

Raport 1.

10 maja 2022

1. Informacje

Dane pochodzą ze strony finance.yahoo.com. Zbiory danych to odpowiednio: X - kurs funta brytyjskiego w przeliczeniu na dolary (GBP/USD), Y - kurs euro w przeliczeniu na dolary (EUR/USD). Do analizy użyto dziennych danych z okresu 5 lat, od 17 kwietnia 2017 roku do 17 kwietnia 2022 roku. Ze zbiorów zostały usunięte wiersze z brakiem danych. Ostateczna liczba wierszy obu zbiorów wynosi 1303.

2. Analiza

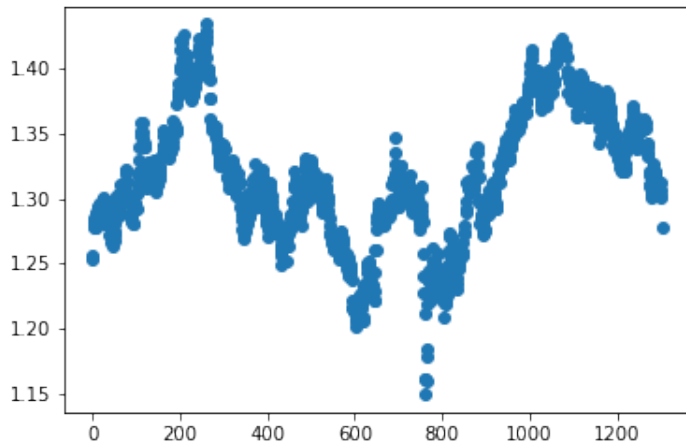
Wszystkie statystyki zostały obliczone za pomocą własnoręcznie napisanych programów z użyciem estymatorów nieobciążonych lub funkcji wbudowanych.

	X	Y
Mediana	1.3120	1.1532
Kwartył rzędu $\frac{1}{4}$	1.2866	1.1207
Kwartył rzędu $\frac{3}{4}$	1.3558	1.1842
Rozstęp międzykwartyłowy	0.0691	0.0635
Rozstęp	0.2845	0.1901
Wariancja	0.0026	0.0017
Odchylenie standardowe	0.0507	0.0415
Współczynnik zmienności	3.85%	3.60%

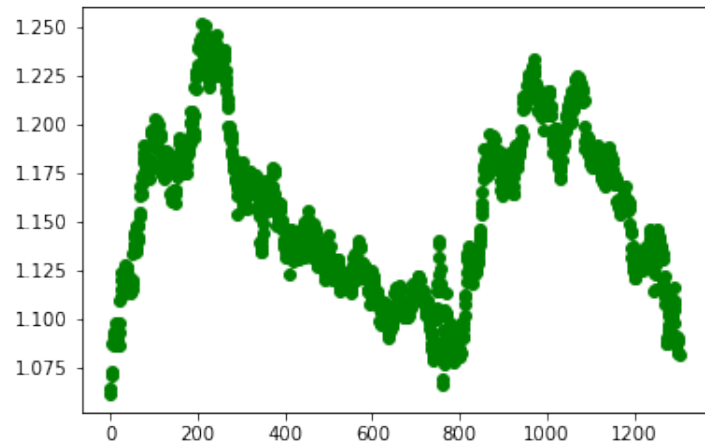
Tabela 1. Tabela charakterystyk zbiorów X i Y

Mediana to wartość, powyżej i poniżej której znajduje się jednakowa liczba obserwacji. Kwantyle spełniają relację $F(x_p) = p$, gdzie p to wartość kwantyla, która nas interesuje. Rozstęp międzykwartyłowy to różnica między kwartylami rzędów $\frac{1}{4}$ oraz $\frac{3}{4}$. Rozstęp jest różnicą między wartością największą i najmniejszą zbioru. Wariancję obliczamy ze wzoru: $\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$. Jest to estymator nieobciążony. Odchylenie standardowe to pierwiastek kwadratowy z wariancji. Współczynnik zmienności to parametr określający miarę zróżnicowania cechy.

Niektóre z charakterystyk obu zbiorów są do siebie dość zbliżone, np. współczynniki zmienności w obu zbiorach różnią się od siebie o 0,25%.



Rysunek 1. Scatterplot zbioru X



Rysunek 2. Scatterplot zbioru Y

Scatterploty zbiorów wykonano w celu zobrazowania jak przebiegały obniżki i wzrosty kursów EUR i GBP. Oś x reprezentują kolejne liczby porządkowe danych.

3. Interpretacja

Średnia	X	Y
Arytmetyczna	1.3177	1.1541
Geometryczna	1.3167	1.1537
Harmoniczna	1.3157	1.1526

Tabela 2. Tabela średnich arytmetycznych, geometrycznych i harmoniczných dla zbiorów X i Y

Obliczone wartości kolejnych średnich zostały zaokrąglone do czterech miejsc po przecinku w celu zaprezentowania nieznacznych różnic między nimi. Średnią arytmetyczną obliczamy ze wzoru:

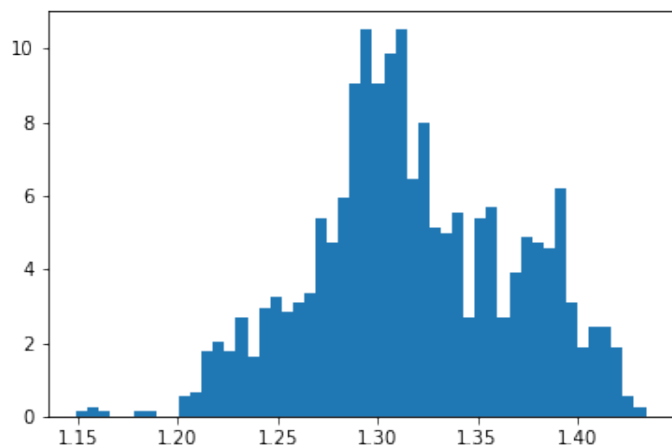
$$\overline{x_a} = \frac{1}{n} \sum_{i=1}^n x_i$$

Średnią geometryczną obliczamy ze wzoru:

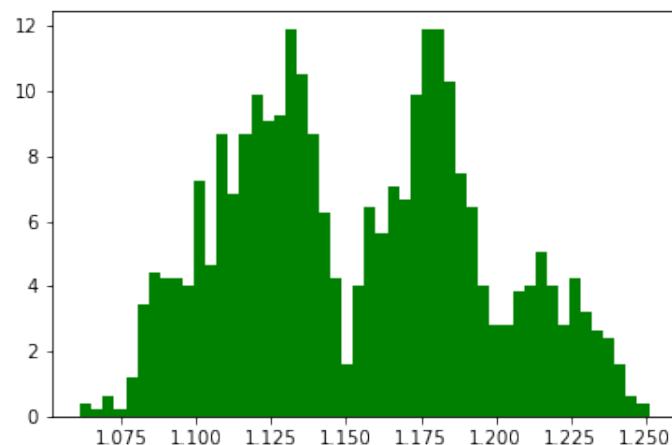
$$\overline{x_g} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Średnią harmoniczną obliczamy ze wzoru:

$$\overline{x_h} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

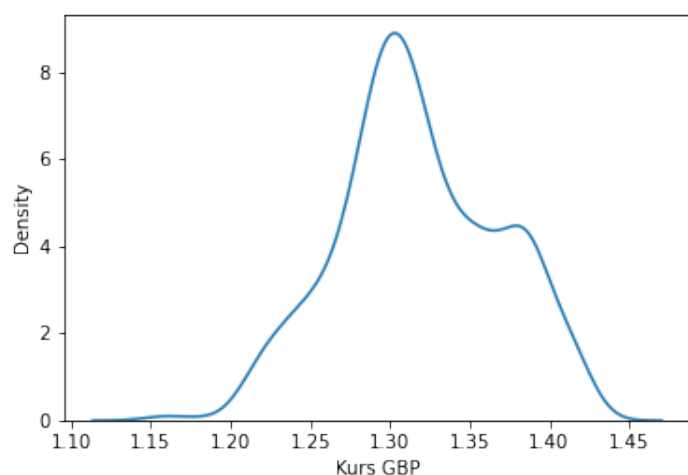


Rysunek 3. Histogram zbioru X

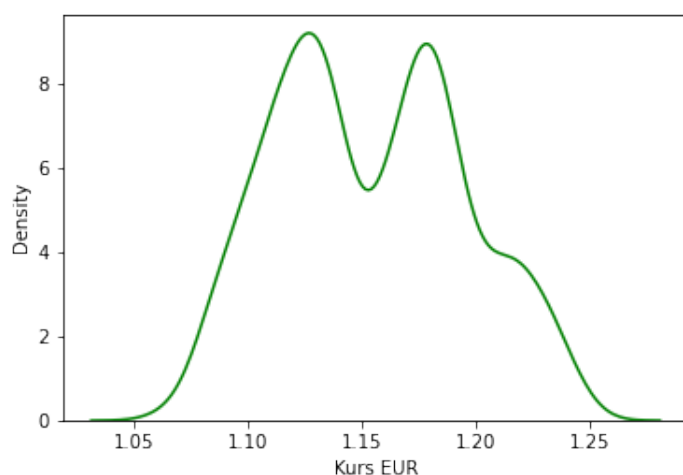


Rysunek 4. Histogram zbioru Y

Histogramy zbiorów X i Y pokazują liczbę obserwacji poszczególnych wartości kursu. W zbiorze X najczęściej występują wartości ze środka zbioru, z zakresu 1,27 – 1,33. Natomiast w zbiorze Y, inaczej niż w przypadku X, mamy dwa najczęściej występujące zakresy: 1,12 – 1,13,5 oraz 1,17 – 1,185. Oczywiście są to przybliżone zakresy, odczytane na podstawie wykresu.

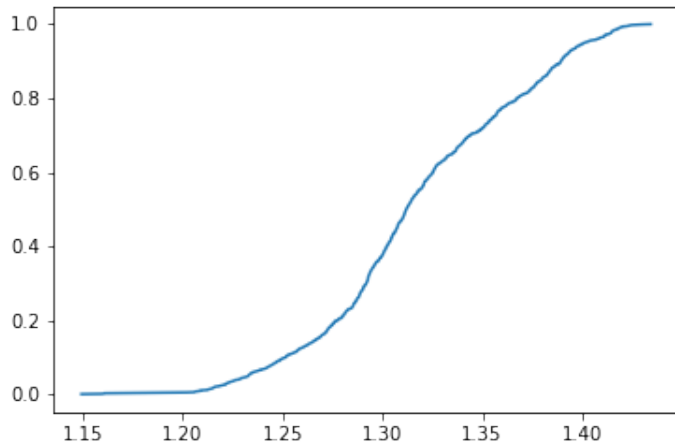


Rysunek 5. Gęstość zbioru X

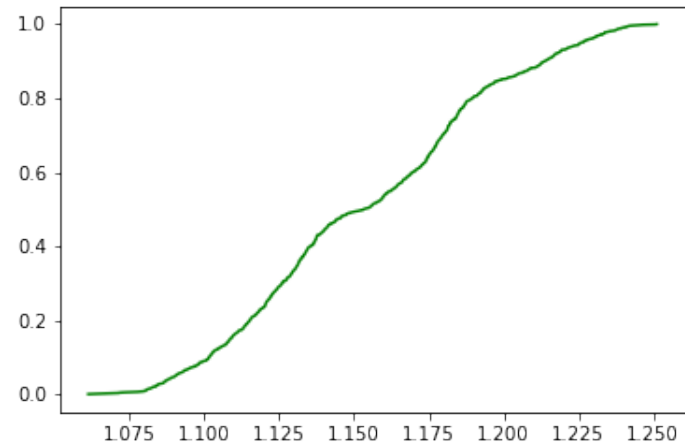


Rysunek 6. Gęstość zbioru Y

Gęstości to wygładzone histogramy i reprezentuje dokładnie to samo za pomocą ciągłej funkcji zamiast słupków. Jak widzimy nie zawierają wszystkich słupków przedstawionych na histogramie. W przypadku zbioru Y linia gęstości nie schodzi w pewnym momencie do odpowiedniego poziomu występowania obserwacji.

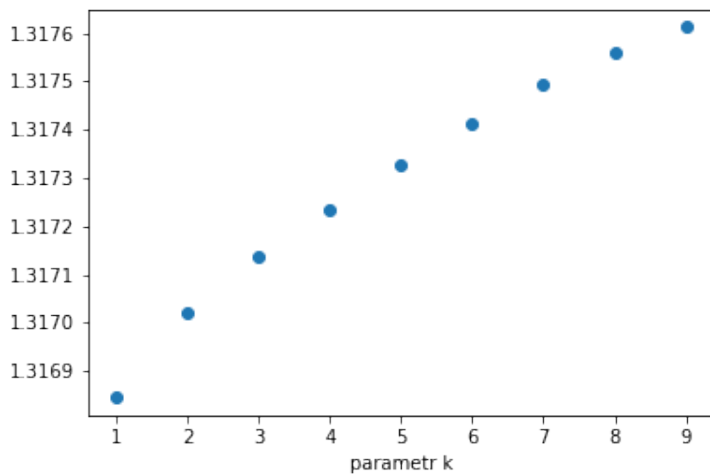


Rysunek 7. Dystrybuanta zbioru X

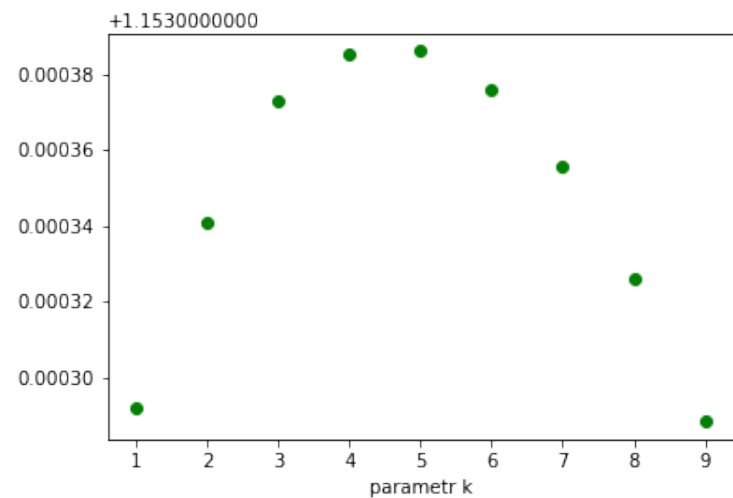


Rysunek 8. Dystrybuanta zbioru Y

Dystrybuanta informuje nas o prawdopodobieństwie wystąpienia wartości mniejszych bądź mniejszych od tej, która nas interesuje $P(Z \leq t)$. I tak na przykład szansa, że ze zbioru X otrzymamy wartość $\leq 1,25$ wynosi ~ 0.1 , w zbiorze Y szansa na otrzymanie tej samej wartości wynosi już 1.



Rysunek 9. Średnia ucinana zbioru X w zależności od parametru k

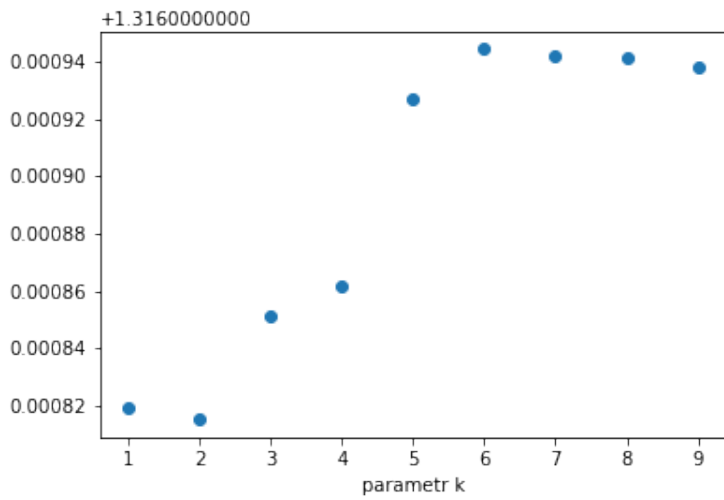


Rysunek 10. Średnia ucinana zbioru Y w zależności od parametru k

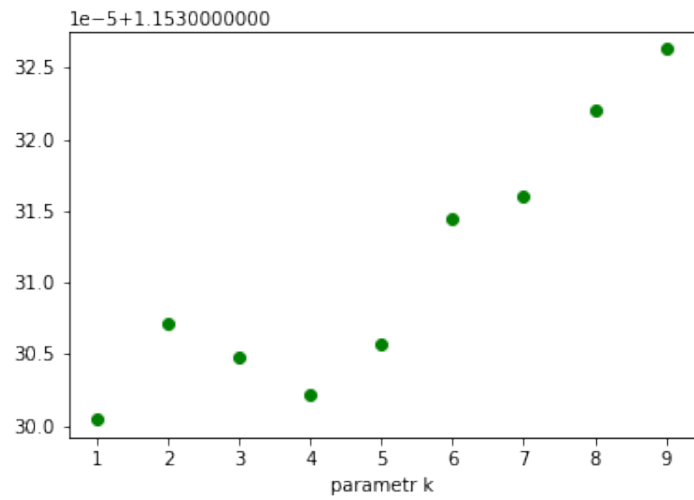
Średnią ucinaną obliczamy ze wzoru:

$$\overline{x_{tk}} = \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} x_i$$

Średnia ucinana zbioru X w zależności od parametru k wzrasta liniowo, natomiast Y - krzywoliniowo.



Rysunek 11. Średnia winsorowska zbioru X w zależności od parametru k



Rysunek 12. Średnia winsorowska zbioru Y w zależności od parametru k

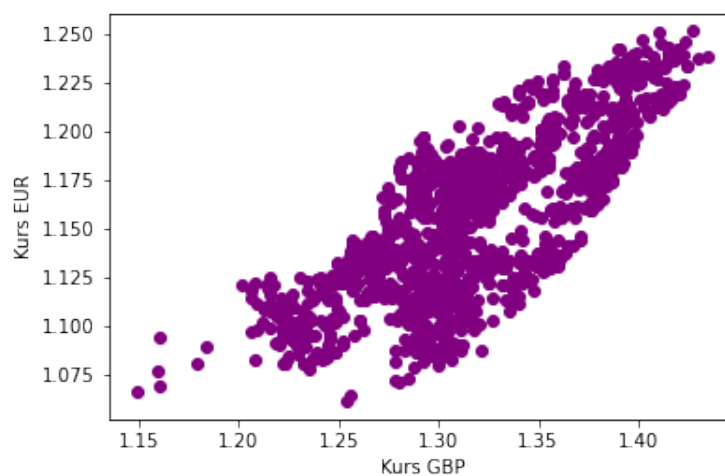
Średnią winsorowską obliczamy ze wzoru:

$$\overline{x_{wk}} = \frac{1}{n} \left[(k+1)x_{k+1} + \sum_{i=k+2}^{n-k-1} x_i + (k+1)x_{n-k} \right]$$

Średniej ucinanej i winsorowskiej używamy, by ograniczyć wpływ wartości odstających na naszą analizę.

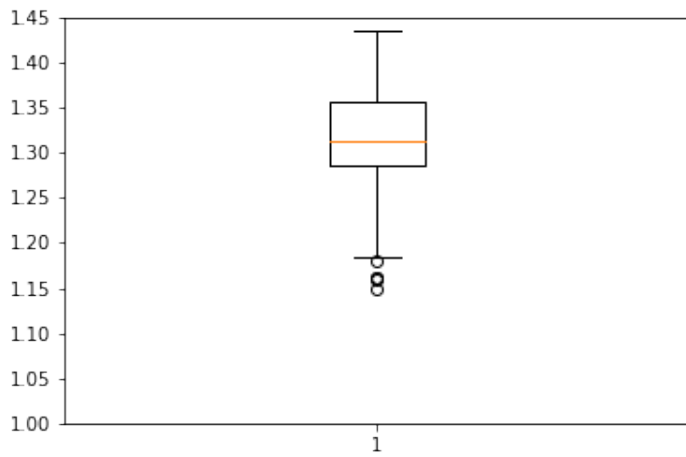
4. Porównanie

Korelacja zbiorów jest dość wysoka i wynosi ~ 0.77 .

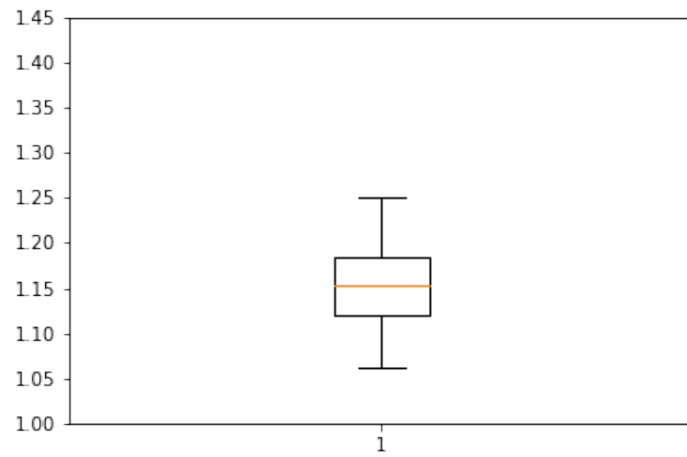


Rysunek 13. Scatterplot zależności Y od X

Wykres pokazuje zależność danych w zbiorze Y od danych w zbiorze X. Jest to zależność liniowa. Gdy kurs EUR rośnie, rośnie też kurs GBP.



Rysunek 14. Boxplot zbioru X



Rysunek 15. Boxplot zbioru Y

Widzimy, że zbiór X ma co najmniej 3 wartości odstające, podczas gdy Y nie ma żadnej. Mediana zbioru X jest przesunięta w stronę pierwszego kwartyli, zbioru Y wydaje się być na dokładnie pomiędzy kwartylami. Najniższa wartość X jest mocniej wysunięta niż najwyższa, podczas gdy Y są rozmieszczone w podobnej odległości od kwartyli. Żaden ze zbiorów nie ma górnej ani dolnej obserwacji ekstremalnej.

5. Podsumowanie

Zakresy, które przyjmują dane ze zbiorów X i Y nie są bardzo rozległe, dlatego nie obserwujemy ogromnych różnic przy wyliczaniu różnych statystyk, na przykład średnich. Na podstawie analizy wykresów oraz obliczeniu współczynnika korelacji odkryliśmy, że kursy walut EUR i GBP są mocno skorelowane. Dane zbioru Y są bardziej skupione niż dane zbioru X, który posiada kilka wartości odstających.