

Raport 2

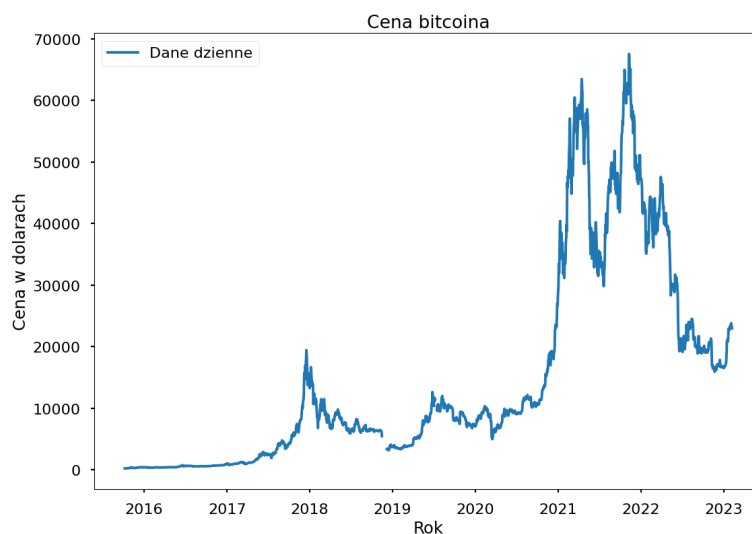
Analiza danych rzeczywistych przy pomocy modelu ARMA

Bartłomiej Gintowt

08.02.2023

1 Wstęp

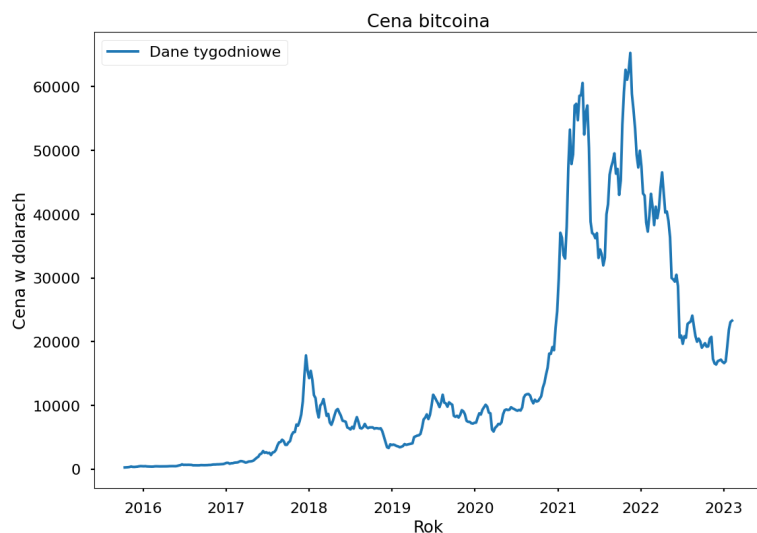
Celem naszego sprawozdania jest analiza wartości kryptowaluty Bitcoin i dopasowanie do niej jak najlepszego szeregu z modelu $ARMA(p,q)$. Z danych będziemy usuwać odpowiednio trend i sezonowość, następnie dobierzemy odpowiednie rzędy modelu oraz parametry. Przetestujemy również czy residua utworzonego modelu będą spełniały założenia białego szumu, takie jak stałość wariancji, średnia równa zero, niezależność oraz czy będą one z rozkładu normalnego. Dane nasze pochodzą ze strony <https://www.kaggle.com/datasets/swaptr/bitcoin-historical-data> i są to wartości Bitcoina aktualizowane co minutę na przestrzeni lat 2015-2023.



Rysunek 1: Wykres rzeczywistych cen Bitcoina

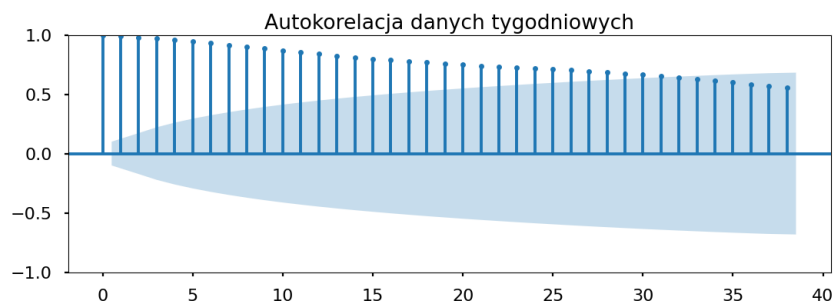
2 Przygotowanie danych do analizy

Dane Bitcoina jak można zauważyć na wykresie numer 1 posiadają pewne braki. Aby się z nimi uporać będziemy bazować na uśrednionych tygodniowych danych Bitcoina.

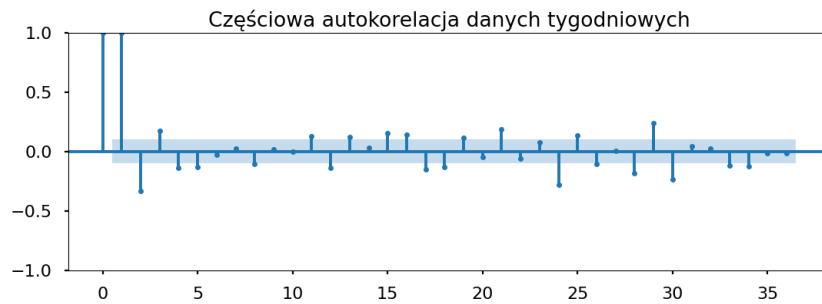


Rysunek 2: Wykres tygodniowych rzeczywistych cen Bitcoina

Zbadamy autokorelację i częściową autokorelację tygodniowych danych aby ocenić ich zależność w celu zbadania stacjonarności.



Rysunek 3: Korelacja danych tygodniowych



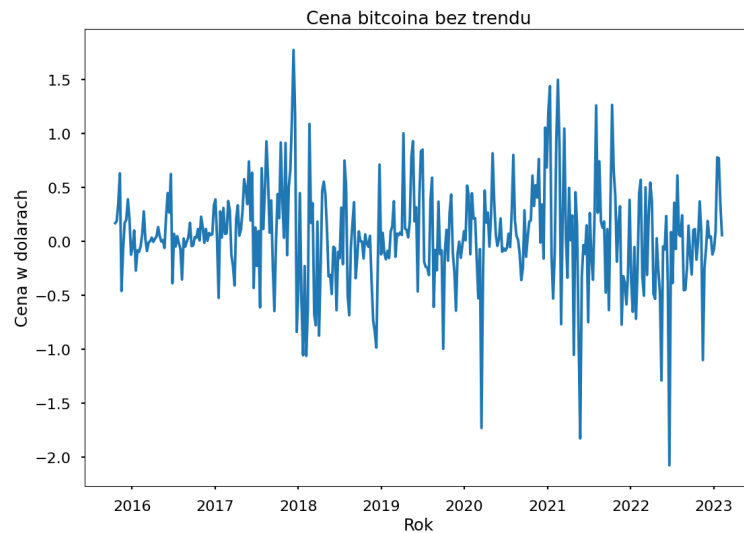
Rysunek 4: Częściowa korelacja danych tygodniowych

Z wykresu obu korelacji wynika, iż dane nie są skorelowane a sam szereg nie jest stacjonarny. Dodatkowo przeprowadzając test Dickey’a-Fullera otrzymujemy p-wartość równą 0.517773. Dla p-wartości mniejszych od 0.05 szereg byłby stacjonarny, zatem wnioski otrzymane z wykresów korelacji zostały wyciągnięte poprawnie. Będziemy chcieli przeprowadzić dekompozycję Walda dla danych w celu uzyskania szeregu stacjonarnego. Dekompozycję rozpoczniemy do nałożenia transformacji Boxa-Coxa na dane w celu stabilizacji wariancji.



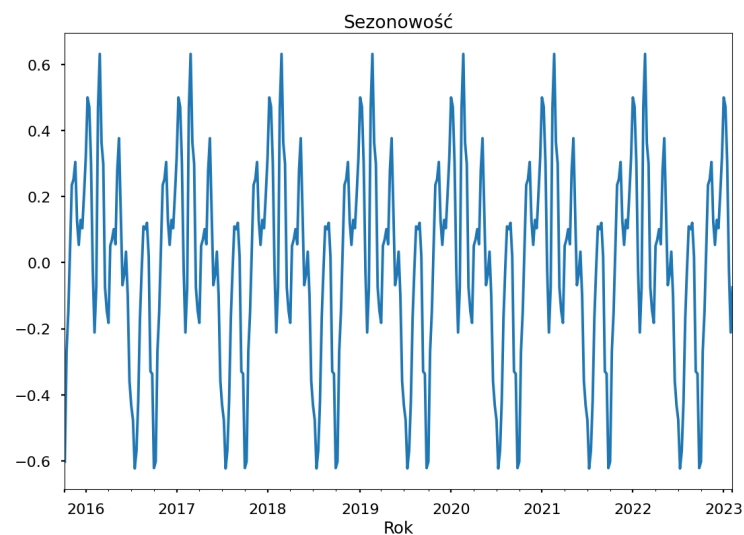
Rysunek 5: Dane po nałożeniu transformacji Boxa-Coxa

Kolejnym celem jest usunięcie trendu za pomocą różnicowania. Przyjmujemy parametr różnicowy równy 1.



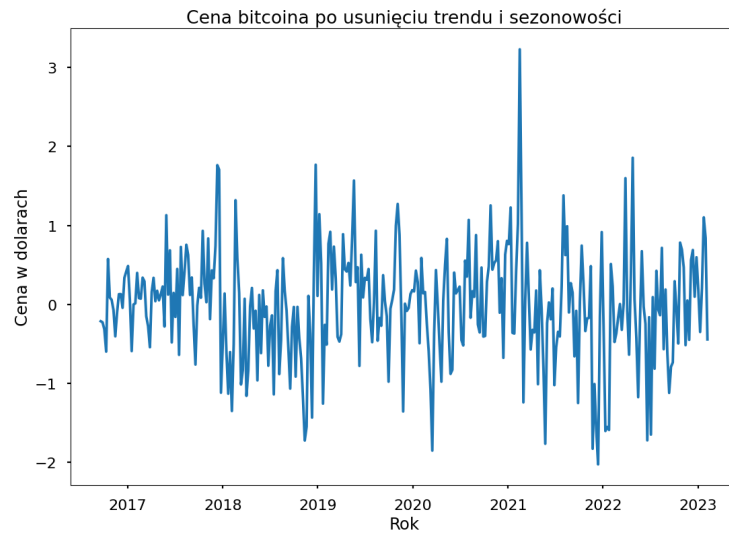
Rysunek 6: Dane po usunięciu trendu

Po usunięciu trendu będziemy chcieli się pozbyć sezonowości danych. Wpierw za pomocą wbudowanej funkcji "seasonal decompose" w języku Python sprawdzimy wykres sezonowości danych.



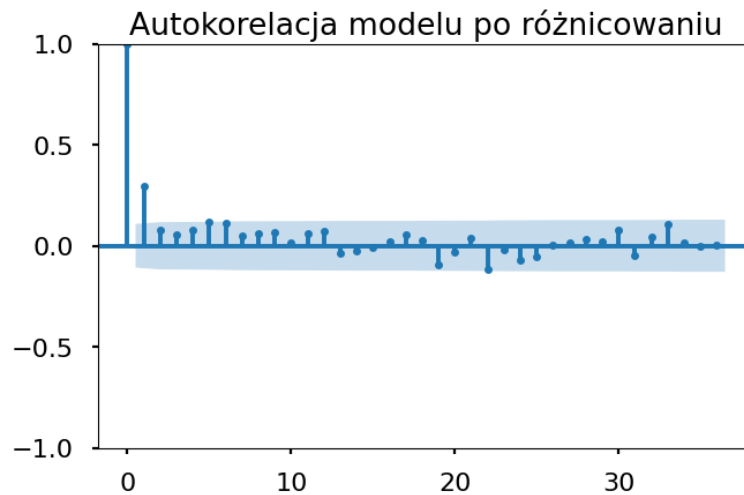
Rysunek 7: Sezonowość danych

Odczytujemy, iż za okres sezonowości możemy przyjąć 1 rok, czyli około 52 tygodnie. Możemy teraz usunąć sezonowość z danych.

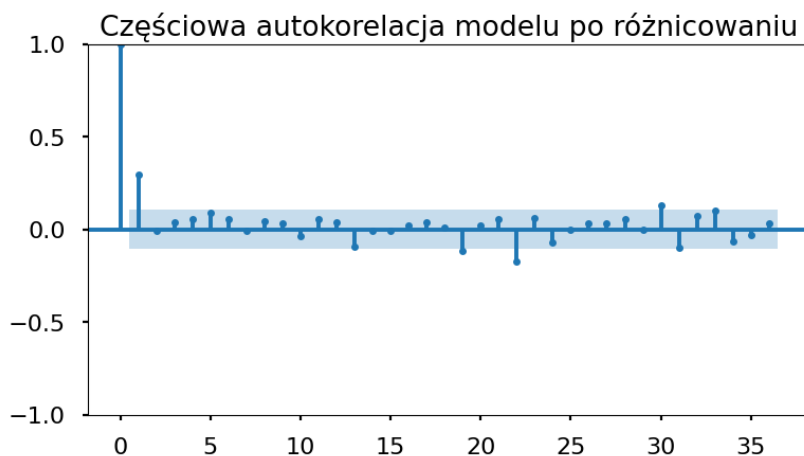


Rysunek 8: Dane po usunięciu trendu i sezonowości

Po przeprowadzonych operacjach na szeregu czasowym ponownie ocenimy za pomocą funkcji autokorelacji i autokorelacji częściowej jego stacjonarność.



Rysunek 9: Autokorelacja danych po różnicowaniu



Rysunek 10: Częściowa autokorelacja danych po różnicowaniu

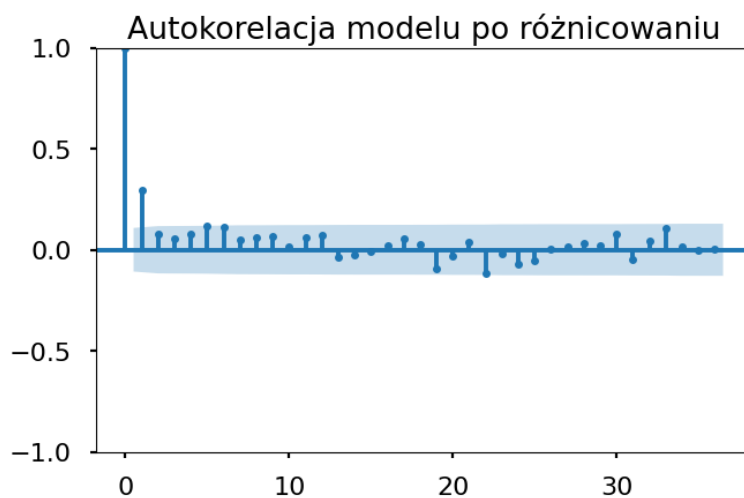
Tym razem otrzymany szereg po przekształceniach jest stacjonarny. W celu upewnienia się czy nasze wnioski są poprawne ponownie przeprowadzimy test Dickey’a-Fullera. Otrzymujemy p-wartość równą 0.0149, zatem potwierdziło się, iż nasz szereg jest stacjonarny.

3 Modelowanie danych przy pomocy ARMA

Mając już stacjonarny szereg czasowy ARMA możemy wyznaczyć jego rzędy p i q . Korzystając z modelu ARIMA z biblioteki "scipy.stats" możemy przy pomocy kryterium informacyjnego Akaikego ocenić, które rzędy będą zwracały najmniejszą wartość w rozważanym teście. Otrzymujemy, iż dane uzyskane po dekompozycji są szeregiem czasowym ARMA(2, 0), czyli AR(2). Wywołanie modelu zwraca nam również wartość parametrów szeregu. W tym przypadku otrzymujemy parametry $\phi_1 = 1.3157, \phi_2 = -0.3177$. Stąd widzimy, że nasz stacjonarny szereg czasowy jest postaci $X_t - 1.3157X_{t-1} + 0.3177X_{t-2} = Z_t$.

4 Ocena dopasowania modelu

W celu sprawdzenia poprawności dopasowania modelu sprawdzimy jak funkcja autokorelacji zachowuje się w przedziałach ufności uzyskanych za pomocą biblioteki "sm.graphics.tsa". Posłużymy się wykresem autokorelacji przedstawionym już wcześniej w raporcie.

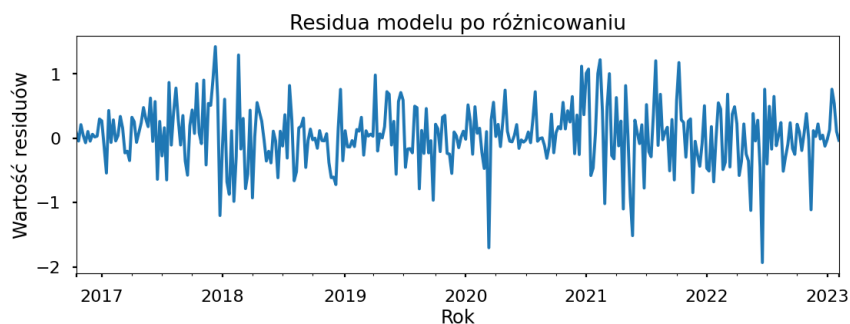


Rysunek 11: Autokorelacja danych po różnicowaniu

Zauważamy, iż wartości korelacji znajdują się w przedziałach ufności. Oznacza to, iż dane są niezależne a sam szereg jest stacjonarny.

5 Weryfikacja założeń dotyczących szumu

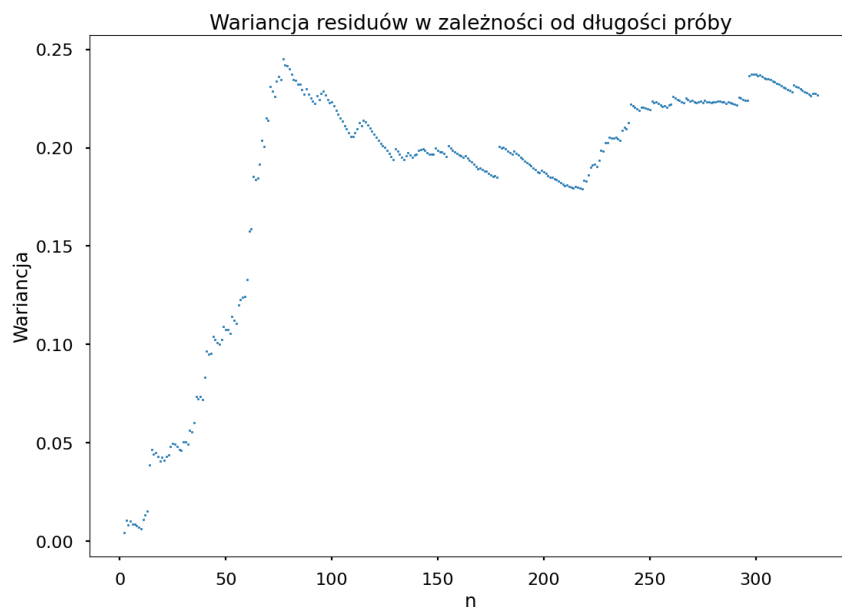
Na początku sprawdziliśmy, czy średnia wartości resztowych jest bliska 0. Wartość średniej jaką otrzymaliśmy wynosi 0.0417199, zatem możemy przyjąć, że residua spełniają warunek dotyczące średniej.



Rysunek 12: Wykres wartości resztowych modelu

Na wykresie numer 12 możemy zauważyć, że wartości residuów oscylują wokół zera.

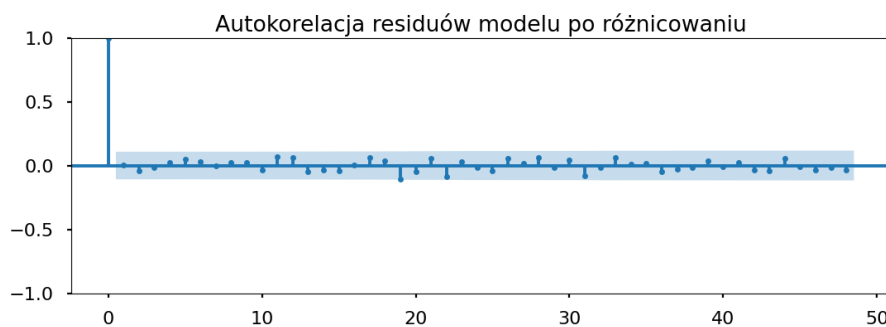
Następnie sprawdziliśmy warunek, czy residua posiadają stałą wariancję.



Rysunek 13: Wykres wariancji residuów w zależności od długości próby

Wykres numer 13 pokazuje, że wariancja residuów zmienia się wraz z zwiększeniem długości próby, co mówi nam, iż wartości resztowe nie posiadają stałej wariancji.

Do sprawdzenia czy residua są od siebie zależne wykonaliśmy wykres autokorelacji.

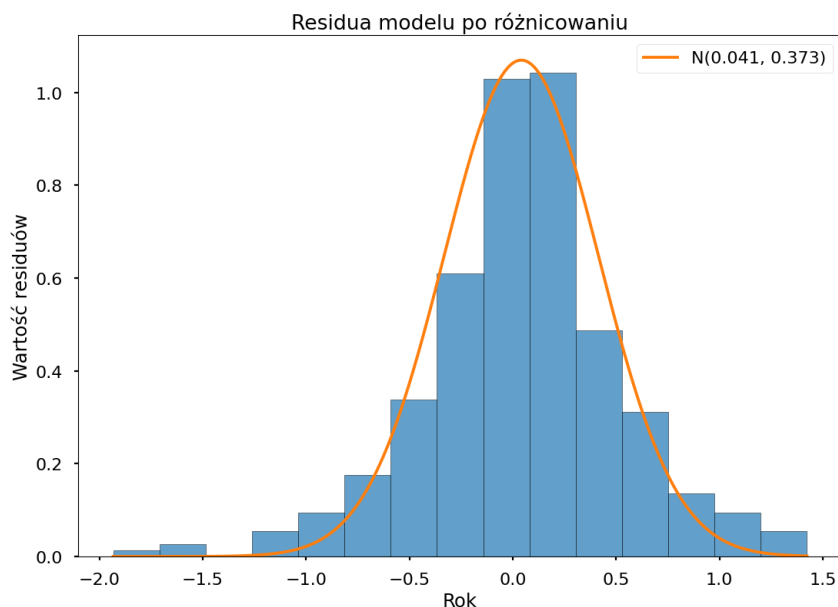


Rysunek 14: Wykres autokorelacji dla wartości resztowych

Wykres numer 14 pokazuje, że wartości resztowe są od siebie niezależne ponieważ, wartość autokorelacji dla $h = 0$ jest równa 1, natomiast dla $h \neq 0$

wartości oscylują wokół zera.

Na koniec sprawdziliśmy założenie dotyczące normalności rozkładu, w celu sprawdzenia tego warunku wykonaliśmy histogram residuów i dopasowaliśmy do niego gęstość rozkładu normalnego.



Rysunek 15: Histogram wartości resztowych modelu

Jak można zauważyć na wykresie numer 15, iż gęstość empiryczna oraz teoretyczna są do siebie podobne, co może wskazywać na to, że residua mają rozkład normalny. Dodatkowo przeprowadziliśmy test Shapiro-Wilka dla którego p-wartość wynosi 0.000135, zatem możemy przyjąć, że wartości resztowe są z rozkładu normalnego

6 Podsumowanie

Przeprowadzając analizę danych wartości Bitcoina na przedziale lat 2015-2023 byliśmy w stanie przeprowadzając dekompozycję Walda uzyskać z niego szereg $AR(2)$ stacjonarny w słabym sensie. Postać szeregu możemy opisać równaniem $X_t - 1.3157X_{t-1} + 0.3177X_{t-2} = Z_t$. Gdzie Z_t spełnia warunek na wartość średniej, niezależność oraz normalność, natomiast nie spełnia założenia stałości wariancji. Nie możemy zatem nazwać go białym szumem naszego modelu $AR(2)$.