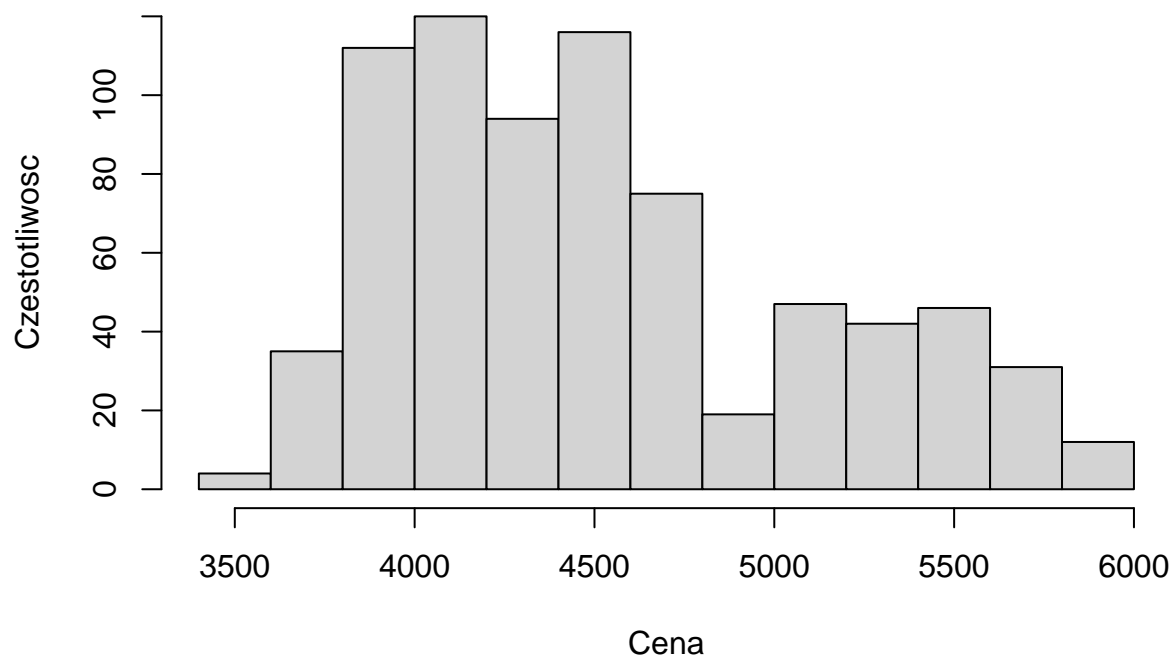
Wykresy danych

Najpierw robię wykres ceny od czasu:



Oraz wyświetlam histogram tych danych:

Histogram cen zamknięcia S&P 500



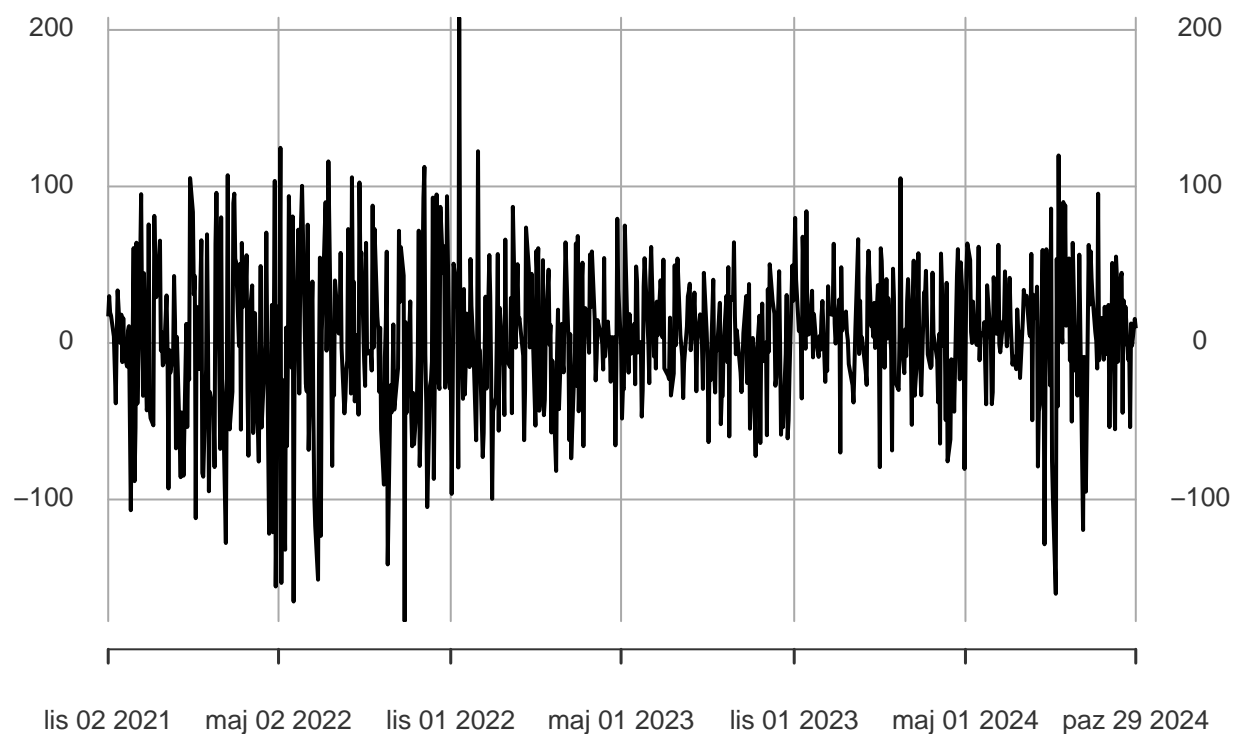
Histogram pokazuje, że rozkład danych nie jest normalny, posiada jedną modę (nie jest bimodalny) oraz że jest prawoskośny. Oznacza to, że o wiele więcej wartości rozkładu znajduje się poniżej średniej, oraz że średnia będzie większa od mediany. Potwierdza to wykres ceny do czasu, gdzie widać jak wartości poszybowały do góry od 2023 roku. Również na górnym wykresie nie rzucają się w oczy żadne wartości odstające.

Różnicowanie szeregu

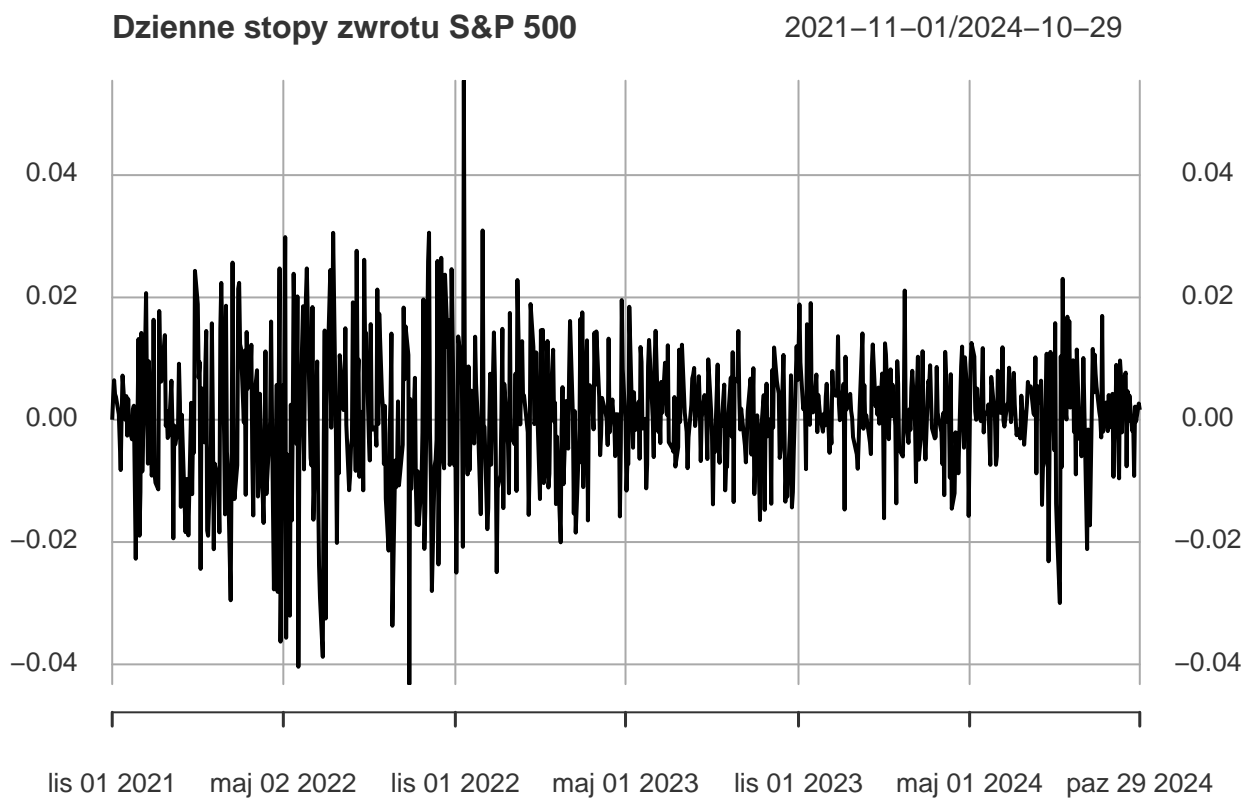
By móc wyliczyć miarę VaR dla moich danych, potrzebuje wpierw obliczyć potencjalny zysk lub stratę z inwestowania w S&P 500. Dlatego różnicuję mój szereg, i wyznaczam dzienną różnicę między cenami indeksu, na podstawie których będę wyliczać potencjalną stratę:

Zróznicowane ceny S&P 500

2021-11-02/2024-10-29



Analogicznie wyliczam dzienne stopy zwrotu cen zamknięcia S&P 500, by móc VaR wyrażać w procentach:



Wartości historyczne

Pierwszą metodą policzenia miary VaR będzie wyliczenie jej jako kwantyla empirycznego z danych.

VaR jako wartość

Dla $\alpha=0.05$:

```
##          5%
## -78.73212
```

Dla $\alpha=0.01$:

```
##          1%
## -130.3202
```

VaR jako procent

Dla $\alpha=0.05$:

```
##          5%
## "-1.75%"
```

Dla $\alpha=0.01$:

```
##          1%
## "-3.10%"
```

Dopasowanie rozkładu normalnego

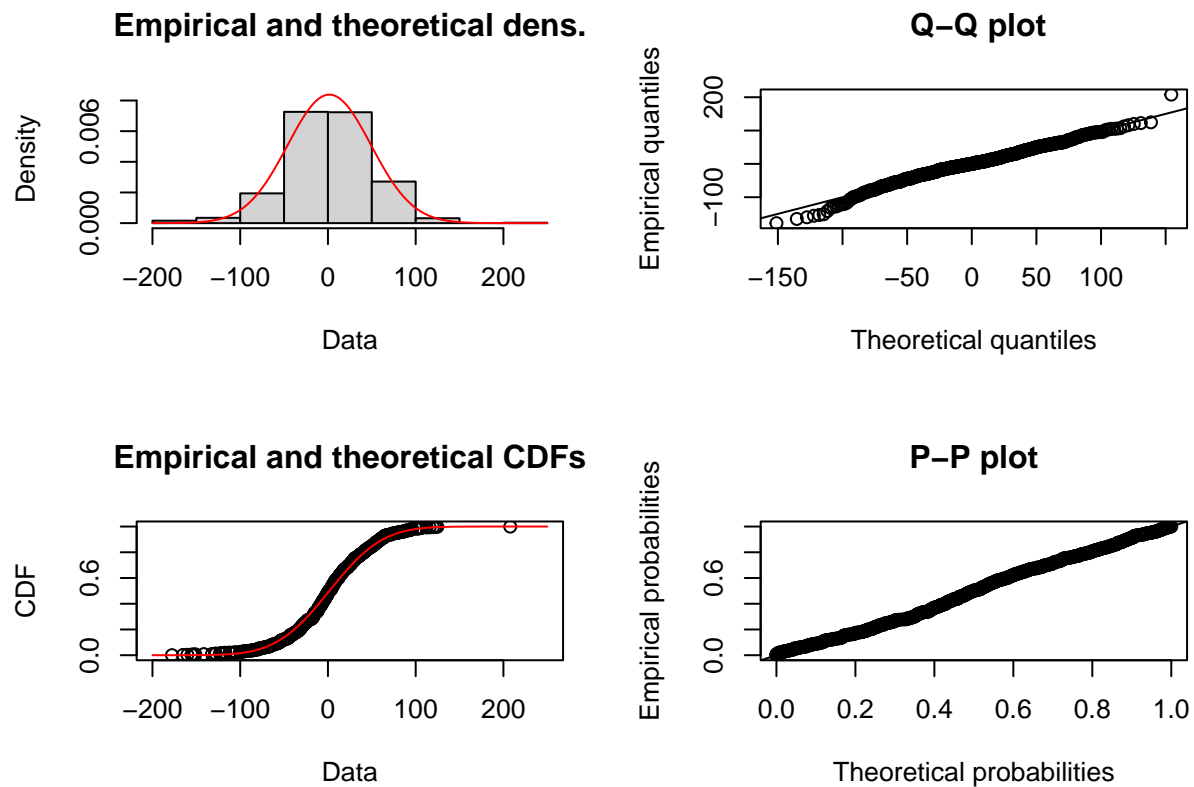
Drugim sposobem będzie dopasowanie rozkładu normalnego do moich danych, i wyznaczenie VaR jako kwantyla tego rozkładu normalnego.

VaR jako wartość

Parametry dopasowanego rozkładu normalnego:

```
##      mean      sd
## 1.621343 47.585788
```

Jego wykresy diagnostyczne:



I wartości VaR-u dla dopasowanego rozkładu normalnego:

- $\alpha=0.05$:

```
## [1] -76.65031
```

- $\alpha=0.01$:

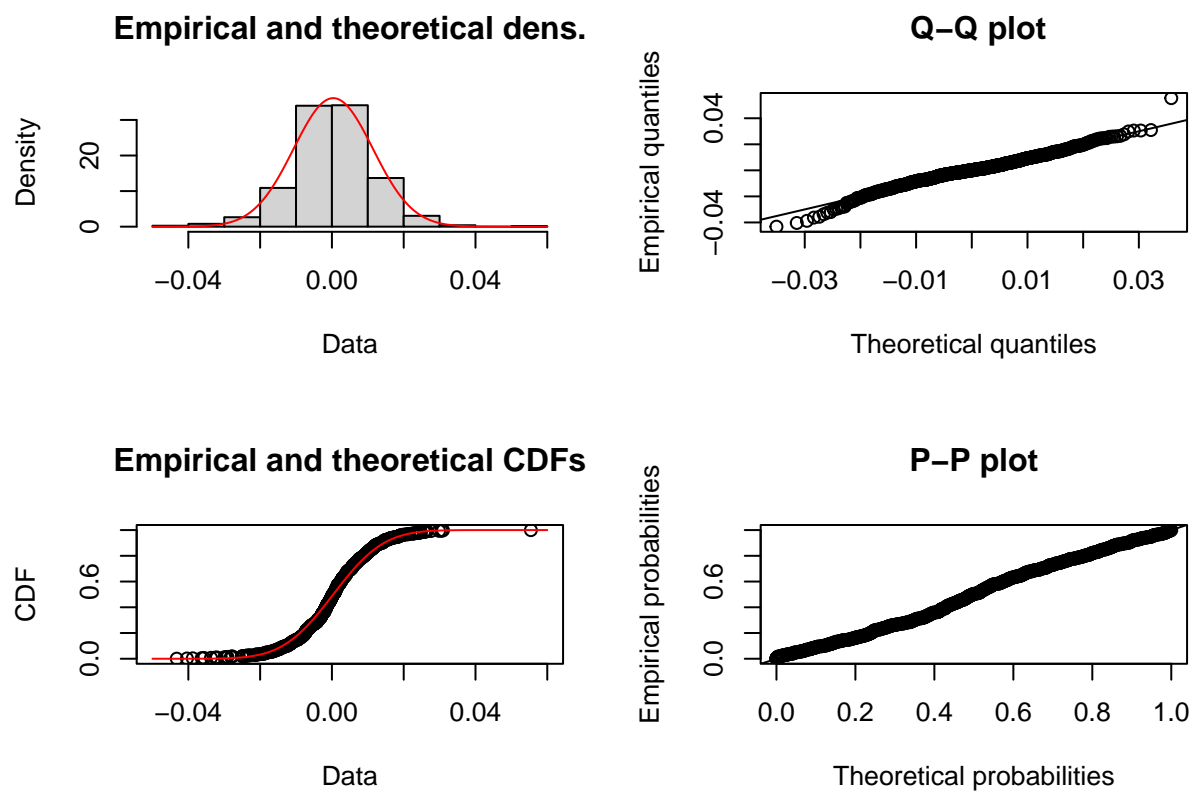
```
## [1] -109.0798
```

VaR jako procent

Analogicznie wpierw parametry dopasowanego rozkładu:

```
##          mean          sd
## 0.0003725124 0.0110445149
```

Następnie wykresy diagnostyczne:



I na koniec wartości VaR:

- $\alpha=0.05$:

```
## [1] "-1.78%"
```

- $\alpha=0.01$:

```
## [1] "-2.53%"
```

Wiarygodność tej metody

Problemem z powyższymi wynikami może być to, że rozkład normalny dopasowujemy na siłę do danych, które niekoniecznie mogą reprezentować taki rozkład. Takie usilne dopasowanie może czasami prowadzić do błędnych wyników. Idealnym pokazem tego jest zastosowanie testu statycznego, który oceni nam dopasowanie naszych danych do rozkładu normalnego. Ja użyję testu Shapiro-Wilka, który ma hipotezę zerową o przynależności naszych danych do rozkładu normalnego:

Wpierw dla danych wartościowych:

```
##  
## Shapiro-Wilk normality test  
##  
## data:  as.numeric(differ)  
## W = 0.98688, p-value = 2.86e-06
```

I od razu dla danych procentowych:

```
##  
## Shapiro-Wilk normality test  
##  
## data:  as.numeric(simple_returns)  
## W = 0.98167, p-value = 4.205e-08
```

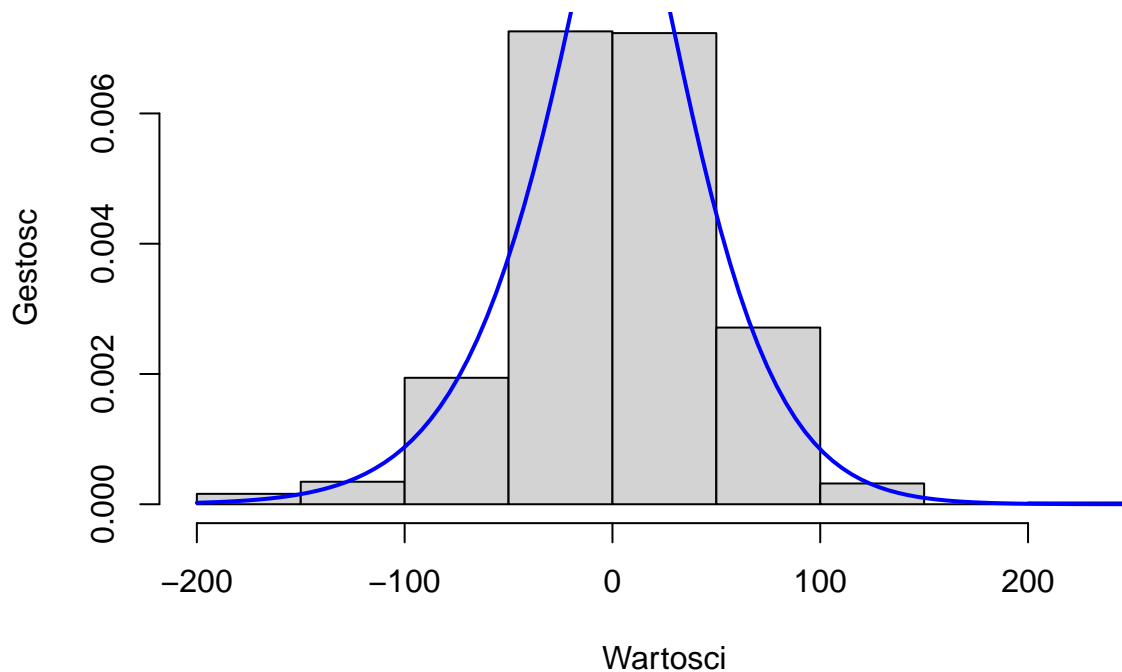
W obu przypadkach p-value jest bardzo małe, co mocno sugeruje, że hipoteza zerowa jest błędna i nasze dane nie reprezentują rozkładu normalnego, zatem można poddawać pod wątpliwość jakość wyliczonego w ten sposób VaR-u.

Dopasowanie dowolnego rozkładu

Trzecim sposobem jest dopasowanie dowolnego rozkładu do naszych danych, i wyznaczenie VaR-u jako kwantyla tegoż rozkładu.

VaR jako wartość

Dopasowuje rozkład do danych o wartościach i patrze na wykres dopasowania rozkładu na histogramie:



Rozkład wygląda na dość dobrze dopasowany. Sprawdzam szczegóły jaki rozkład został tutaj dopasowany:

```
## [1] "SEP4"                                "skew exponential power type 4"

## $mu
## [1] 2.892599
##
## $sigma
## [1] 52.05922
##
## $nu
## [1] 1.336366
##
## $tau
## [1] 1.486594
```

I wyliczam VaR jako kwantyle tego rozkładu:

- $\alpha=0.05$:

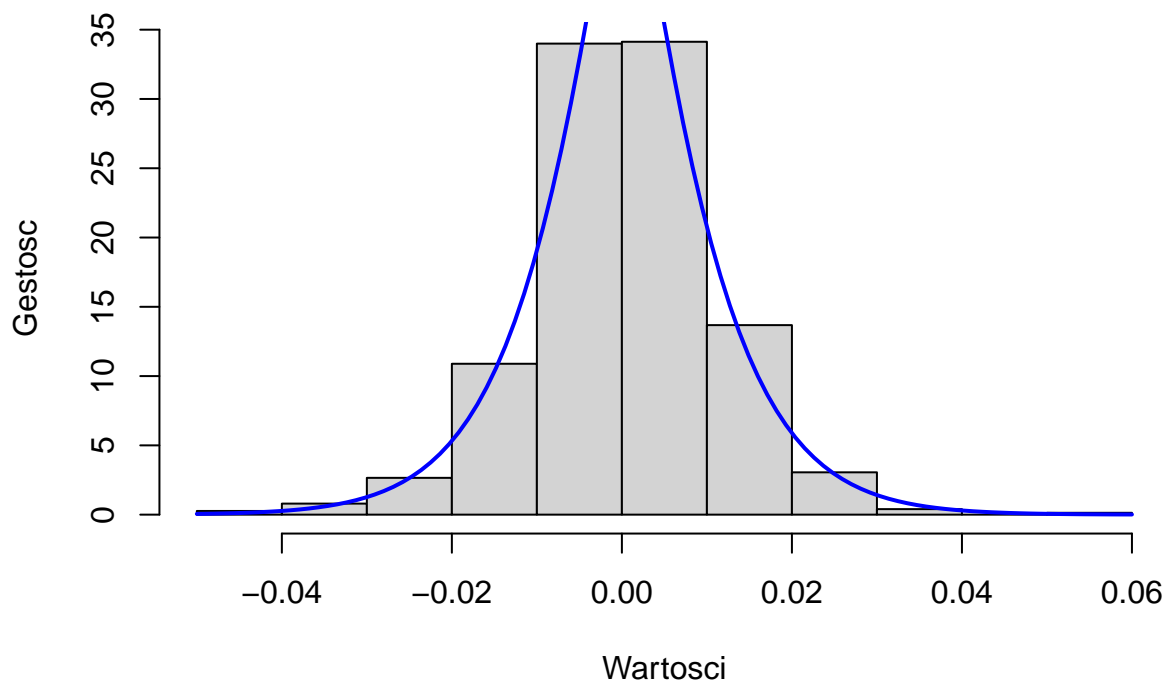
```
## [1] -79.81331
```

- $\alpha=0.01$:

```
## [1] -125.7982
```

VaR jako procent

Analogicznie postępuje dla danych procentowych, najpierw wykres dopasowania:



Ponownie wygląda on dobrze, więc sprawdzam szczegóły rozkładu:

```
## [1] "PE" "Power Exponential"
```

```
## $mu  
## [1] 0.0003712494  
##  
## $sigma  
## [1] 0.01104842  
##  
## $nu  
## [1] 1.234227
```

I wyliczam VaR jako kwantyle tego rozkładu:

- $\alpha = 0.05$:

```
## [1] "-1.78%"
```

- $\alpha = 0.01$:

```
## [1] "-2.86%"
```

Porównanie wyników

Wartościowych z $\alpha=0.05$:

Metoda historyczna:

```
##          5%  
## -78.73212
```

Metoda z rozkładem normalnym:

```
## [1] -76.65031
```

Metoda z dopasowaniem rozkładu:

```
## [1] -79.81331
```

Wartościowych z $\alpha=0.01$:

Metoda historyczna:

```
##          1%  
## -130.3202
```

Metoda z rozkładem normalnym:

```
## [1] -109.0798
```

Metoda z dopasowaniem rozkładu:

```
## [1] -125.7982
```

Procentowych z $\alpha=0.05$:

Metoda historyczna:

```
##          5%  
## "-1.75%"
```

Metoda z rozkładem normalnym:

```
## [1] "-1.78%"
```

Metoda z dopasowaniem rozkładu:

```
## [1] "-1.78%"
```

Procentowych z $\alpha=0.01$:

Metoda historyczna:

```
##          1%  
## "-3.10%"
```

Metoda z rozkładem normalnym:

```
## [1] "-2.53%"
```

Metoda z dopasowaniem rozkładu:

```
## [1] "-2.86%"
```

Interpretacja wyników

W wynikach mamy bardzo duży przekrój zależności, niektóre wyniki się pokrywają, a niektóre różnią się o naprawdę spore wartości. Najważniejsze jest jednak fakt, że w każdym przypadku VaR 99-procentowy wyznaczony dowolną metodą przekraczał wartością wszystkie VaR-y 95-procentowe, co daje nadzieje na dobre jakościowe wyniki. I dla wartości i dla procentów różnice między różnymi metodami dla VaR-u 95% są bardzo małe, a dla VaR-u 99% są o wiele większe, co też brzmi logicznie gdyż im dalszy ogon tym ciężiej go dokładnie wymodelować. Z wyników wydają się, że metoda z rozkładem normalnym daje “najdziwniejsze” i najbardziej niedoszacowane w porównaniu do reszty wyniki, poza jednym wyjątkiem dla wartości procentowych VaR-u 95 gdzie mamy jedyne pokrycie wartości.