

# Złoto i srebro

Julia Wołk-Łaniewska, Agata Żabska

8 maja 2023

## 1 Wstęp

W niniejszym projekcie zajmiemy się analizą danych dotyczących cen złota i srebra w kolejnych dniach od 30 grudnia 1977 do 31 grudnia 2012, to jest 9132 dni. Dane zostały zaczerpnięte ze strony: <https://vincentarelbundock.github.io/Rdatasets/doc/AER/GoldSilver.html>.

Pojedynczy wiersz zawiera:

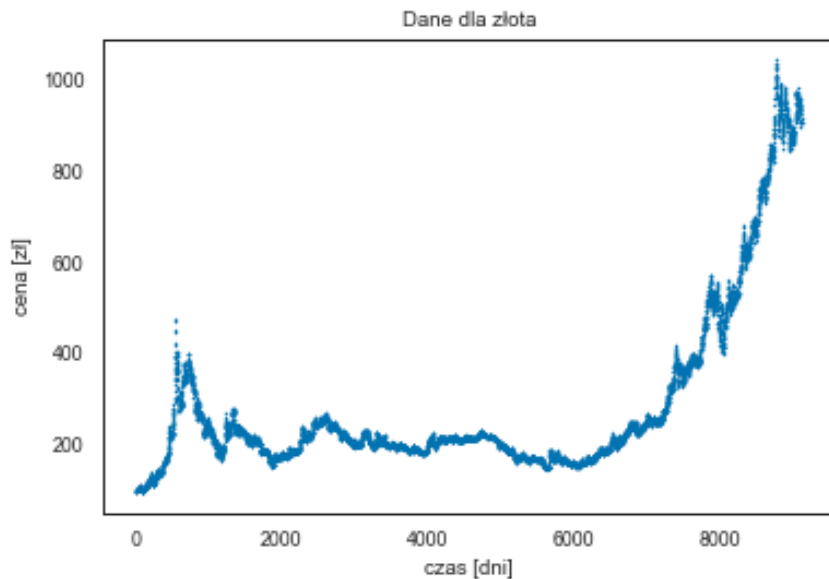
- datę,
- cenę złota [zl],
- cenę srebra [zl].

Poniżej zajmiemy się analizą statystyczną cen złota i srebra jako odrębnych zmiennych losowych, a następnie ich analizą porównawczą. Wszystkie obliczenia oraz wykresy zostały wykonane w *Pythonie*.

## 2 Analiza danych

### 2.1 Ceny złota

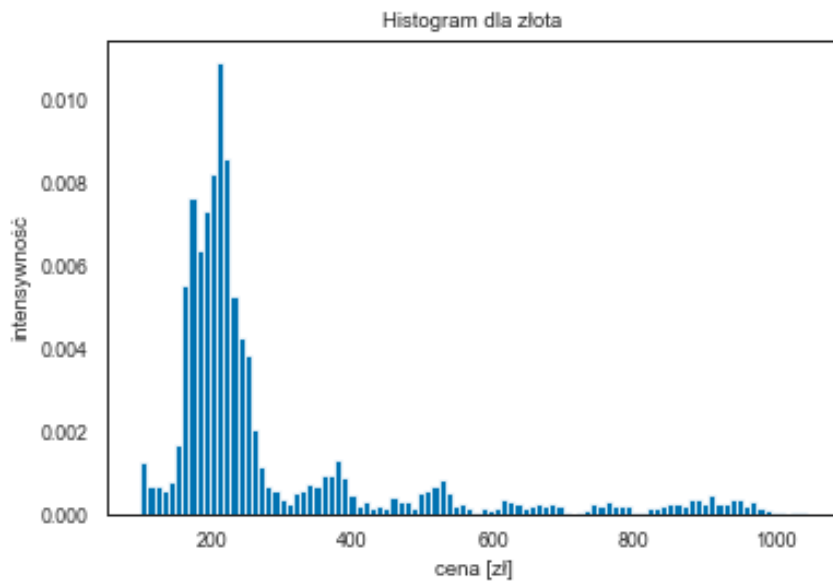
W tej części zajmiemy się analizą cen złota. Na poniższym wykresie (Rysunek 1) umieszone zostały ceny złota w kolejnych dniach obserwacji.



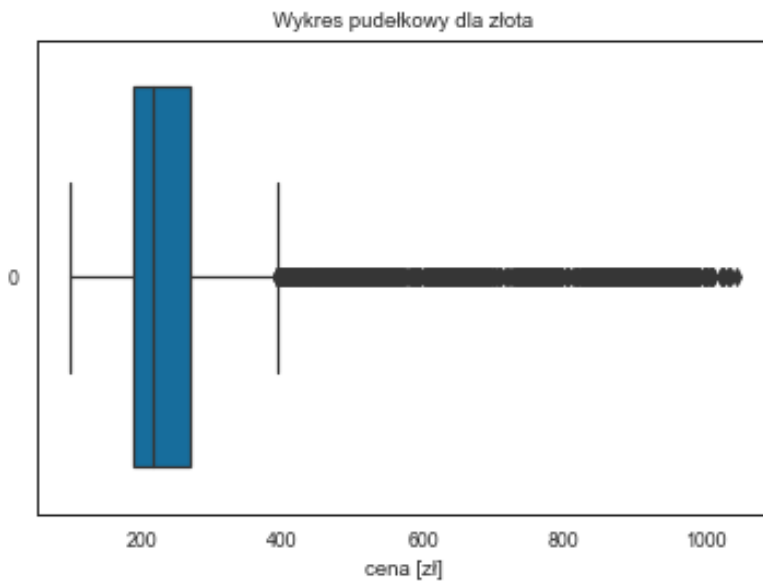
Rysunek 1: Ceny złota.

Jak możemy zauważyć, ceny w pierwszej części pomiarów rosły. Następnie zaliczyły spadek i przez pewien okres utrzymywały się na stosunkowo stabilnym poziomie. W ostatniej części ceny rosły aż do osiągnięcia wartości maksymalnych.

Sporządziliśmy również histogram (Rysunek 2) oraz wykres pudełkowy (Rysunek 3).



Rysunek 2: Histogram cen złota.



Rysunek 3: Wykres pudełkowy cen złota.

Widzimy, że zarówno histogram, jak i wykres pudełkowy, potwierdzają dokonane przez nas obserwacje. Na histogramie możemy zauważyć, że najwięcej wartości oscyluje w okolicy 200 zł. Zdecydowanie mniej danych możemy zaobserwować na przedziale [0zł, 200zł] oraz w okolicy 400 zł i 560 zł. Pojawiają się tam jednak pewne zagęszczenia. Histogram nie maleje jednostajnie, zamiast tego obserwujemy cykliczne wzrosty i spadki. Pokrywa się to z wykresem cen złota w kolejnych dniach (Rysunek 1). Rozważając wykres pudełkowy, możemy dojść do podobnych wniosków. Mianowicie, że najwięcej wartości znajduje się w przedziale od 200 zł do 280 zł. Pozostałe oscylują między 100 zł i 400 zł. Występuje również dużo wartości odstających, które osiągają nawet około 1000 zł.

Na podstawie zebranych danych obliczyliśmy podstawowe statystyki. Ich wartości znajdują się w tabeli poniżej (Tabela 1).

Charakterystyka	Wartość
Wartość minimalna	97.45 zł
Wartość maksymalna	1043.89 zł
Dolny kwartył (Q1)	186.38 zł
Mediana (Q2)	215.155 zł
Górny kwartył (Q3)	268.3725 zł
Rozstęp międzykwartyłowy	81.9925 zł
Rozstęp	946.44 zł
Wariancja	36228.291311378925 zł <sup>2</sup>
Odchylenie standardowe	190.33730929951417 zł
Kurtoza (dla r.norm. = 3)	7.067217820965545
Współczynnik skośności	2.191654800817666
Współczynnik zmienności	66.23364610030305%

Tabela 1: Charakterystyki liczbowe dla złota.

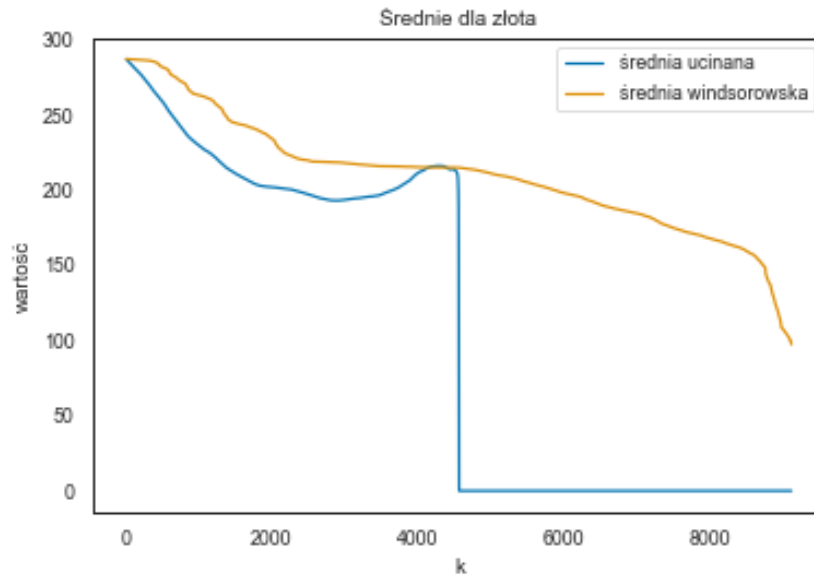
Otrzymane wyniki potwierdzają nasze ogólne rozważania. Wartość minimalna i maksymalna wynoszą odpowiednio około 100 zł i 1000 zł, co ma przełożenie na rozstęp, który jest duży. Kwantyle, w tym mediana wynosząca 215.155 zł, są skupione wokół wartości 200 zł, stąd rozstęp międzykwartyłowy jest niewielki. Dane są jednak mocno rozproszone. Możemy tak wnioskować na podstawie wariancji, a także odchylenia standardowego, które przyjmują duże wartości. Z kolei kurtoza jest większa niż dla rozkładu normalnego. Sugeruje to, że dane mają tendencję do dużych odchyleń. Na podstawie współczynnika skośności możemy stwierdzić, że rozkład jest prawostronnie skośny. Współczynnik zmienności wynosi 66.23364610030305% i informuje nas o silnej zmienności.

Obliczyliśmy wartości średnich: arytmetycznej, geometrycznej i harmonicznej (Tabela 2) oraz sporządziliśmy wykres zależności wartości średniej ucinanej i windsorowskiej od ilości zaniechanych danych (Rysunek 4). Wyniki tych obliczeń umieściliśmy poniżej:

Średnia	Wartość [zł]
Arytmetyczna	287.37253723171267
Geometryczna	249.2936938697913
Harmoniczna	226.5546747651117

Tabela 2: Średnie dla złota.

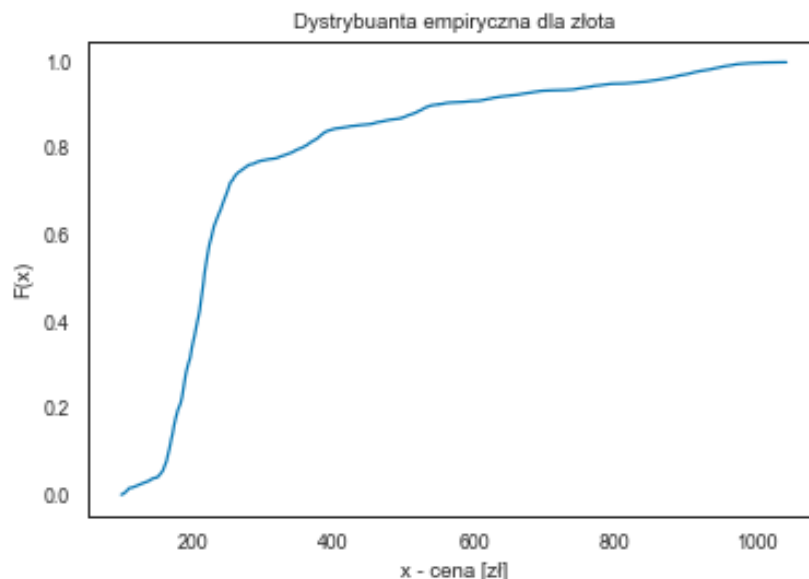
Jak możemy zauważyć, średnie oscylują w okolicy 250 zł. Średnia arytmetyczna, która jest najbardziej wrażliwa na dane odstające, ma zdecydowanie najwyższą wartość. Średnia harmoniczna, która z kolei jest najmniej na to podatna, ma wartość najbardziej zbliżoną do mediany.



Rysunek 4: Średnie dla złota.

Wartości średniej ucinanej i windsorowskiej generalnie maleją. Ciekawe zachowanie przejawia średnia ucinana, która na początku maleje, a następnie nieco wzrasta. Przyczynę możemy odnaleźć na wykresie cen złota w czasie (Rysunek 1). Początkowo, ucinając mało danych, zaniedbujemy głównie wartości wyższe od mediany, przez co średnia ucinana maleje. Gdy jednak zaczynamy pomijać również niższe wartości (przy ciągłej obecności większych w końcowej fazie obserwacji), osiągamy wzrost. Po przekroczeniu  $k = 4000$ , średnia ucinana pozostaje na poziomie 0. Jest to spowodowane ucięciem wszystkich danych. Średnia windsorowska zachowuje się bardziej przewidywalnie. Na całym swoim przebiegu szybciej bądź wolniej maleje. Możemy zauważyć charakterystyczne wypukłości, o których wspominaliśmy wcześniej.

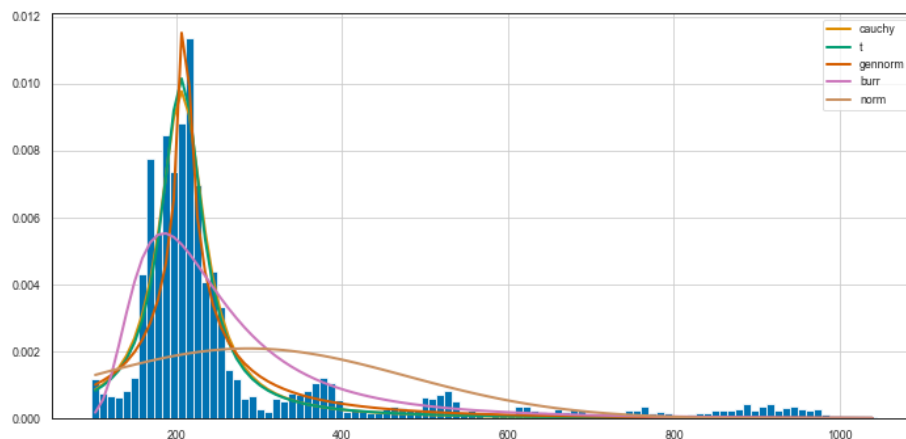
Następnie za pomocą funkcji *ECDF* wygenerowaliśmy wykres dystrybuanty empirycznej (Rysunek 5).



Rysunek 5: Dystrybuanta empiryczna cen złota.

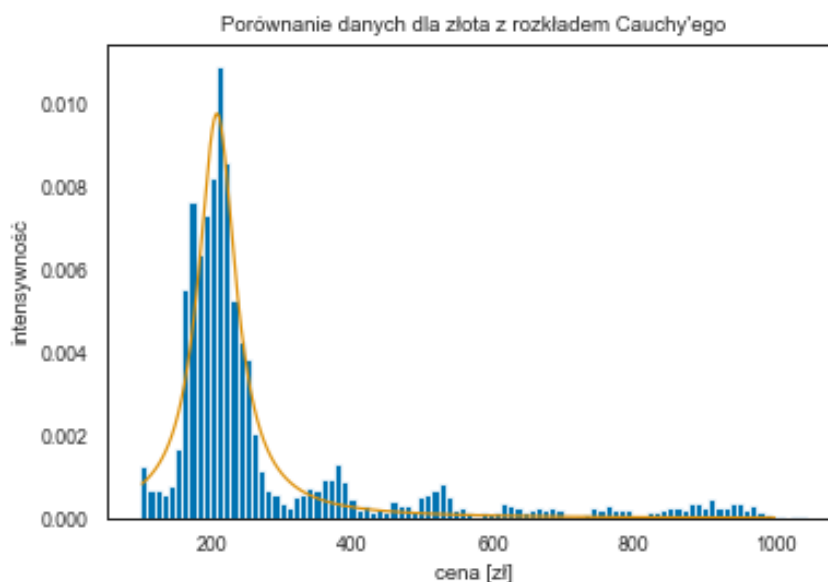
Dystrybucja w pierwszej fazie bardzo szybko rośnie, ponieważ prawdopodobieństwo wylosowania wartości do około 270 zł jest wysokie. W dalszej części wzrost zdecydowanie spowalnia, jest to spowodowane mniejszą ilością wartości wysokich w porównaniu ze średnimi.

Dalej, za pomocą biblioteki *fitter*, szukaliśmy rozkładu, który przybliży nasze dane. Pod uwagę braliśmy wiele rozkładów, jednak najbardziej interesujące okazały się: uogólniony rozkład Gaussa (tu: gennorm), t-Studenta, Burr'a oraz Cauchy'ego. Dla porównania uwzględniony został również rozkład normalny. Na poniższym wykresie (Rysunek 6) możemy zaobserwować wszystkie wymienione wyżej rozkłady w porównaniu z histogramem dla danych.



Rysunek 6: Gęstości rozkładów zmiennych losowych vs histogram cen złota.

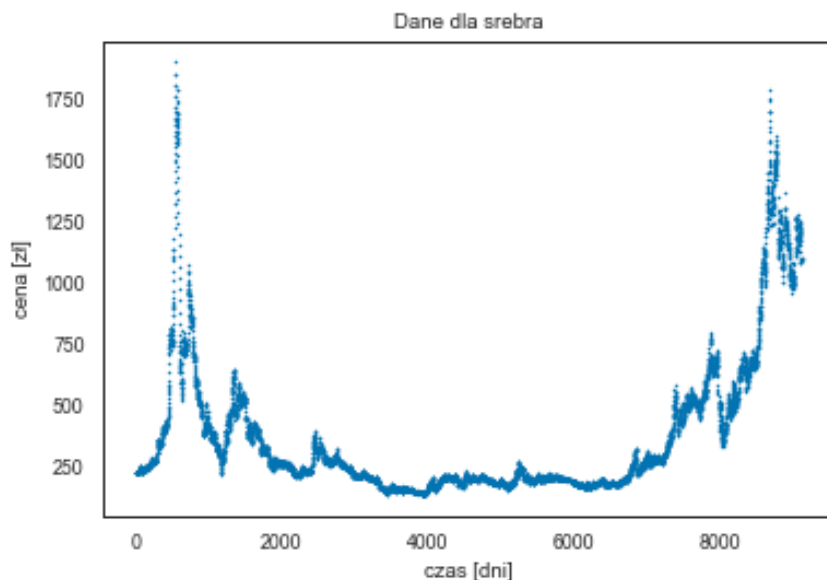
Do wyznaczenia rozkładu, który najdokładniej przybliży nasze dane, wykorzystaliśmy dwie metody: SSE (sumy błędów kwadratowych) oraz BIC (Bayesowskie kryterium informacyjne Schwartz). Obydwie wykazały, że najbardziej odpowiednią aproksymacją jest rozkład Cauchy'ego z parametrem położenia  $x_0 = 206.5577803504877$  oraz parametrem skali  $\gamma = 32.57608210316859$ . Poniżej znajduje się porównanie histogramu cen złota z gęstością rozkładu Cauchy'ego (Rysunek 7).



Rysunek 7: Rozkład Cauchy'ego vs histogram cen złota.

## 2.2 Ceny srebra

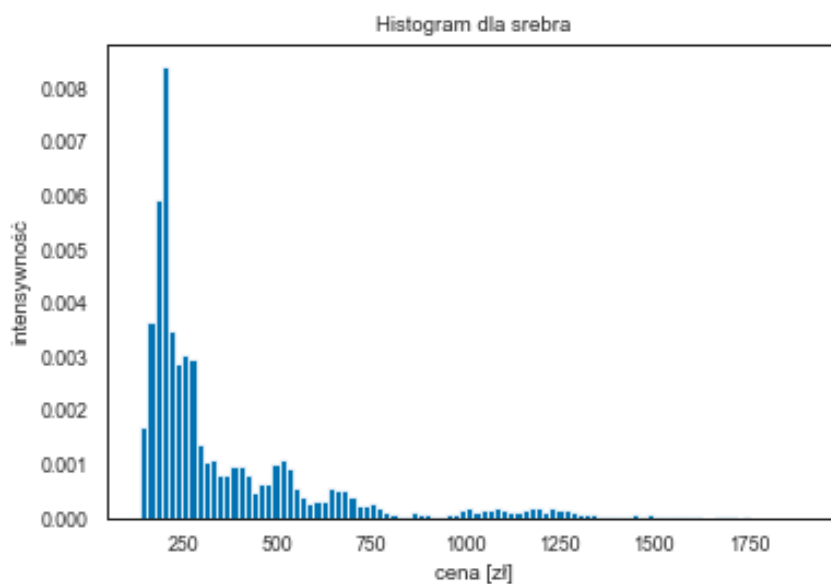
W tej części przeprowadzimy analogiczną analizę cen srebra. Na poniższym wykresie (Rysunek 8) umieszczone zostały ceny srebra w kolejnych dniach obserwacji. .



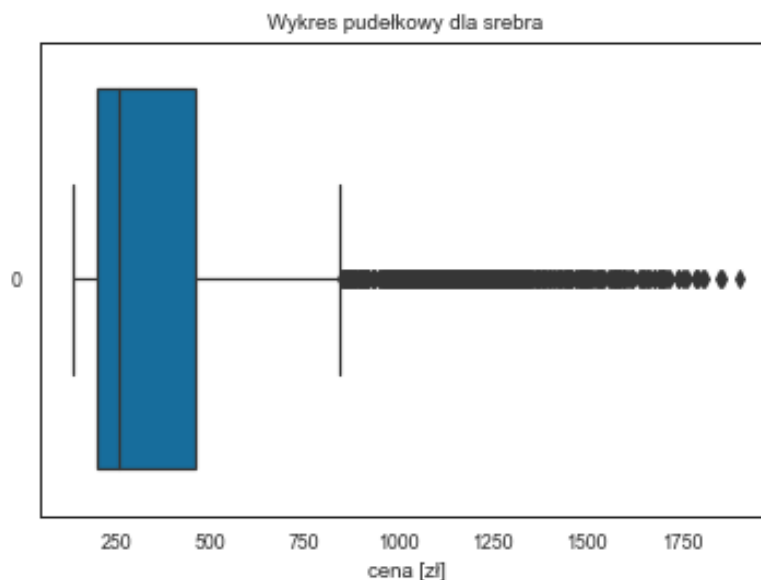
Rysunek 8: Ceny srebra.

Możemy zauważyć, że srebro zarówno w początkowej, jak i końcowej fazie przyjmuje znacznie wyższe wartości od pozostałych pomiarów. Widzimy, że wartości maleją w kierunku środka. Pojawiają się charakterystyczne wzrosty i spadki, jednak jest ich mniej niż w przypadku złota. Ciekawą obserwacją może być fakt, że rozważany wykres jest stosunkowo symetryczny.

Sporządziliśmy również histogram (Rysunek 9) oraz wykres pudełkowy (Rysunek 10).



Rysunek 9: Histogram cen srebra.



Rysunek 10: Wykres pudełkowy cen srebra.

Analizując histogram możemy zauważyć, że najwięcej danych przyjmuje wartości w okolicy 200 zł, następnie histogram maleje, ponieważ im wyższe są wartości, tym mniej danych je przyjmuje. Widzimy jednak pewne zagęszczenia, które odpowiadają charakterystycznym wzrostom i spadkom obserwowanym na wykresie cen srebra w czasie (Rysunek 8).

W przypadku srebra wykres pudełkowy również potwierdza nasze wstępne obserwacje. Najwięcej danych mieści się w przedziale od około 200 zł do około 500 zł. Mediana znajduje się w okolicy wartości 250 zł. Jak możemy zaobserwować, odstęp pomiędzy Q3 a górnym wąsem wykresu pudełkowego jest duży. Szczególnie w porównaniu z odstępem pomiędzy Q1 a wartością minimalną. Obserwujemy również szeroki przedział danych odstających, których wartości dochodzą do około 1900 zł.

Na podstawie zebranych danych obliczyliśmy podstawowe statystyki, ich wartości umieściliśmy w tabeli poniżej (Tabela 3).

Charakterystyka	Wartość
Wartość minimalna	137.3 zł
Wartość maksymalna	1901.93 zł
Kwartyl dolny (Q1)	197.7775 zł
Mediana (Q2)	255.36 zł
Kwartyl górny (Q3)	459.195 zł
Rozstęp międzykwartylowy	261.4175 zł
Rozstęp	1764.63 zł
Wariancja	85154.492019371285 zł <sup>2</sup>
Odchylenie standardowe	291.81242608801165 zł
Kurtoza (dla r.norm. = 3)	7.697812750318201
Współczynnik skośności	2.1686087235417357
Współczynnik zmienności	77.47511619273328%

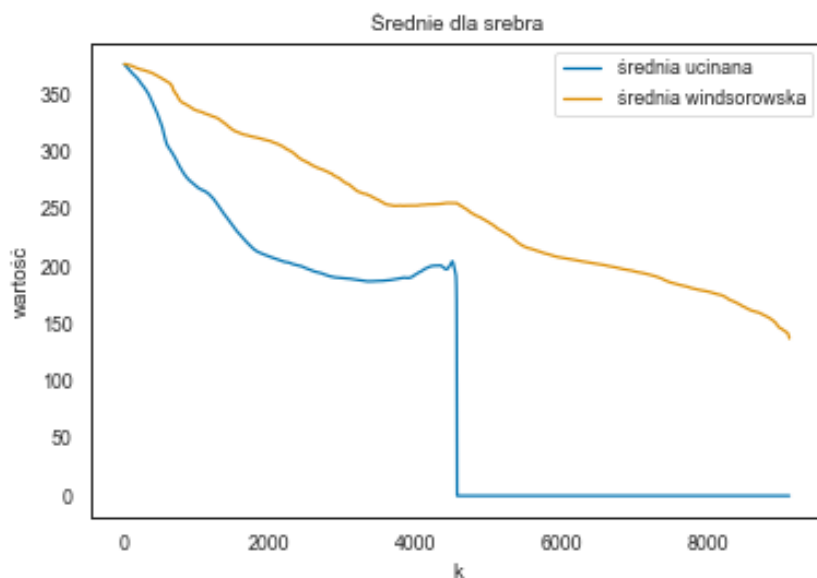
Tabela 3: Charakterystyki liczbowe dla srebra.

Analizując otrzymane wyniki, widzimy ich pokrycie z wcześniejszymi rozważaniami. Minimalna wartość wynosi 137.3 zł, z kolei maksymalna to aż 1901.93 zł. Odzwierciedlenie tego jest widoczne w rozstępie, który jest równy 1764.63 zł oraz w wariancji i odchyleniu standardowym, które przyjmują bardzo wysokie wartości. Mediana to około 250 zł, co pokrywa się z informacjami odczytanymi z wykresu pudełkowego. Odległość między kwartylami, a co za tym idzie- rozstęp międzykwartylowy, jest dużo mniejsza niż rozstęp. W tym przypadku kurtoza również przekracza wartość dla rozkładu normalnego, co wskazuje na skoncentrowanie danych wokół średniej i większą szansę na pojawienie się ogonów. Współczynnik zmienności, który wynosi około 77%, wskazuje na silną zmienność cen srebra. Z kolei współczynnik skośności informuje nas o tym, że rozkład jest prawostronnie skośny.

Natępnie wyznaczyliśmy wartości średnich: arytmetycznej, geometrycznej i harmonicznej (Tabela 4) oraz sporządziliśmy wykres zależności wartości średniej ucinanej i windsorowskiej od ilości zaniechanych danych (Rysunek 11).

Średnia	Wartość [zł]
Arytmetyczna	376.6530989925536
Geometryczna	307.77947590290717
Harmoniczna	267.5771378547969

Tabela 4: Średnie dla srebra.



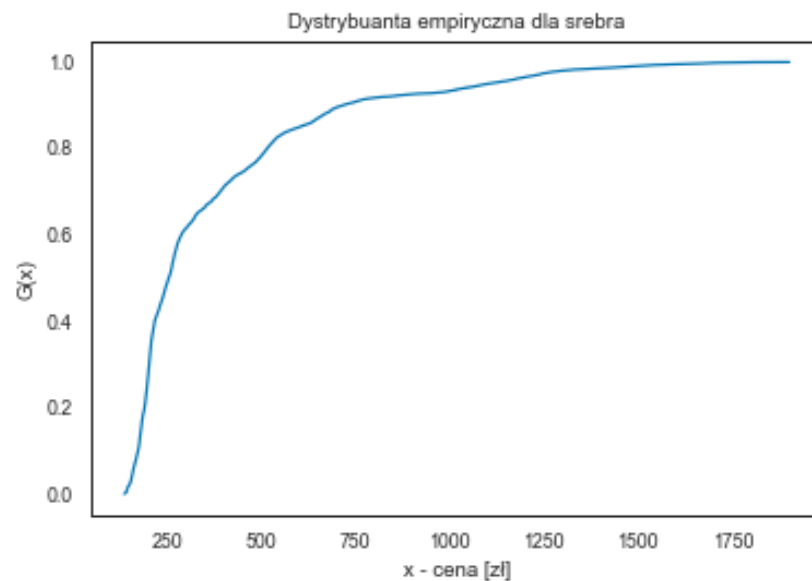
Rysunek 11: Średnie dla srebra.

Rozważając otrzymane rezultaty, możemy zauważyć, że ich wartości są wyższe niż mediana. Szczególnie znacząco jest to zauważalne w przypadku średniej arytmetycznej, która jest najbardziej podatna na dane odstające. Natomiast średnia harmoniczna jest zbliżona do mediany.

Wykres średniej ucinanej i winsorowskiej dla srebra prezentuje podobne zachowanie, jak w przypadku złota. Wartości średniej ucinanej początkowo maleją, ponieważ zaniechujemy odstające wartości, następnie rosną. Jest to jednak mniejszy wzrost niż mogliśmy obserwować podczas analizy przeprowadzonej dla poprzedniego surowca. Powodem tego może być większa symetria danych.



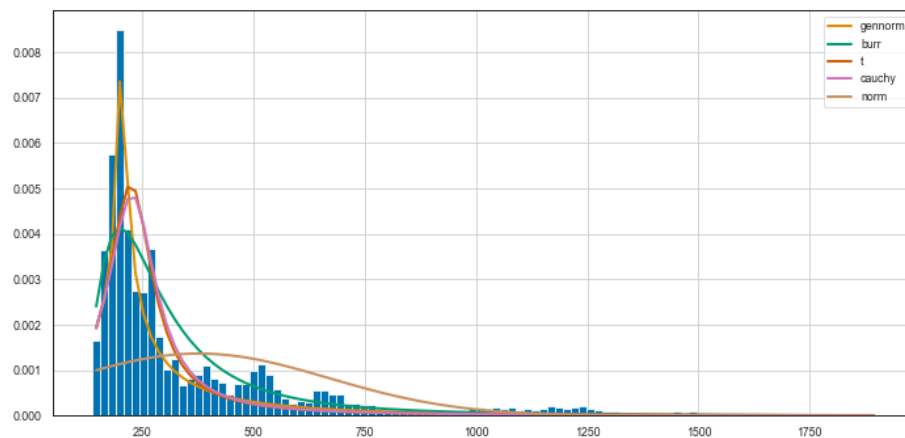
Następnie za pomocą funkcji *ECDF* wygenerowaliśmy dystrybuantę empiryczną (Rysunek 12).



Rysunek 12: Dystrybuanta empiryczna cen srebra.

Przeglądając się przebiegowi dystrybuanty, możemy zauważyć, że w początkowej fazie dość intensywnie rośnie, aby później spowolnić swój wzrost. Jednak to spowolnienie następuje względnie jednostajnie.

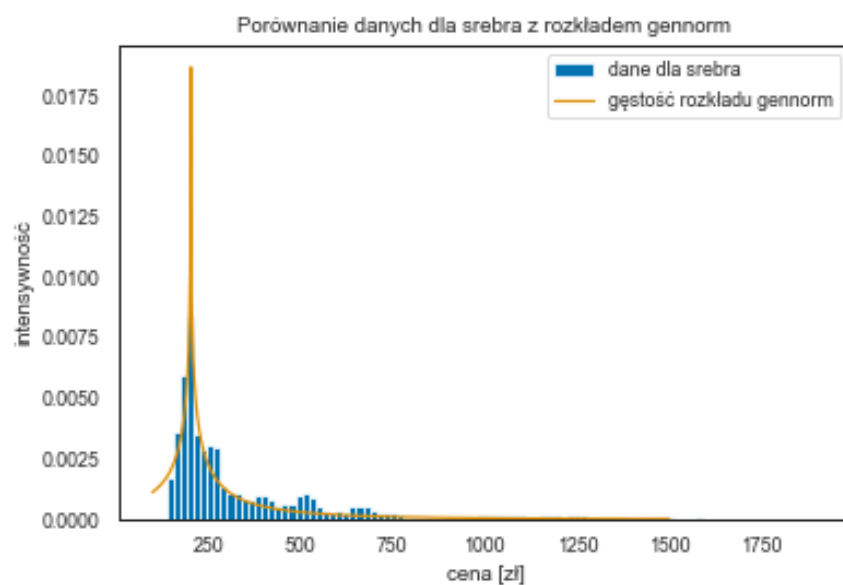
Ponownie za pomocą biblioteki *fitter* dopasowywaliśmy znane rozkłady zmiennych losowych do naszych danych. Pod uwagę braliśmy rozkłady: normalny, gennormalny, t-Studenta, Burr'a oraz Cauchy'ego. Na poniższym wykresie (Rysunek 13) widnieją wszystkie wymienione rozkłady w porównaniu z histogramem cen srebra.



Rysunek 13: Gęstości rozkładów zmiennych losowych vs histogram cen srebra.

Metody SSE oraz BIC ponownie okazały się zgodne i wykazały, że w tym przypadku najdokładniejszą aproksymacją będzie uogólniony rozkład Gaussa z parametrem położenia  $\mu = 204.25999999996662$ , parametrem skali  $\alpha = 5.861847066773678$  oraz parametrem kształtu  $\beta = 0.362872297970977$ .

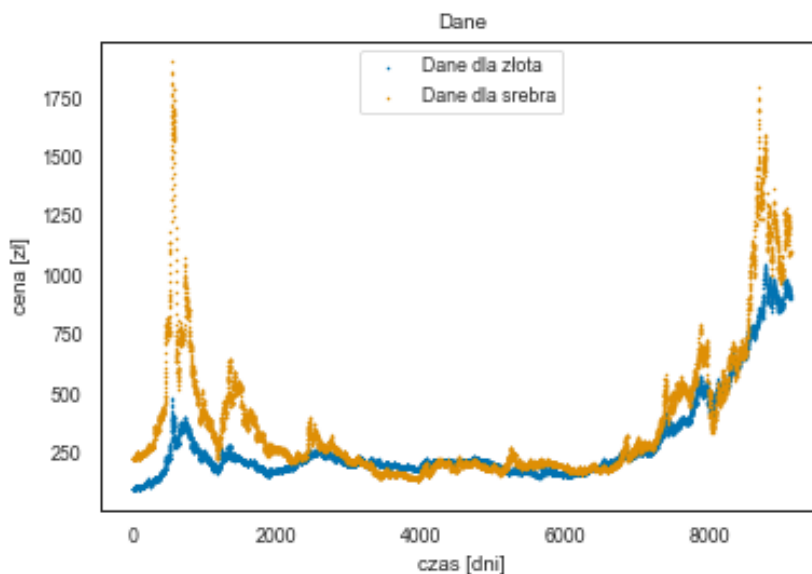
Na wykresie (Rysunek 14) widzimy porównanie histogramu do gęstości wyżej wspomnianego rozkładu.



Rysunek 14: Gęstość uogólnionego rozkładu Gaussa vs histogram cen srebra

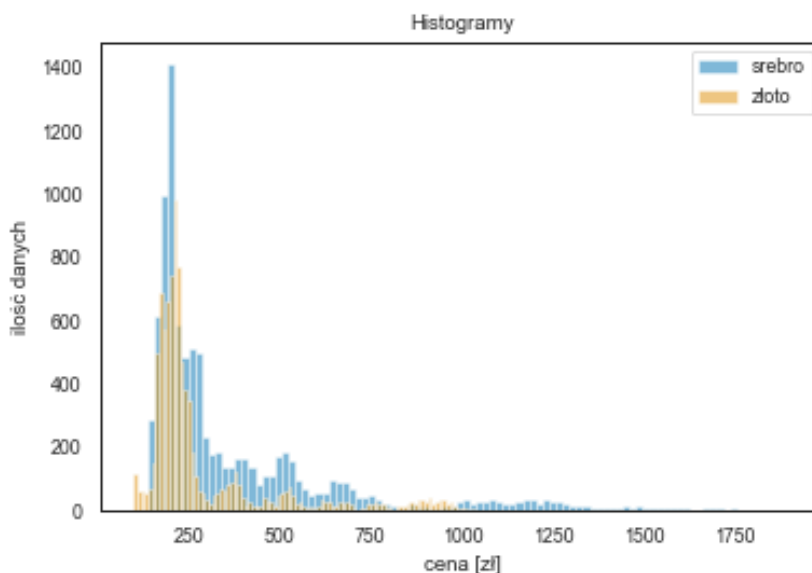
### 3 Analiza porównawcza

Kolejny etap naszych rozważań to analiza porównawcza. Zestawiliśmy ze sobą ceny surowców w czasie (Rysunek 15), ich wykresy pudełkowe (Rysunek 16), histogramy (Rysunek 17) oraz dystrybuanty empiryczne (Rysunek 18). Umieściliśmy je poniżej.



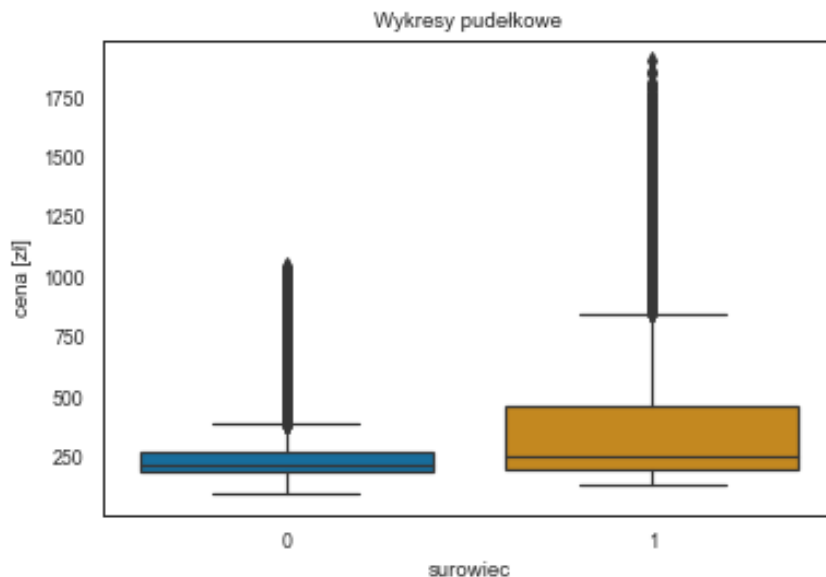
Rysunek 15: Ceny surowców.

Analizując wykresy cen surowców możemy zauważyć, że wartości dla srebra są zazwyczaj większe. W początkowej fazie ceny złota przewyższają nieco swoją medianę, natomiast ceny srebra osiągają swoje maksimum. Następnie obie serie danych przez dłuższy czas oscylują wokół zbliżonych wartości, po czym zaliczają znaczący wzrost dla ostatniego etapu obserwacji. Wahania cen srebra są zdecydowanie bardziej zauważalne. Widzimy charakterystyczne wzrosty i spadki, które mają dużo większe amplitudy niż w przypadku złota. Wykres cen srebra jest stosunkowo symetryczny w przeciwieństwie do wykresu cen drugiego surowca.



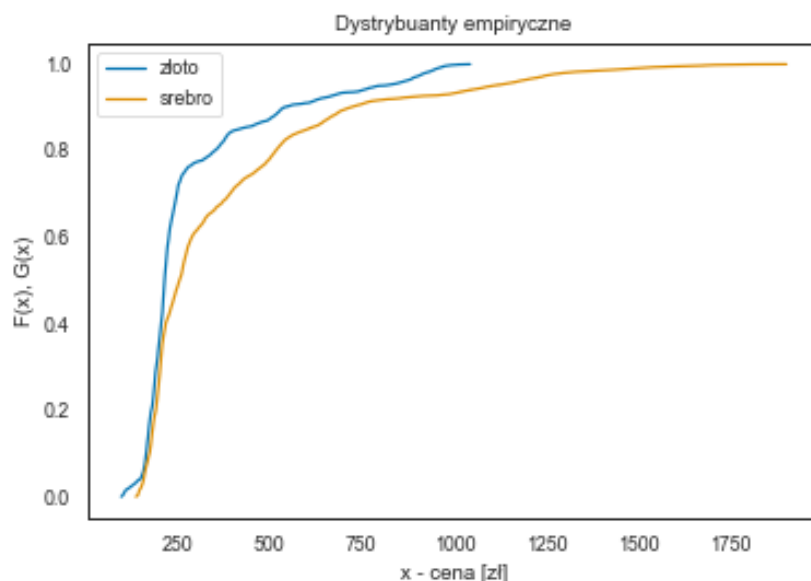
Rysunek 16: Histogramy cen surowców.

Na podstawie porównania histogramów możemy stwierdzić, że srebro przyjmuje dużo wyższe wartości, a histogram dość jednostajnie maleje do zera. Z kolei złoto przyjmuje niższe wartości w porównaniu ze srebrem, a spadek ilości danych przyjmujących dane wartości jest bardzo gwałtowny. Kolejnym spostrzeżeniem jest fakt, że zagęszczenia danych, o których wspominaliśmy w przypadku obu surowców, następują zazwyczaj na tych samych przedziałach cenowych.



Rysunek 17: Wykresy pudełkowe cen surowców.

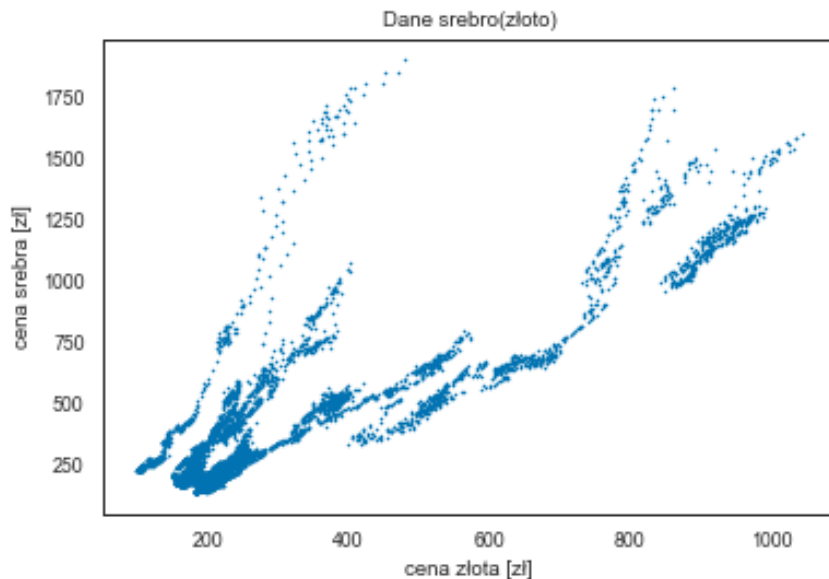
Analizując wykresy pudełkowe możemy zauważyć, że ceny srebra przyjmują zdecydowanie szerszy zakres. Zarówno rozstęp, jak i rozstęp międzykwartyłowy osiągają dużo większe wartości w przypadku tego surowca. Co ciekawe, Q1 oraz mediana znajdują się na podobnej wysokości. Różnica w wartościach rozstępów wynika ze znacząco wyższej wartości Q3 oraz maksimum. Interesujący jest fakt, że wartości odstających jest mniej więcej tyle samo w przypadku obu rozkładów, minimalnie więcej dla srebra.



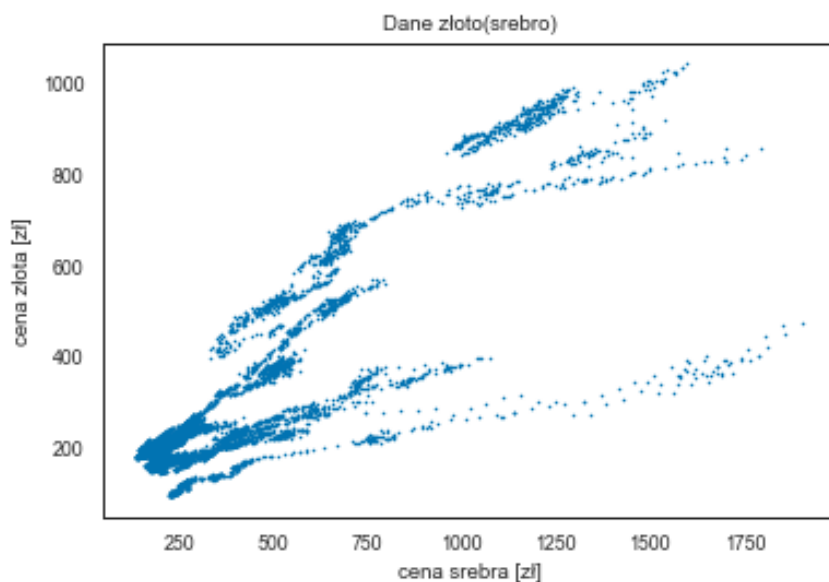
Rysunek 18: Dystrybuanty empiryczne cen surowców.

Rozważając dystrybuanty empiryczne obu surowców możemy zauważyć, że w przypadku złota przyjmuje ona wyższe wartości oraz o wiele szybciej dochodzi do 1. Jest to spowodowane faktem, że skoro złoto przyjmuje niższe wartości oraz są one bardziej skupione wokół mediany niż w przypadku srebra, to większe jest prawdopodobieństwo ich natrafienia. Z kolei fakt, że dystrybuanta empiryczna cen złota dużo szybciej osiąga wartość 1 spowodowany jest tym, że wartość maksymalna dla złota jest dużo niższa, a co za tym idzie prawdopodobieństwo osiągnięcia cen wyższych niż 1100zł jest zerowe, zatem nie następuje wzrost dystrybunaty. Kolejną obserwacją jest fakt, że po gwałtownym wzroście w pierwszej fazie w przypadku obu surowców, w dalszej części dystrybunaty dla srebra rośnie bardziej jednostajnie.

Wykonaliśmy wykresy rozproszenia danych tj. wykresy cen surowców względem siebie (Rysunek 19, Rysunek 20).



Rysunek 19: Zależność srebro(złoto).



Rysunek 20: Zależność złoto(srebro).

Na powyższych wykresach możemy zauważyć, że nie występują punkty, dla których wartość cen złota jest wysoka a srebra niska. Istnieje jednak sytuacja przeciwna. Są to pojedyncze punkty, dla których wartość dla złota mieści się w zakresie od 200 zł do 400 zł przy wysokich cenach srebra. Odpowiada to początkowemu fragmentowi wykresu. Rozważając wykres zależności złota od srebra możemy zaobserwować, że dane kumulują się w pobliżu punktu (250 zł, 200 zł). Odpowiada to centralnej części wykresu przebiegu danych (Rysunek 15), w której znajduje się najwięcej obserwacji. Następnie widzimy nieregularny wzrost zarówno dla argumentów, jak i wartości. Możemy również stwierdzić, że pomimo wzrostu wartości na obu osiach wzrost względem osi y-owej odpowiadającej srebru jest większy.

Ostatecznie obliczyliśmy wartości współczynników korelacji Pearsona oraz Spearmana. Otrzymaliśmy:

- współczynnik korelacji Pearsona: 0.8529206333022279,
- współczynnik korelacji Spearmana: 0.7035059468074513.

Na podstawie wyników możemy stwierdzić, że ceny srebra i złota są w dużym stopniu skorelowane, ponieważ wyniki, są bardziej zbliżone do 1 niż 0. Nie jest to jednak zależność liniowa, co moglibyśmy założyć w przypadku osiągnięcia współczynnika korelacji Pearsona równego 1. Współczynnik korelacji Spearmana wskazuje na silną, monotoniczną zależność zmiennych (niekoniecznie liniową).

## 4 Podsumowanie i wnioski

1. Przeprowadziliśmy analizę statystyczną dla przebiegu cen złota i srebra w kolejnych dniach. Następnie wykonaliśmy analizę porównawczą tych zmiennych względem siebie. Rozważając dane osobno zauważyliśmy, że wartości cen są dość rozproszone, jednak największe skupienie obserwacji występuje w okolicy mediany. Co ciekawe, mediana osiąga dość zbliżoną wartość w przypadku obu surowców, podobne zjawisko zaobserwowaliśmy w przypadku Q1 oraz minimum. Wartości Q3 oraz maksimum są zdecydowanie większe w przypadku cen srebra. Skutkuje to większym rozstępem i rozstępem międzykwartylowym. Możemy również powiedzieć, że średnio ceny srebra są wyższe. Kolejnym ciekawym faktem jest to, że w przypadkach obu surowców mamy mniej więcej tyle samo wartości odstających, których jest sporo.
2. Wnioski wysuwające się z przeprowadzonej analizy są nieco zaskakujące, ponieważ ogólnie przyjętym w społeczeństwie przekonaniem jest fakt, że złoto jest zdecydowanie droższe.
3. Wykonując z kolei analizę porównawczą doszliśmy do wniosków, że oprócz początkowej fazy obserwacji, w której to ceny srebra były zdecydowanie wyższe, dane przyjmowały zbliżone wartości, szczególnie w centralnej fazie obserwacji. Na końcu srebro ponownie przyjęło wysokie wartości, jednak pokryło się to ze wzrostem dla złota.
4. Z naszych obliczeń wynika, że dane są w dużym stopniu skorelowane. Nie jest to zaskakujący rezultat, ponieważ rynek sprzedaży złota i srebra jest podobny. Bardzo wiele czynników wpływa na kształtowanie się cen tych surowców. W większości przypadków czynniki te się pokrywają, co skutkuje podobnym zachowaniem rozkładów tych danych.