

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Andrzej Żernaczuk
Nr albumu 407240

Dekompozycja i prognozowanie szeregów czasowych

Informatyka i Ekonometria

Projekt wykonany w ramach
konwersatorium z Analizy Szeregów Czasowych
pod kierunkiem dr A. Dzik-Walczak

Warszawa, sierpień 2023

Spis Treści

WSTĘP	2
1. ANALIZA SZEREGU NIESEZONOWEGO - IBM	3
1.1. Dekompozycja notowań giełdowych IBM	3
1.2. Model ARIMA	5
1.3. Model Holta	9
1.4. Porównanie modeli	10
2. ANALIZA SZEREGU SEZONOWEGO - ENERGIA NA RYNKU WEW. NORWEGII	11
2.1. Dekompozycja dostępnej energii	11
2.2. Model SARIMA	13
2.3. Model Holta-Wintersa	16
2.4. Porównanie modeli	18
3. Podsumowanie	18

WSTĘP

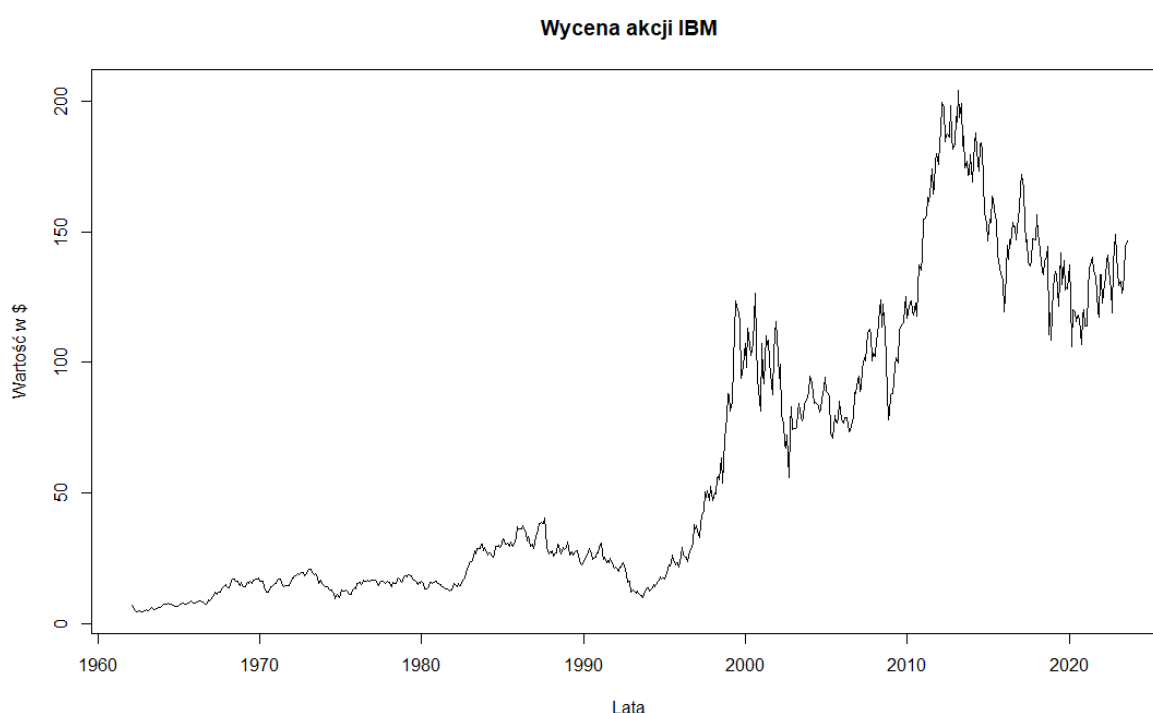
W niniejszej pracy podjęto się analizy, dekompozycji i prognozowania dwóch szeregów czasowych. Pierwszym z nich jest wartość wyceny akcji pierwszego dnia miesiąca spółki *International Business Machines Corporation (IBM)* na nowojorskiej giełdzie. Indeksy giełdowe są od zawsze poddawane próbom analizy i przewidzenia wartości jako najprostszy (oczywiście tylko w teorii) sposób zarobku. Drugim szeregiem poddanym badaniu jest miesięczna ilość energii dostępnej na wewnętrznym rynku Norwegii. Jest to oczywiście szereg sezonowy, można spodziewać się największego zapotrzebowania na kilowatogodziny w trakcie zimy. W ramach predykcji wykorzystane zostały modele ARIMA i SARIMA. Wszystkie operacje na danych zostały wykonane za pomocą języka programowania R w programie RStudio.

1. ANALIZA SZEREGU NIESEZONOWEGO - IBM

1.1. Dekompozycja notowań giełdowych IBM

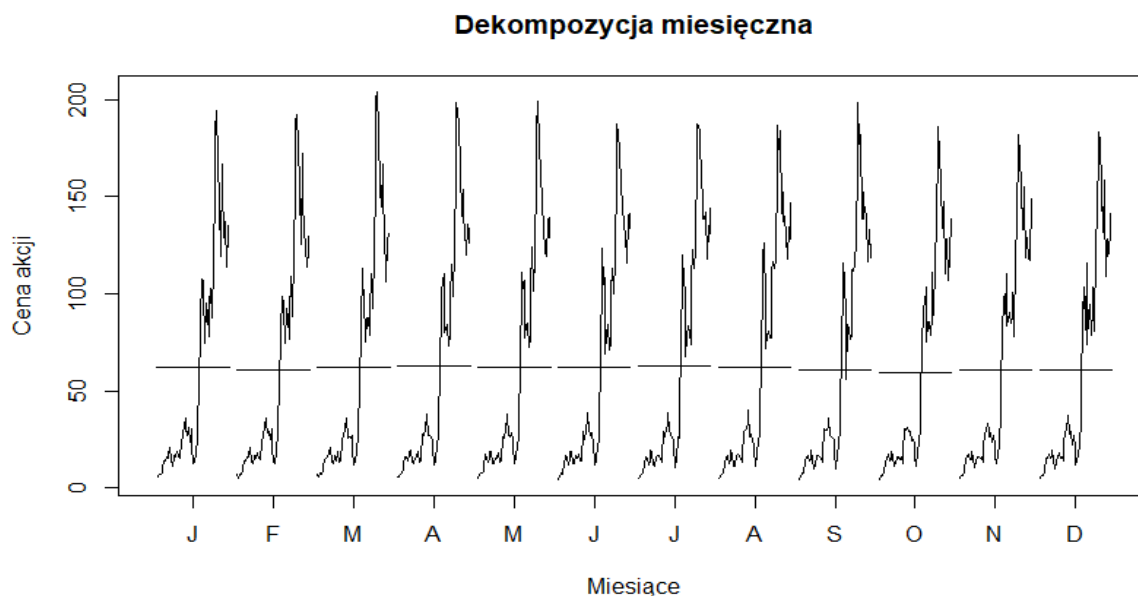
Pierwszym szeregiem analizowanym w niniejszej pracy jest wycena notowań giełdowych IBM w pierwszym dniu miesiąca pomiędzy lutym 1962 a sierpniem 2023. Na poniższym wykresie przedstawione zostały ceny zamknięcia na przestrzeni lat.

Rys. 1. Wycena akcji IBM.

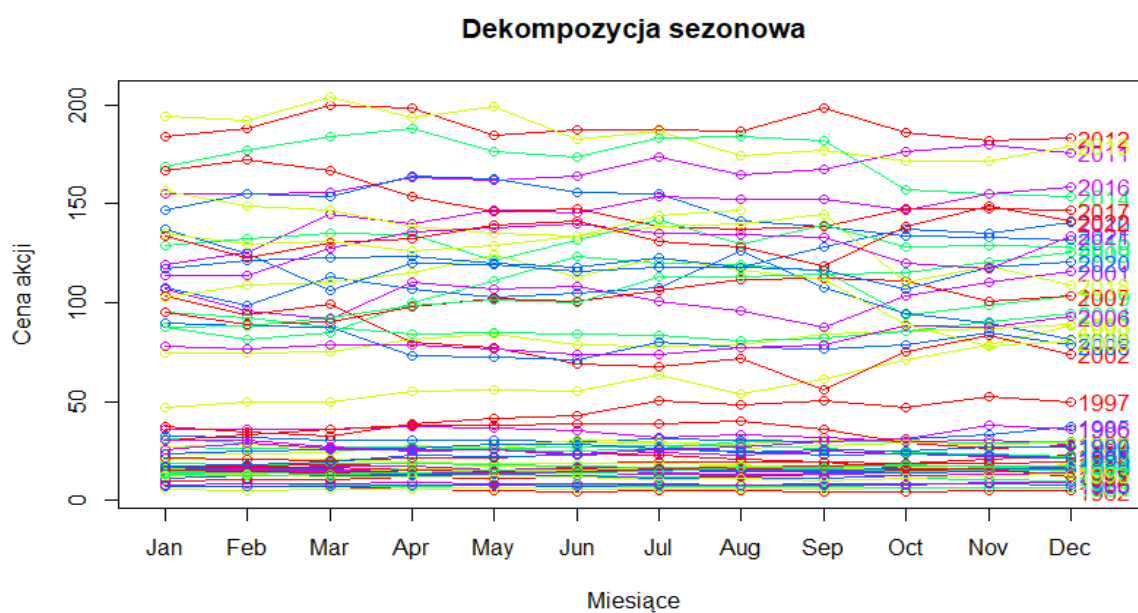


Wstępna analiza empiryczna sugeruje trend wzrostowy, co może odpowiadać intuicji podpowiadającej, że firma będąca jednym z gigantów technologicznych powinna na przestrzeni lat średnio zyskiwać na wartości. Pierwszy z widocznych znaczących wzrostów zaczyna się w okolicach roku 1995 a kończy tuż przed 2000. W latach 1998-2000 na giełdzie amerykańskiej miała miejsce bańka nazywana w anglojęzycznym internecie *dotcom bubble*. Jej początki, zidentyfikowane po czasie miały miejsce właśnie na przestrzeni lat 1995 i 1996, zaś tak zwane “pęknięcie” na przełomie lat 1999 i 2000. Następny znaczący spadek miał miejsce w 2008 roku czyli w momencie pęknięcia drugiej szeroko opisananej bańki, tym razem na rynku nieruchomości, która spowodowała de facto kryzys finansowy na skalę światową. Ostatnia z najbardziej widocznych zmian przypada na okres wybuchu epidemii covid-19.

Rys. 2. Dekompozycja miesięczna akcji IBM.



Rys. 3. Dekompozycja sezonowa akcji IBM.



Założenie o nie sezonowości potwierdzają wykresy widoczne na rysunku drugim. Na pierwszym z nich możemy zaobserwować prawie równą średnią wartość akcji dla każdego z miesięcy, na drugim widać że nie istnieją wahania sezonowe na przestrzeni lat. Można także zauważyć widoczny wzrost wyceny akcji w XXI wieku.

1.2. Model ARIMA

Celem zastosowania modelu ARIMA czyli *AutoRegressive Integrated Moving Average* jest predykcja na podstawie danych historycznych. Jako horyzont przewidywania przyjęto czas ostatnich trzech miesięcy, w tym celu podzielono dane na zbiór treningowy obejmujący okres 02.1962 - 05.2023 i zbiór testowy od czerwca do sierpnia 2023 roku. Poniżej przedstawiono wyniki testowania hipotezy o stacjonarności szeregu.

Tabela 1. Wyniki testów na stacjonarność niezróżnicowanego szeregu czasowego IBM.

Rodzaj testu	Wartość statystyki testowej	P-value	Interpretacja wyniku
ADF	-2.0186	0.5705	Niestacjonarność
KPSS	8.8406	0.01	Niestacjonarność

Wartość p-value dla testu Dickey-Fullera jest wyższa niż poziom istotności 5% co sprawia że nie można odrzucić hipotezy zerowej o niestacjonarności szeregu czasowego. W przypadku testu Kwiatkowskiego-Phillipsa-Schmidta-Shina p-value jest niższe niż 0.05 więc odrzucona została hipoteza o stacjonarności szeregu.

Po potwierdzeniu niestacjonarności szeregu nadano mu postać stacjonarną za pomocą dwukrotnego zróżnicowania go na bazie jednego okresu, po czym ponownie wykonano oba testy.

Rys. 4. Zróżnicowany szereg IBM.

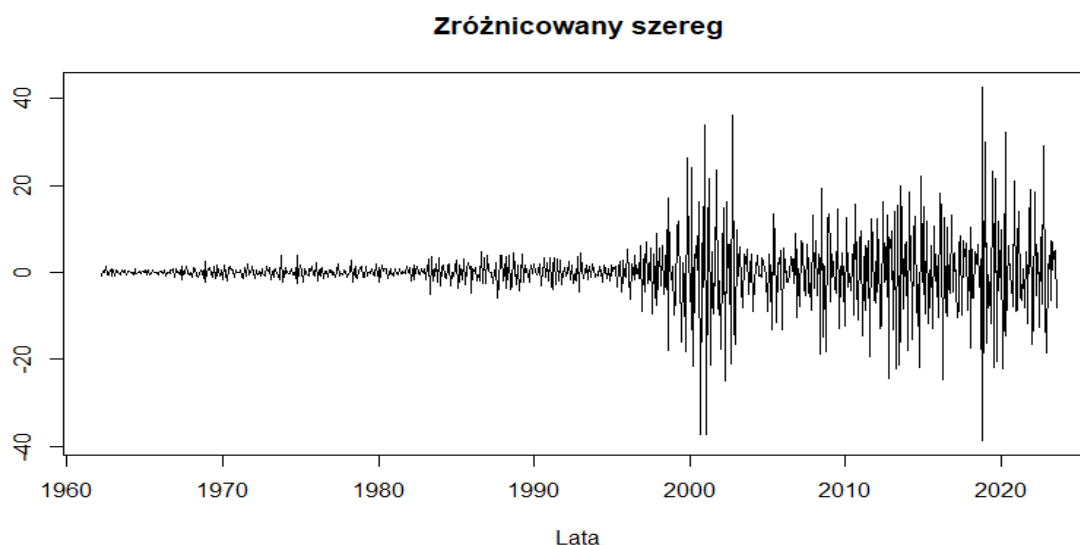


Tabela 2. Wyniki testów na stacjonarność zróżnicowanego szeregu czasowego IBM.

Rodzaj testu	Wartość statystyki testowej	P-value	Interpretacja wyniku
ADF	-16.487	0.01	Stacjonarność
KPSS	0.0051	0.1	Stacjonarność

Z powyższych wyników można wywnioskować, że szereg jest stacjonarny. P-value testu ADF jest w tym przypadku mniejsze od wartości 0.05 co powoduje odrzucenie hipotezy o niestacjonarności, zaś wynik testu KPSS jasno wskazuje na potwierdzenie hipotezy zerowej o stacjonarności badanego szeregu czasowego.

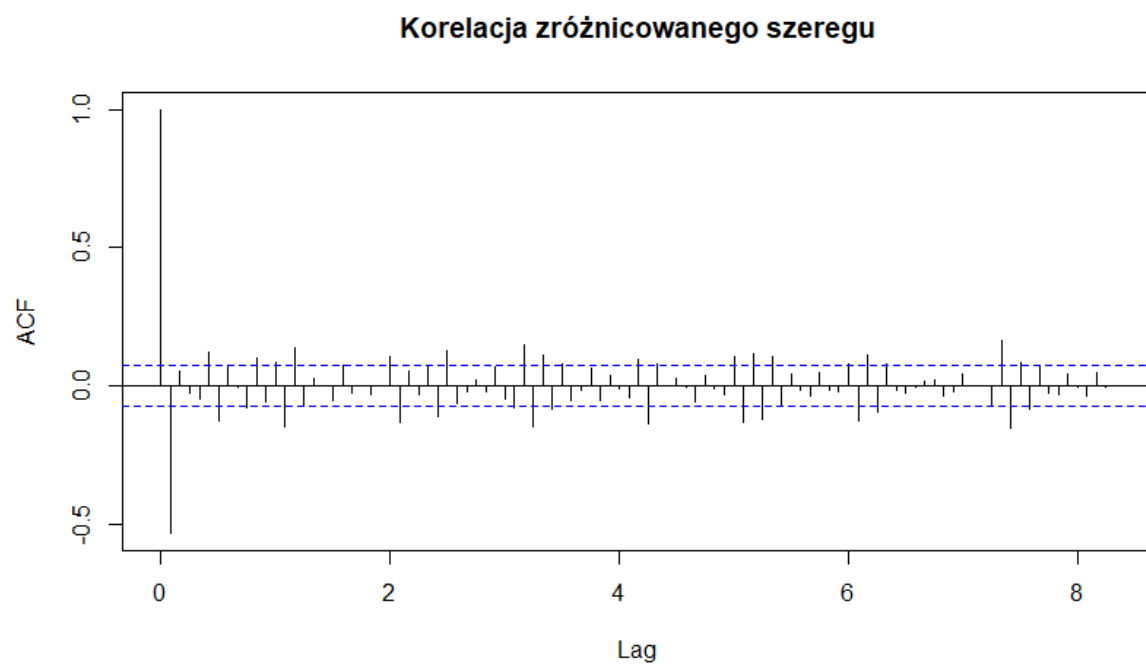
Przeprowadzony został również test Boxa-Ljunga by sprawdzić czy szereg ma charakterystykę białego szumu. P-value wyniosło mniej niż $2.2e-16$, co jest wartością znacznie niższą niż przyjęty poziom istotności. W wyniku tego odrzucono hipotezę zerową o braku autokorelacji.

Tabela 3. Wynik testu na Boxa-Ljunga.

Rodzaj testu	Wartość statystyki testowej	P-value	Interpretacja wyniku
Box-Ljung	261.72	$< 2.2e-16$	Brak charakterystyki białego szumu

Po spełnieniu założenia o stacjonarności szeregu możliwe staje się wykorzystanie modelu ARIMA do prognozowania wysokości cen akcji. Szereg został zróżnicowany dwukrotnie o jeden okres więc parametr d przyjął wartość 2. By określić parametry q oraz p posłużono się odpowiednio wykresami korelacji (ACF) i cząstkowej autokorelacji (PACF).

Rys. 5. ACF zróżnicowanego szeregu IBM.



Rys. 6. PACF zróżnicowanego szeregu IBM.

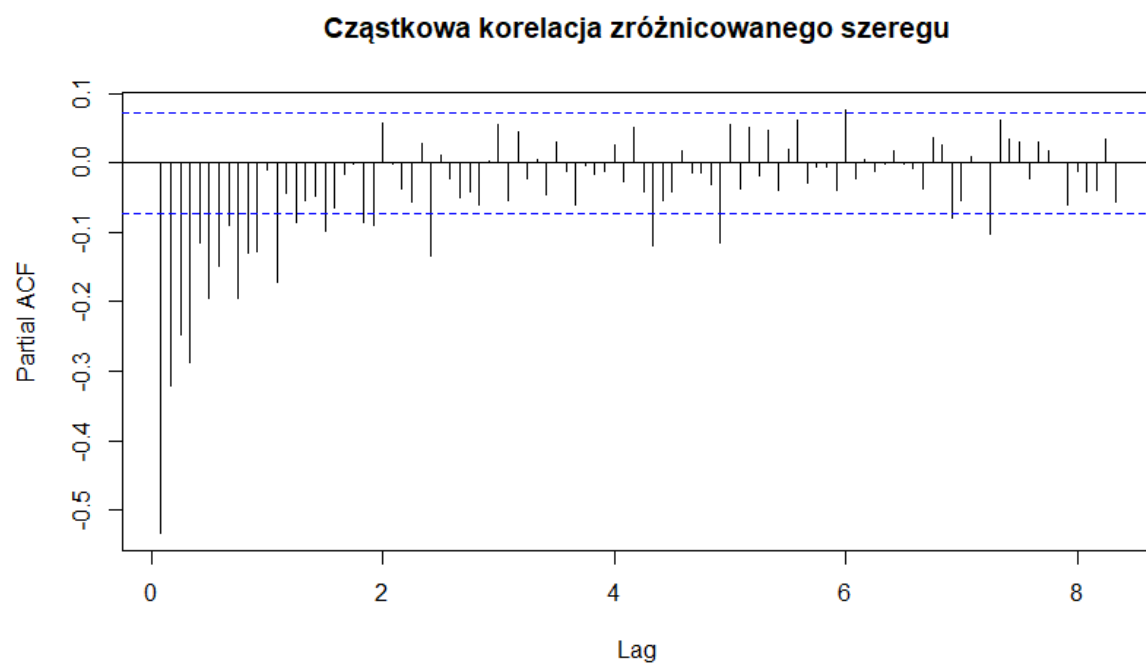


Tabela 4 Porównanie modeli ARIMA dla IBM.

Model	AIC	BIC
ARIMA(0, 2, 0)	5113.8	5118.4
ARIMA(1, 2, 0)	4869.68	4878.88
ARIMA(0, 2, 1)	4545.53	4554.73

Model ARIMA(0,2,0) mimo bycia słabszym pod względem AIC oraz BIC okazał się najlepszym modelem jeżeli chodzi o wartości błędów, dlatego został wybrany.

Tabela 5. Błędy prognozy dla modelu ARIMA(0, 2, 0) dla IBM.

Statystyka	Wartość
RMSE	9.3719
MAE	8.5300
MAPE	5.9301
MASE	0.9089

Rys. 7. Predykcja za pomocą modelu ARIMA dla IBM.

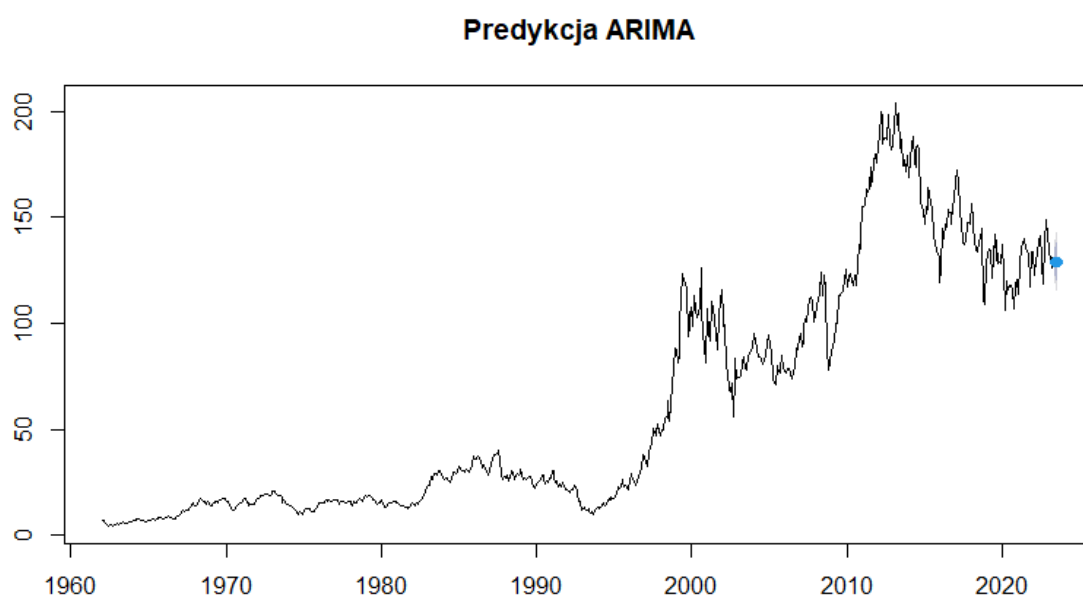


Tabela 6. Predykcje modelu ARIMA dla IBM.

Daty	Predykcja	low 80	high 80	low 95	high 95
Jun 2023	130.77	120.68262	140.8574	115.34268	146.1973
Jul 2023	132.95	110.39393	155.5060	98.45348	167.4465
Aug 2023	135.13	97.38648	172.8735	77.40628	192.8537

1.3. Model Holta

Tabela 7. Błędy prognozy dla modelu Holta dla IBM.

Statystyka	Wartość
RMSE	13.8080
MAE	12.7187
MAPE	8.8520
MASE	1.3542

Rys. 8. Predykcja za pomocą modelu Holta dla IBM.

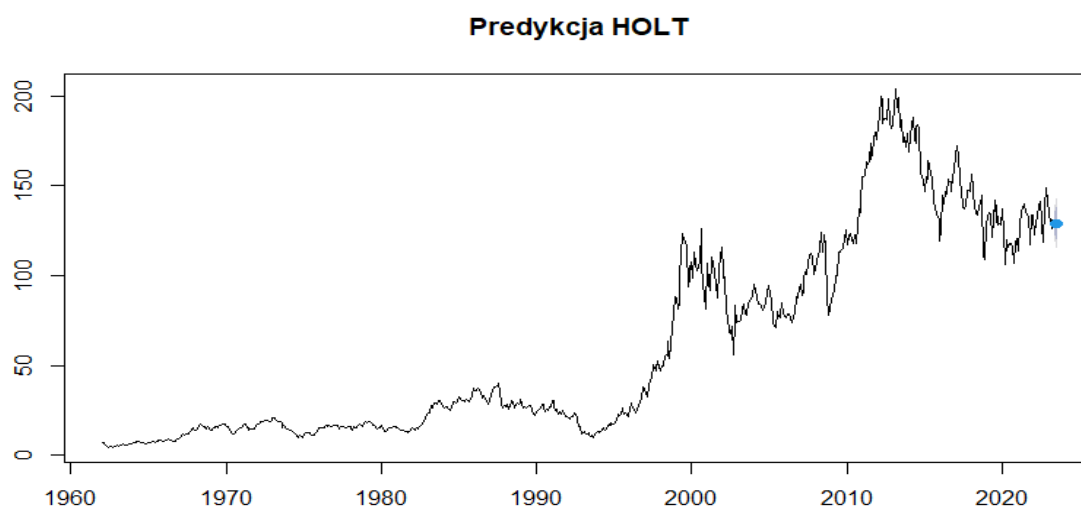


Tabela 8. Predykcje Holta dla IBM.

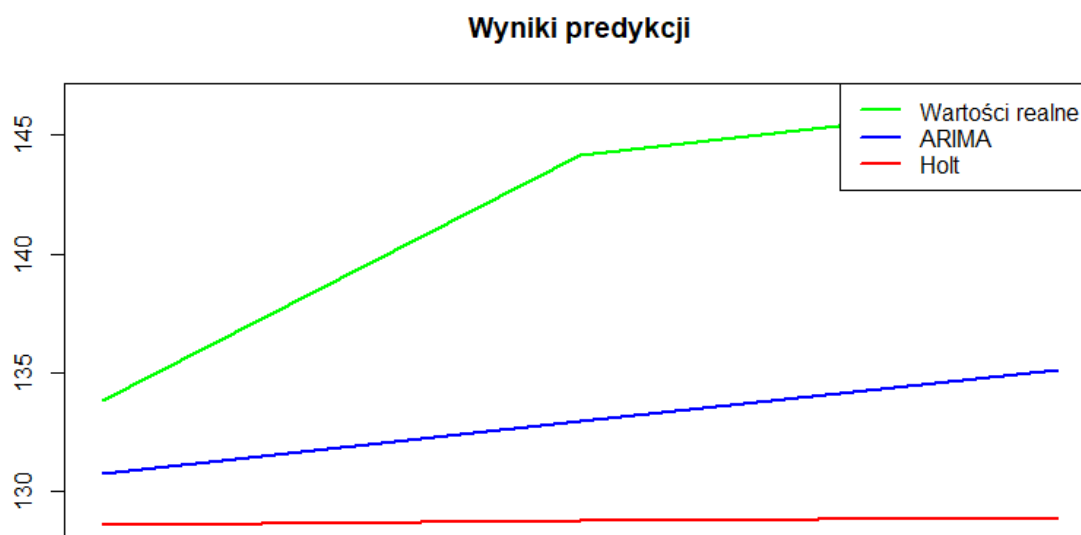
Daty	Predykcja	low 80	high 80	low 95	high 95
Jun 2023	128.5957	121.8113	135.3801	118.2199	138.9715
Jul 2023	128.7613	119.6592	137.8635	114.8408	142.6819
Aug 2023	128.9270	117.9872	139.8667	112.1960	145.6579

Model Holta okazał się stabilniejszy w predykcji.

1.4. Porównanie modeli

Na poniższym wykresie można zauważyć istotną różnicę w predykcjach modeli. Metoda Holta okazała się bezużyteczna przewidując praktycznie zerową zmienność danych, jednak ARIMA mimo błędów oszacowania dobrze oddała trend modelu, trafnie przewidując wzrost.

Rys. 9. Porównanie wyników dla IBM.

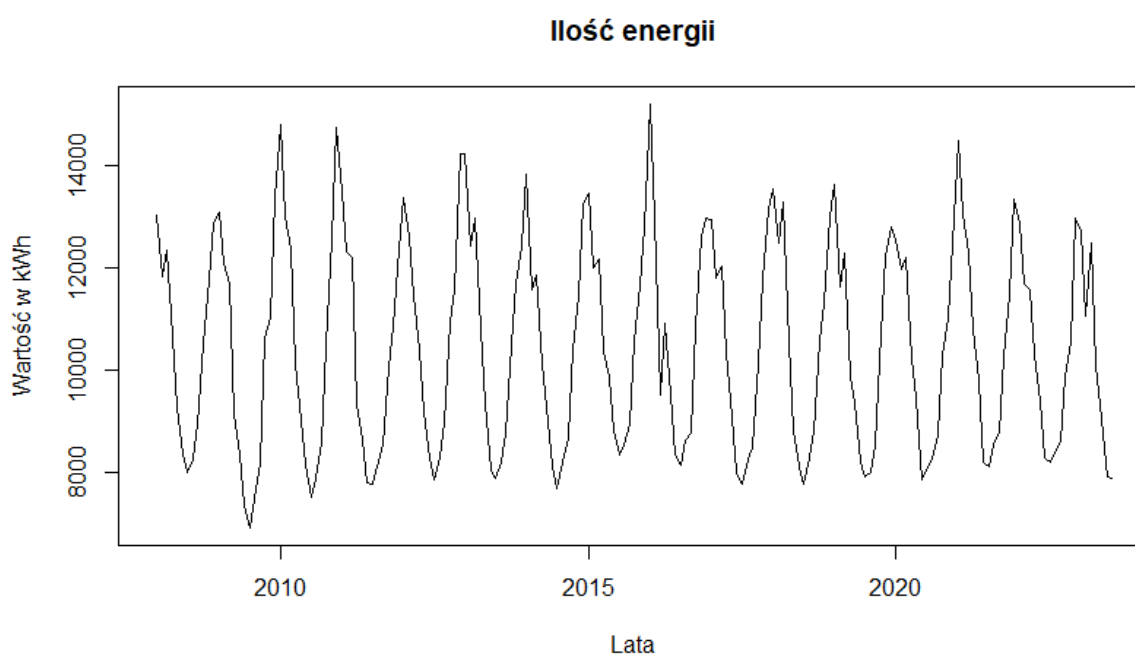


2. ANALIZA SZEREGU SEZONOWEGO - ENERGIA NA RYNKU WEW. NORWEGII

2.1. Dekompozycja dostępnej energii

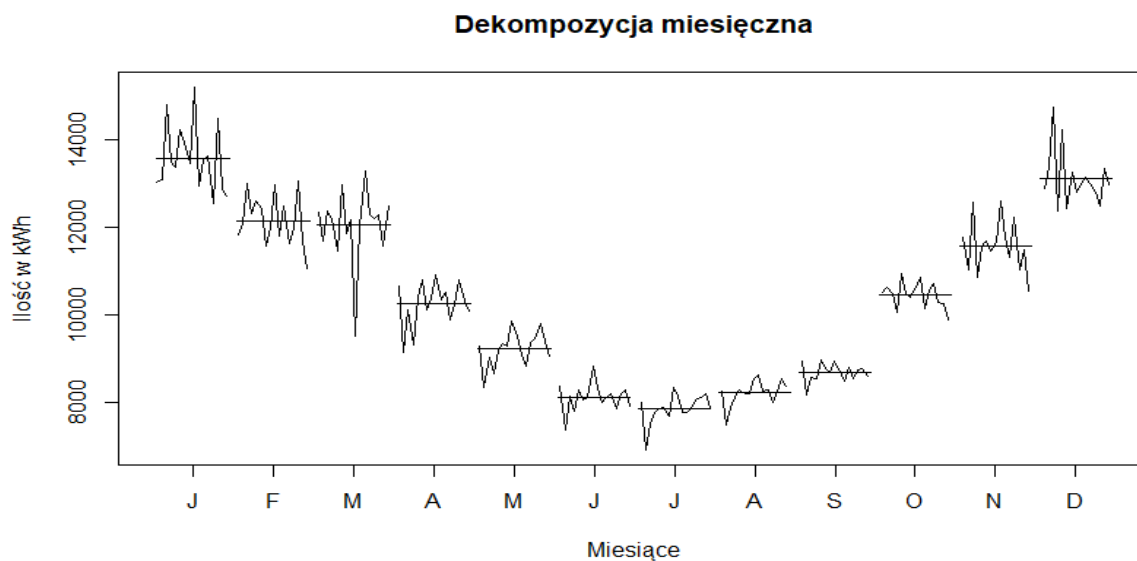
Drugim szeregiem analizowanym w niniejszej pracy jest ilość energii dostępnej na rynku wewnętrznym w Norwegii pomiędzy styczniem 2008 a lipcem 2023. Na poniższym wykresie przedstawione zostały wartości w danych miesiącach.

Rys. 10. Ilość energii dostępnej w Norwegii.

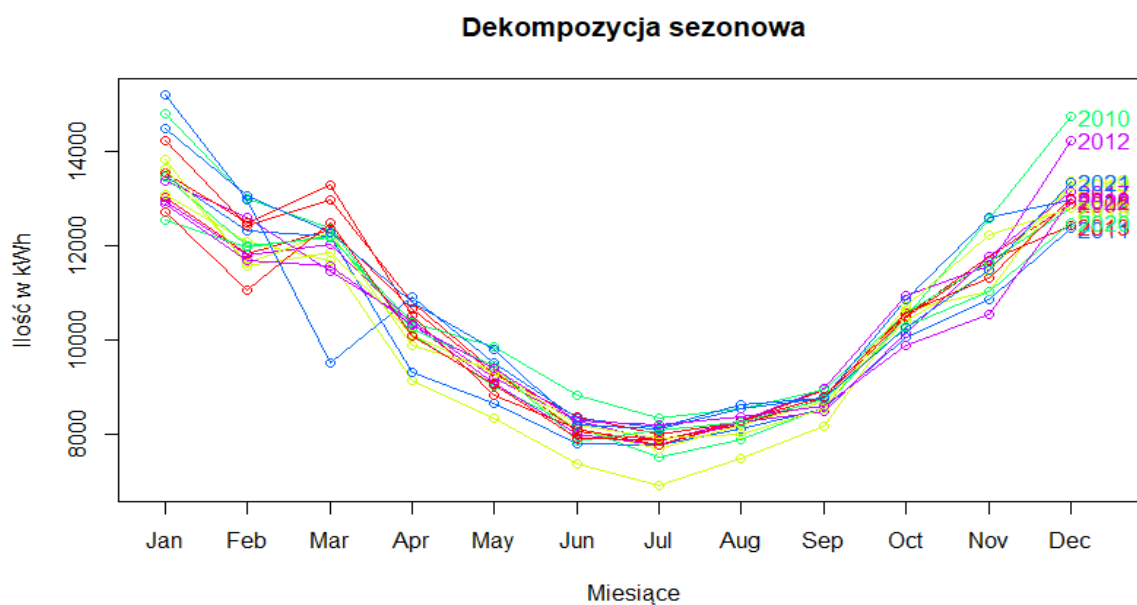


Na pierwszy rzut oka widoczna jest sezonowość, można zaobserwować brak generalnego trendu, ilość dostępnej energii wydaje się utrzymywać nad dość podobnym poziomie. Intuicyjne są wyższe wartości w miesiącach zimowych, można założyć że nadwyżka energii jest zużywana na cele grzewcze, w warunkach złej pogody ludzie preferują zostawać w swoich domach. Jak widać w sezonach letnich zapotrzebowanie na energię spada, może to świadczyć o mniejszej popularności sprzętów AGD służących do obniżania temperatury lub też ich wyższej efektywności.

Rys. 11. Dekompozycja miesięczna ilości energii dostępnej w Norwegii.



Rys. 12. Dekompozycja sezonowa ilości energii dostępnej w Norwegii.



Założenie o sezonowości szeregu potwierdzają obydwa wykresy. Na wykresie miesięcznym widać, że średnia wartość ilości energii dostępnej na rynku najniższa jest w miesiącach letnich, zaś najwyższa w przypadku miesięcy zimowych czyli od listopada do lutego. W przypadku dekompozycji sezonowej widać, że dane układają się w zbliżony do siebie sposób, bez znacznych obserwacji odstających. Zaobserwować można wcześniej wspomniany lekki trend malejący - lata późniejsze są bliżej osi X.

2.2. Model SARIMA

W przypadku szeregów z sezonowością do predykcji służy specjalny wariant modelu ARIMA czyli SARIMA (*Seasonal AutoRegressive Integrated Moving Average*). Tak jak w poprzednim rozdziale dane zostały sprawdzone za pomocą testów ADF i KPSS.

Tabela 9. Wyniki testów na stacjonarność szeregu czasowego energii.

Rodzaj testu	Wartość statystyki testowej	P-value	Interpretacja wyniku
ADF	-14.394	0.01	Stacjonarność
KPSS	0.01175	0.1	Stacjonarność

Na podstawie wyników testów Dickey-Fullera i KPSS można określić szereg mianem stacjonarnego. W pierwszym przypadku odrzucono hipotezę zerową o niestacjonarności szeregu, w drugim zaś brak jest podstaw do odrzucenia hipotezy o stacjonarności. Również w tym przypadku wyniki testu Boxa-Ljunga sugerują brak charakterystyk typowych dla białego szumu. Aby pozbyć się sezonowości zastosowano dwukrotne pojedyncze różnicowanie, pierwsze z opóźnieniem równym 12 okresów, drugie z równym 1.

Tabela 10. Wyniki testów na stacjonarność zróżnicowanego szeregu czasowego energii.

Rodzaj testu	Wartość statystyki testowej	P-value	Interpretacja wyniku
ADF	-8.1113	0.01	Stacjonarność
KPSS	0.015044	0.1	Stacjonarność

Tabela 11. Wynik testu Boxa-Ljunga.

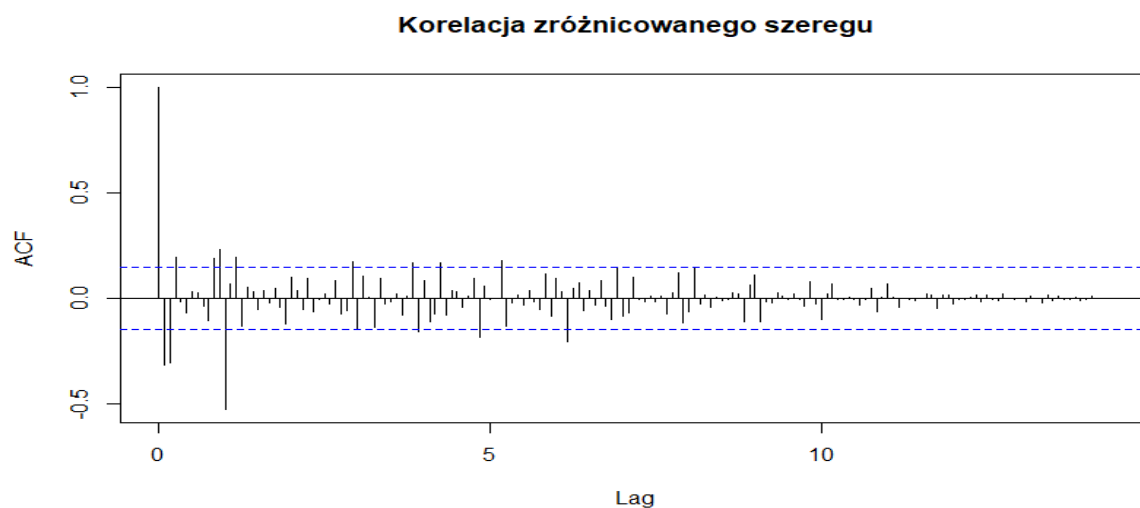
Rodzaj testu	Wartość statystyki testowej	P-value	Interpretacja wyniku
Box-Ljung	112.89	$< 2.2e-16$	Brak charakterystyki białego szumu

Po zróżnicowaniu szereg jest oczywiście dalej stacjonarny, nie posiada charakterystyki białego szumu, oraz nie wykazuje sezonowości.

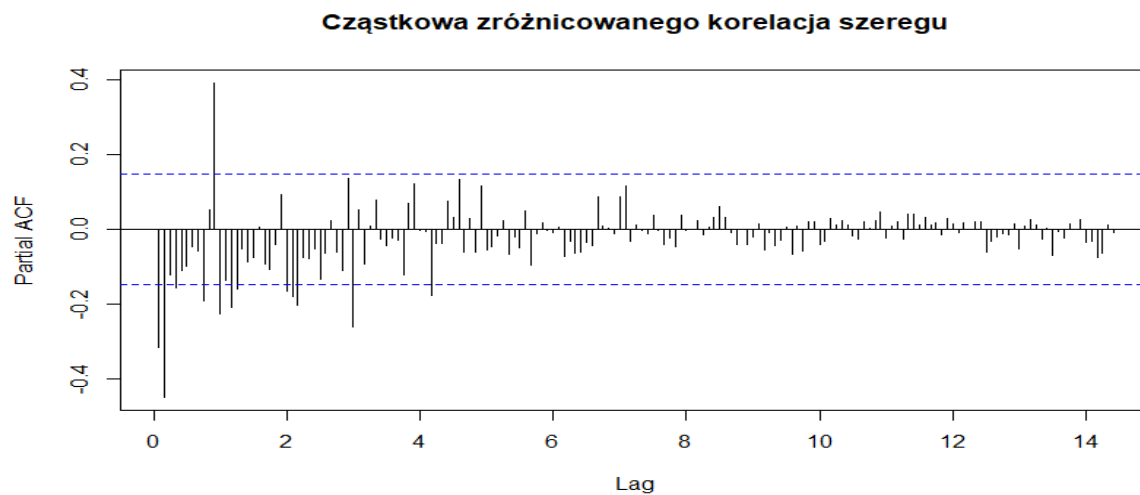
Rys. 13. Zróżnicowany szereg energii.



Rys. 14. Korelacja zróżnicowanego szeregu ilości energii dostępnej w Norwegii.



Rys. 15. Częstkowa korelacja zróżnicowanego szeregu ilości energii dostępnej w Norwegii.



Po analizie powyższych danych uzasadnione staje się użycie modelu SARIMA do prognozy szeregu czasowego z parametrem d równym 2 jako, że dwukrotnie została zastosowana metoda różnicowania. Dobranie pozostałych parametrów zrealizowano za pomocą analizy korelogramów widocznych poniżej. Został przeprowadzony automatyczny dobór parametrów, który został uwzględniony w poniższej tabeli jako pierwsza propozycja.

Tabela 12 Porównanie modeli SARIMA dla energii.

Model	AIC	BIC
SARIMA(1, 2, 0)(0,1,0) (auto)	2780.28	2786.44
SARIMA(2, 2, 0)(0,1,0)	2713.54	2722.78
SARIMA(2, 2, 1)(0,1,0)	2628.36	2640.68

Biorąc pod uwagę wyniki zarówno AIC i BIC wybrany został model SARIMA o parametrach (2, 2, 1)(0, 1, 0). Jego statystyki prezentują się następująco:

Tabela 13. Błędy prognozy modelu SARIMA dla energii.

Statystyka	Wartość
RMSE	475.9605
MAE	393.2544
MAPE	3.792634
MASE	0.6816050

Rys. 16. Predykcja za pomocą modelu SARIMA dla energii.

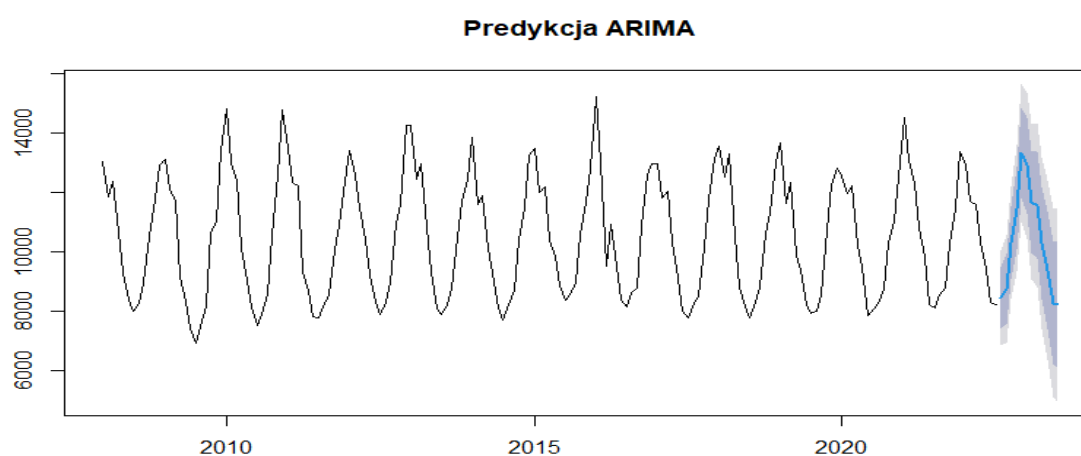


Tabela 14. Wyniki predykcji modelu SARIMA dla energii.

Data	Predykcja	low 80	high 80	low 95	high 95
Aug 2022	8438.172	7405.221	9471.124	6858.409	10017.94
Sep 2022	8755.826	7569.643	9942.010	6941.715	10569.94
Oct 2022	10289.266	9059.393	11519.140	8408.337	12170.20
Nov 2022	11453.635	10063.962	12843.309	9328.313	13578.96
Dec 2022	13335.627	11811.135	14860.118	11004.118	15667.14
Jan 2023	12904.801	11301.190	14508.411	10452.290	15357.31
Feb 2023	11677.613	9975.759	13379.467	9074.851	14280.37
Mar 2023	11564.374	9759.167	13369.581	8803.548	14325.20
Apr 2023	10301.335	8410.823	12191.848	7410.047	13192.62
May 2023	9386.137	7411.262	11361.012	6365.827	12406.45
Jun 2023	8270.829	6209.398	10332.260	5118.142	11423.52
Jul 2023	8204.888	6062.343	10347.432	4928.149	11481.63

2.3. Model Holta-Wintersa

Drugim modelem predykcyjnym wykorzystanym w analizie ilości energii na rynku wewnętrznym Norwegii był model multiplikatywny Holta-Wintersa. Został on wybrany kosztem addytywnego ze względu na lepsze wyniki.

Tabela 15. Błędy prognozy dla modelu Holta dla IBM.

Statystyka	Wartość
RMSE	581.4851
MAE	430.1163
MAPE	4.0119
MASE	0.7455

Rys. 17. Predykcja za pomocą modelu Holta dla energii.

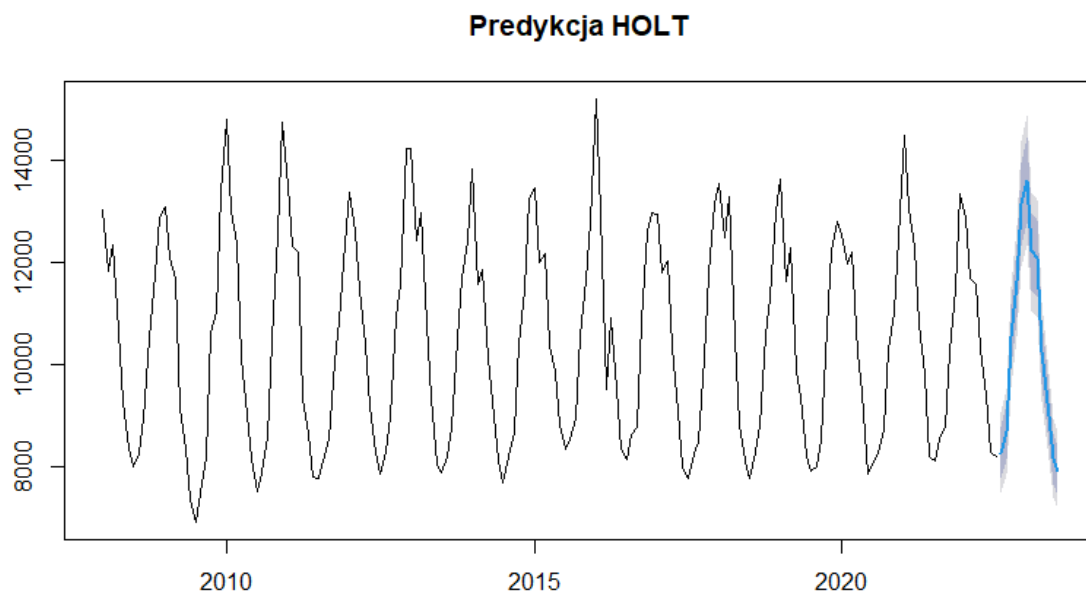


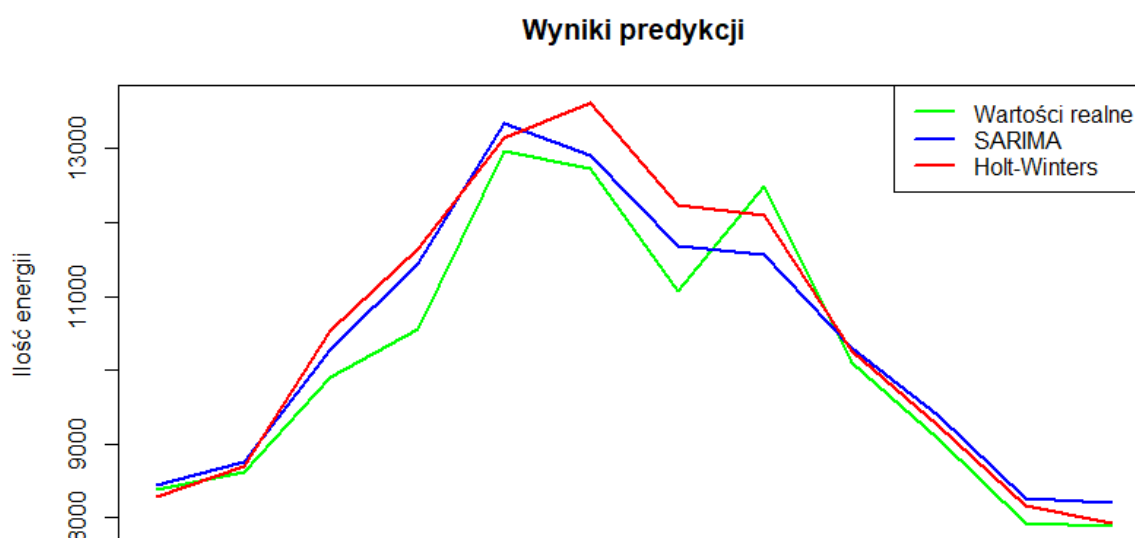
Tabela 8. Predykcje Holta dla energii.

Data	Predykcja	low 80	high 80	low 95	high 95
Aug 2022	8279.272	7771.012	8787.532	7501.955	9056.588
Sep 2022	8696.786	8162.660	9230.911	7879.911	9513.660
Oct 2022	10544.668	9896.751	11192.585	9553.764	11535.571
Nov 2022	11643.470	10927.686	12359.255	10548.772	12738.168
Dec 2022	13150.258	12341.424	13959.092	11913.253	14387.263
Jan 2023	13625.442	12786.923	14463.962	12343.038	14907.847
Feb 2023	12232.427	11479.202	12985.652	11080.469	13384.386
Mar 2023	12106.113	11360.215	12852.010	10965.361	13246.864
Apr 2023	10264.621	9631.783	10897.459	9296.779	11232.464
May 2023	9258.470	8687.284	9829.656	8384.917	10132.024
Jun 2023	8171.519	7667.040	8675.998	7399.984	8943.053
Jul 2023	7919.559	7430.280	8408.839	7171.271	8667.848

2.4. Porównanie modeli

Na poniższym wykresie można zauważyć niewielką różnicę w predykcjach modeli. Metoda SARIMA wydaje się nieco lepiej oddawać trend szeregu, jej wartości predykcji przez większość czasu są bliższe wartościom realnym, choć różnica między predykcjami Holta-Wintersa nie jest duża

Rys. 18. Porównanie wyników dla energii.



3. Podsumowanie

Po raz kolejny potwierdza się trudność przewidywania akcji firm na giełdzie, gdyby tak nie było na pewno już dawno ktoś by to wykorzystał. Metoda Holta okazała się bezużyteczna w przypadku akcji IBM, metoda ARIMA przewidziała poprawnie trend jednak ze znacznym odchyłem w stosunku do wartości rzeczywistych. W przypadku szeregu sezonowego obie metody wykazały zaskakująco dokładne prognozy, z minimalną przewagą SARIMY.

Oświadczenie autora pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca została napisana samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Data

01.09.23

Podpis autora pracy

Andrzej Zimnik