

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Zoia Romas

Nr albumu: 443827

Analiza szeregów czasowych

Projekt zaliczeniowy na zajęcia
z Analizy Szeregów Czasowych, gr. nr 304
prowadzone przez dr Piotra Żocha
Semestr letni 2023/2024

Warszawa, czerwiec 2024

Oświadczenie autora pracy

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

Data

16.06.2024

Podpis autora pracy

Loia Romas

WSTĘP

W pracy przedstawiono analizę dwóch szeregów czasowych pochodzących ze strony Federal Reserve Economic Data (FRED). Pierwszy z nich jest szeregiem o charakterze niesezonowym. Dotyczy miesięcznych zmian światowej ceny oliwy z oliwek w dolarach amerykańskich za tonę od stycznia 2000 roku do grudnia 2023 roku. Drugi szereg jest szeregiem sezonowym i przedstawia zmienność wartości miesięcznych dla sprzedaży detalicznej w kategoriach takich jak meble, wyposażenie domu, elektronika i sprzęt AGD w milionach dolarów amerykańskich, również dla okresu od stycznia 2000 roku do grudnia 2023.

Dla każdego z przedstawionych szeregów dokonana została dekompozycja, następnie dopasowany został model z klasy ARIMA (w przypadku szeregu niesezonowego) lub SARIMA (w przypadku szeregu sezonowego) po przeprowadzeniu niezbędnych elementów analizy, a także dopasowano modele ekstrapolacyjne. W ramach oceny uzyskanych prognoz przeprowadzono porównanie, zarówno empiryczne, jak i przy pomocy odpowiednich wskaźników. Każdy z szeregów został podzielony na część in-sample oraz out-of-sample w celu zbadania jakości prognozy. W przypadku szeregu sezonowego próbka out-of-sample wyniosła 12 najnowszych obserwacji (pełny cykl sezonowy), natomiast dla szeregu niesezonowego próbka ta wyniosła 3 najnowsze obserwacje (3 miesiące). Na potrzeby niniejszej pracy założono poziom istotności na poziomie 5%. Analiza szeregów czasowych zostało wykonane przy użyciu Demetra oraz środowiska RStudio wraz z dostępnymi w nim bibliotekami.

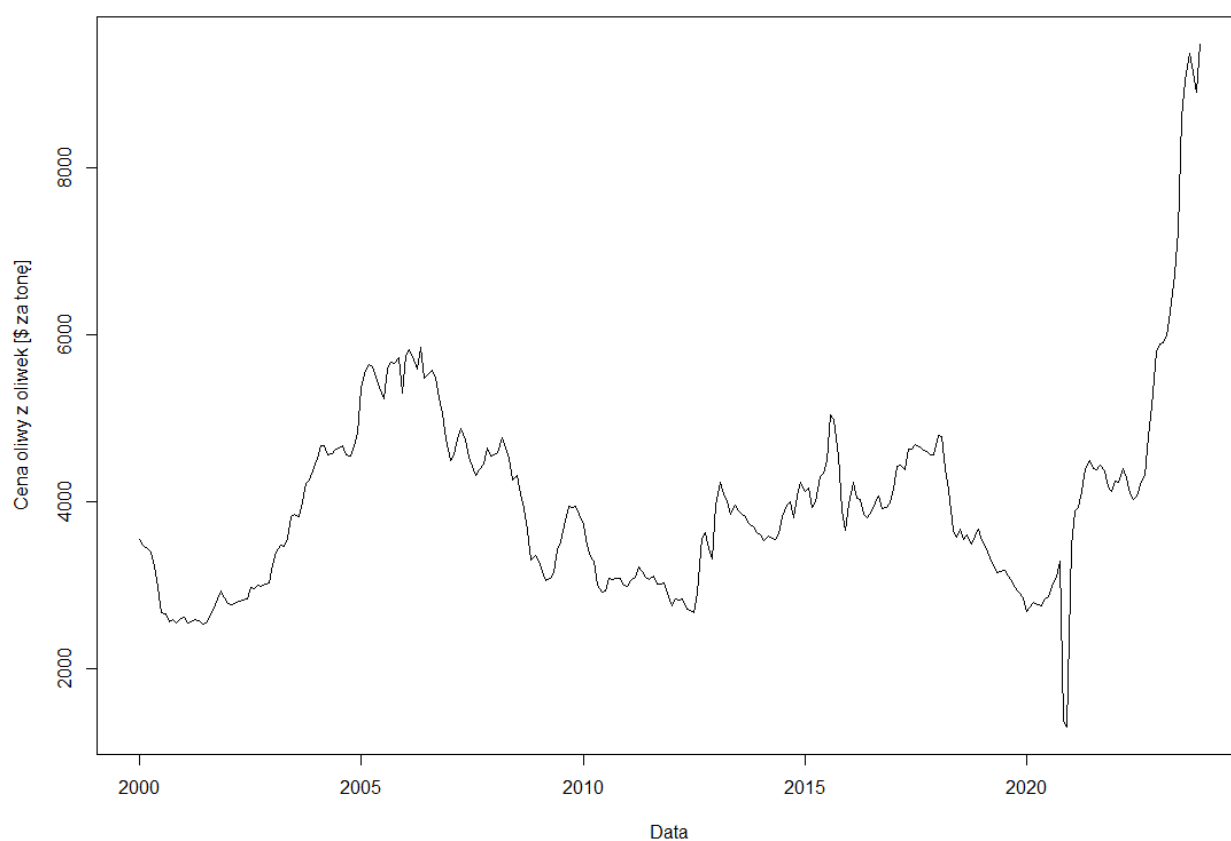
ROZDZIAŁ I

Szereg niesezonowy

1.1. Opis szeregu

Wybrany szereg niesezonowy przedstawia zmienność światowej ceny oliwy z oliwek. Źródłem tych danych jest baza danych Federal Reserve Economic Data (FRED). Badany szereg obejmuje łącznie 288 obserwacji. Pierwsza obserwacja występuje w styczniu 2000 roku, natomiast ostatnie przypada na grudzień 2023 roku. Wartości obserwacji wyrażone są w dolarach amerykańskich za tonę oliwy z oliwek. Na Rys. 1 przedstawiony został wykres obserwacji w czasie.

Rys. 1. Wykres szeregu niesezonowego

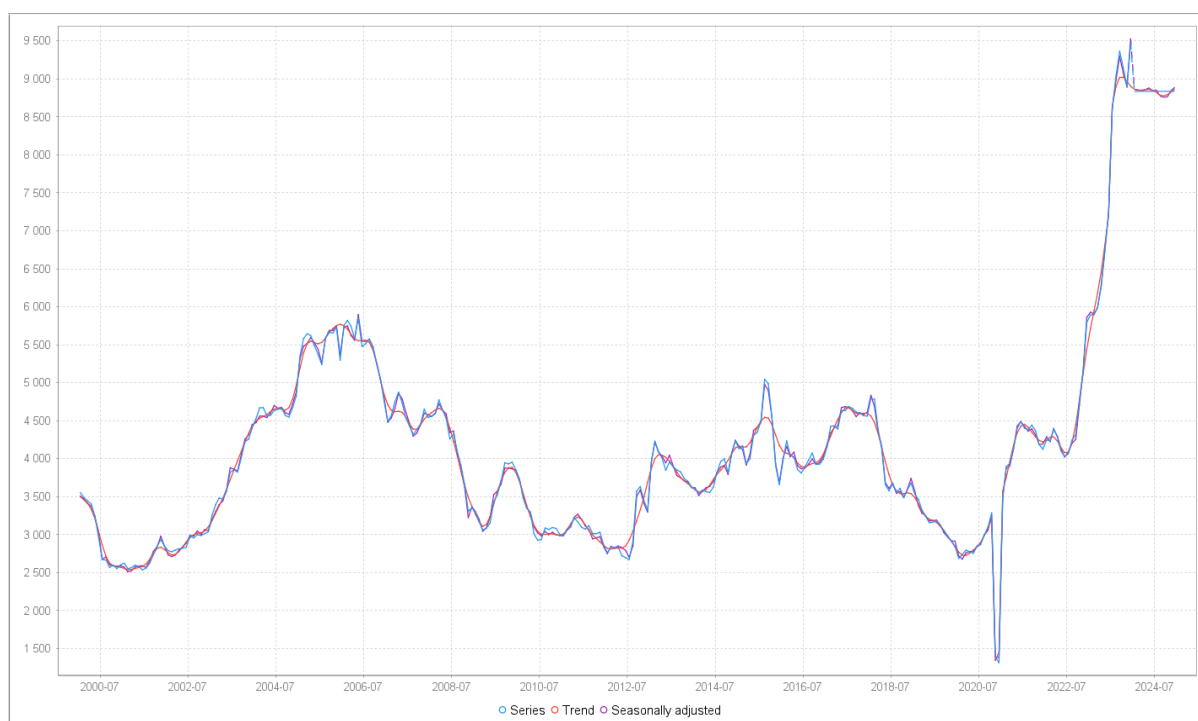


Źródło: Opracowanie własne przy pomocy RStudio.

1.2. Dekompozycja szeregu czasowego

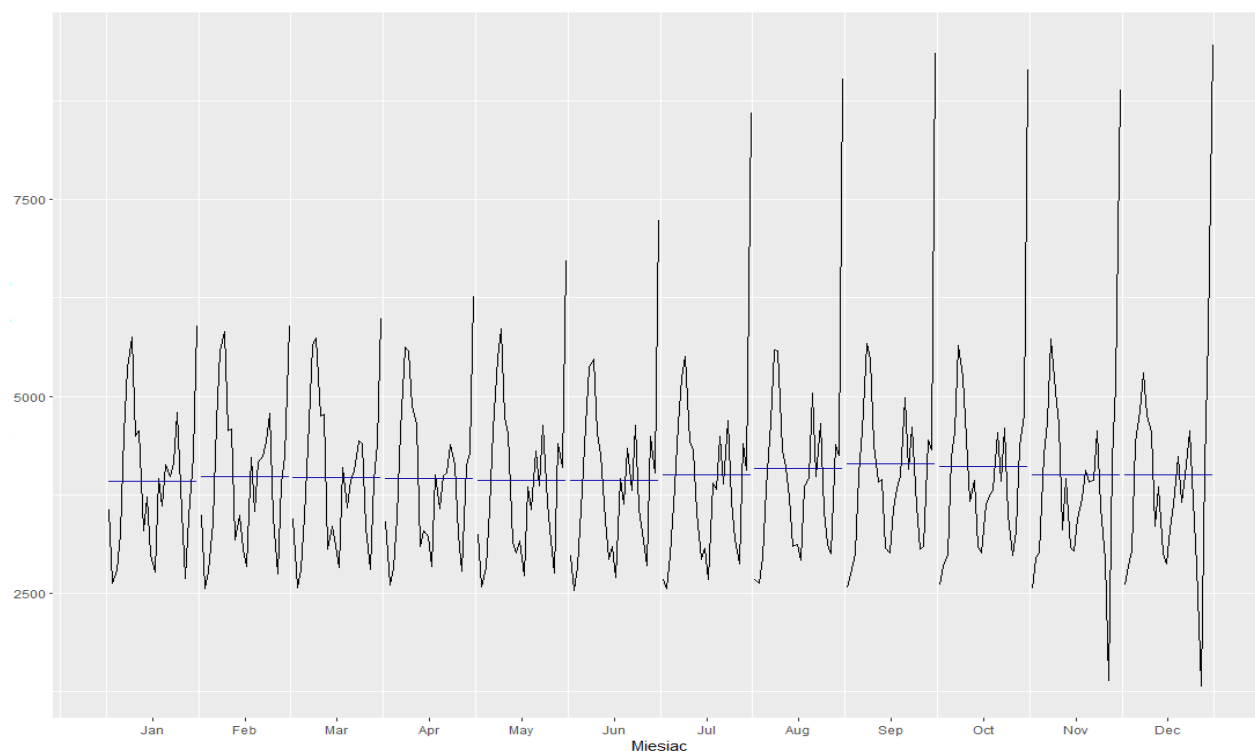
Na podstawie Rys. 1 nie można określić jednorodnego trendu, co potwierdza dekompozycja szeregu przedstawiona na Rys. 2 wykonana przy użyciu programu JDemetra+. Nie widać także wyraźnej sezonowości zgodnie z dekompozycją według miesięcy na Rys. 3. Potwierdza to analiza szeregu w programie JDemetra+, zgodnie z którą przeprowadzone testy w kierunku sezonowości nie potwierdziły jej występowania. Analiza wykazała również sześć obserwacji odstających, z czego trzy o charakterze długotrwałym (level shift), które wystąpiły w listopadzie 2020 roku, styczniu 2021 roku i lipcu 2023 roku. Pozostałe trzy obserwacje odstające miały charakter jednorazowy (additive outlier), wystąpiły w grudniu 2005 roku, maju 2006 roku i grudniu 2023 roku.

Rys. 2. Dekompozycja szeregu niesezonowego



Źródło: Opracowanie własne przy pomocy JDemetra+.

Rys. 3. Dekompozycja szeregu nie sezonowego według poszczególnych miesięcy



Źródło: Opracowanie własne przy pomocy RStudio.

1.3. Model ARIMA

W celu budowy modelu, który pozwala na prognozowanie analizowanego szeregu niesezonowego posłużono się modelem ARIMA (ang. Autoregressive Integrated Moving Average), zbudowanego przy wykorzystaniu procedury Boxa-Jenkinsa (identyfikacja – estymacja – diagnostyka – prognozowanie). Model ten składa się z procesów: autoregresyjnego $AR(p)$, średniej ruchomej $MA(q)$ oraz stopnia integracji $I(d)$. Do przeprowadzenia prognozy obserwacje w szeregu niesezonowym muszą zostać podzielone na okresy in-sample oraz out-of-sample. W niniejszej pracy na potrzeby analizy próbka out-of-sample dla szeregu o charakterze niesezonowym stanowi obserwacje dla 3 ostatnich miesięcy dostępnych w próbie danych ze względu na to, że model ARIMA może być wykorzystywany jedynie do prognozowania krótkoterminowego. Głównym założeniem, które musi być spełnione dla tego modelu to stacjonarność badanego szeregu, należy więc sprawdzić to w pierwszej kolejności przy pomocy odpowiednich testów formalnych oraz analizy graficznej korelogramów.

1.3.1. Stacjonarność szeregu

Do weryfikacji stacjonarności szeregu posłużono się w pierwszej kolejności testem Dickey-Fullera (DF) w trzech wersjach – podstawowej, błędzenie losowe z dryfem oraz z trendem deterministycznym. Hipoteza zerowa testu DF niezależnie od wersji stwierdza, że weryfikowana zmienna jest błędzeniem przypadkowym (czyli nie wykazuje stacjonarności). Na podstawie uzyskanych wyników w każdym z testów nie ma podstaw do odrzucenia hipotezy zerowej o niestacjonarności szeregu, gdyż statystyki testowe są dodatnie, wynoszą 1,58, 1,1647, 0,8394 (przy wartościach krytycznych na poziomie -1,95, -2,87, -3,42) odpowiednio dla klasycznego testu DF, z dryfem oraz z dryfem i trendem. W przypadku ostatniego testu statystyki trendu wskazuje na możliwą obecność trendu, ale dopiero na poziomie istotności 10%.

Ze względu na występującą autokorelację reszt nie można wyciągnąć wniosku na podstawie testu DF, w związku z czym zastosowano jego rozszerzoną wersję – ADF. Do weryfikacji występowania autokorelacji reszt wykorzystano test Breuscha-Godfrey. Wyniki wspomnianych testów zostały przedstawione w Tabeli 1. Zastosowany został wariant testu ze stałą. Zgodnie z Tabelą 1, dla DF ($k=0$) występuje autokorelacja reszt, zatem wymagane jest wykorzystanie testu ADF. Na podstawie wyników zauważyć można, że problem autokorelacji wyeliminowany został dla rozszerzenia w wysokości 2. Przy statystyce testowej na poziomie 0,875 dla $k=2$ (brak autokorelacji reszt) w teście ADF wartość krytyczna na poziomie istotności 5% wyniosła -1,95. Zatem nie ma podstaw do odrzucenia hipotezy zerowej stwierdzającej, że szereg jest niestacjonarny. Wymagane jest zatem różnicowanie zmiennej.

Tabela 1. Wyniki przeprowadzonych testów DF i ADF dla szeregu niesezonowego

Liczba opóźnień k	Test DF/ADF		p-value dla danego rzędu korelacji w teście B-G			
	Statystyka testowa	p-value	1	2	3	4
0	1.1647	0.990	4.91e-05	3.76e-07	3.77e-07	1.38e-07
1	-0.1665	0.935	3.72e-01	5.71-04	1.14e-04	2.96e-03
2	0.8751	0.990	0.695	0.851	0.298	0.416

Źródło: Opracowanie własne.

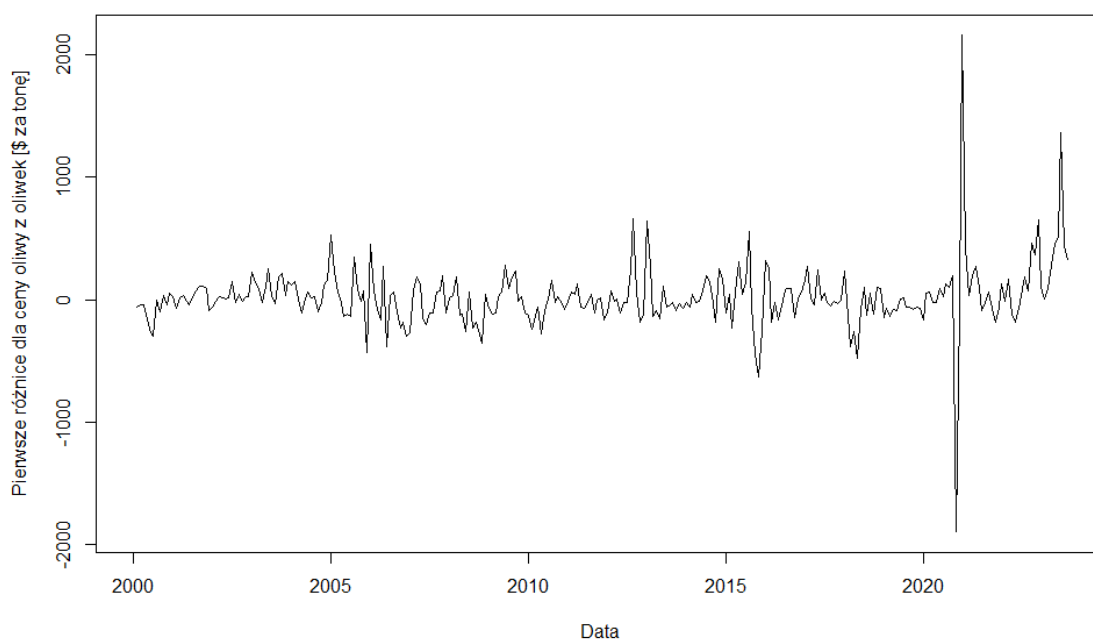
Wyniki tych samych testów DF i ADF przeprowadzonych ponownie, tym razem dla pierwszych różnic zmiennej przedstawione zostały w Tabeli 2. Możemy zauważyć, że w przypadku pierwszego zróżnicowania szeregu już przy pierwszym opóźnieniu ($k=1$) mamy brak autokorelacji reszt, a statystyka testowa testu ADF na poziomie -12,4262 (wartość krytyczna = -1,95) pozwala nam odrzucić hipotezę zerową o niestacjonarności. Jak widać z wyników pierwsze zróżnicowanie zmiennej niesezonowej wystarczyło, aby otrzymać szereg stacjonarny. Możemy więc kontynuować analizę. Wykres przedstawiający szereg po pierwszym różnicowaniu przedstawiony został na Rys. 4.

Tabela 2. Wyniki przeprowadzonych testów DF i ADF dla pierwszych różnic zmiennej

Liczba opóźnień k	Test DF/ADF		p-value dla danego rzędu korelacji w teście B-G			
	Statystyka testowa	p-value	1	2	3	4
0	-128034	< 0.01	0.4068	5.74e-04	1.17e-03	3.06e-03
1	-12.4262	< 0.01	0.7485	0.8167	0.2494	0.3337
2	-8.2699	< 0.01	0.9015	0.9760	0.9972	0.9989

Źródło: Opracowanie własne.

Rys. 4. Wykres zmiennej niesezonowej po pierwszym różnicowaniu



Źródło: Opracowanie własne przy pomocy RStudio.

Wniosek co do stacjonarności szeregu potwierdza również wynik testu KPSS, dla którego wartość statystyki testowej wyniosła 0,378 przy wartości krytycznej 0,463 na poziomie istotności 5%. Nie ma więc podstaw do odrzucenia hipotezy zerowej stwierdzającej stacjonarność szeregu. Na podstawie przeprowadzonych testów (ADF i KPSS) możemy stwierdzić, że zmienna ta jest zintegrowana rzędu 1 ($d=1$).

1.3.2. Identyfikacja, estymacja oraz diagnostyka

Po analizie przeprowadzonej w poprzednim podrozdziale wiemy, że szereg jest stacjonarny po zastosowaniu pierwszych różnic. Parametr d modelu ARIMA wynosi więc 1, ponieważ analizowany szereg jest zintegrowany rzędu 1. Aby przeprowadzić procedurę Boxa-Jenkinsa w celu wybrania odpowiednich p i q do modelu ARIMA (p, d, q) należy sprawdzić, czy proces nie jest białym szumem przy pomocy testów Ljunga-Boxa i Boxa-Pierce'a. Na podstawie wyników wspomnianych testów przedstawionych w Tabeli 3 odrzucamy hipotezę zerową, która mówi o tym, że proces jest białym szumem.

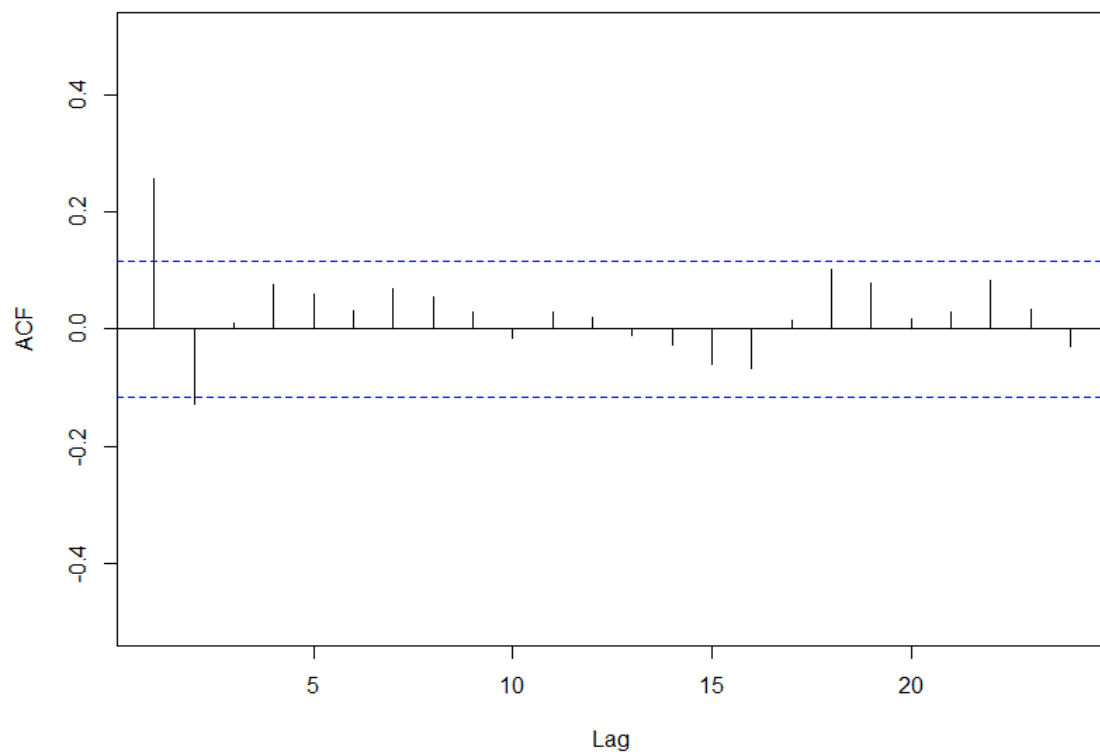
Tabela 3. Test Ljunga-Boxa i Boxa-Pierce'a dla zróżnicowanego szeregu niesezonowego

Test	Statystyka testowa	p-value	Wniosek
Ljunga-Boxa	40.74	0.01778	Proces nie jest białym szumem.
Boxa-Pierce'a	39.49	0.02419	Proces nie jest białym szumem.

Źródło: Opracowanie własne.

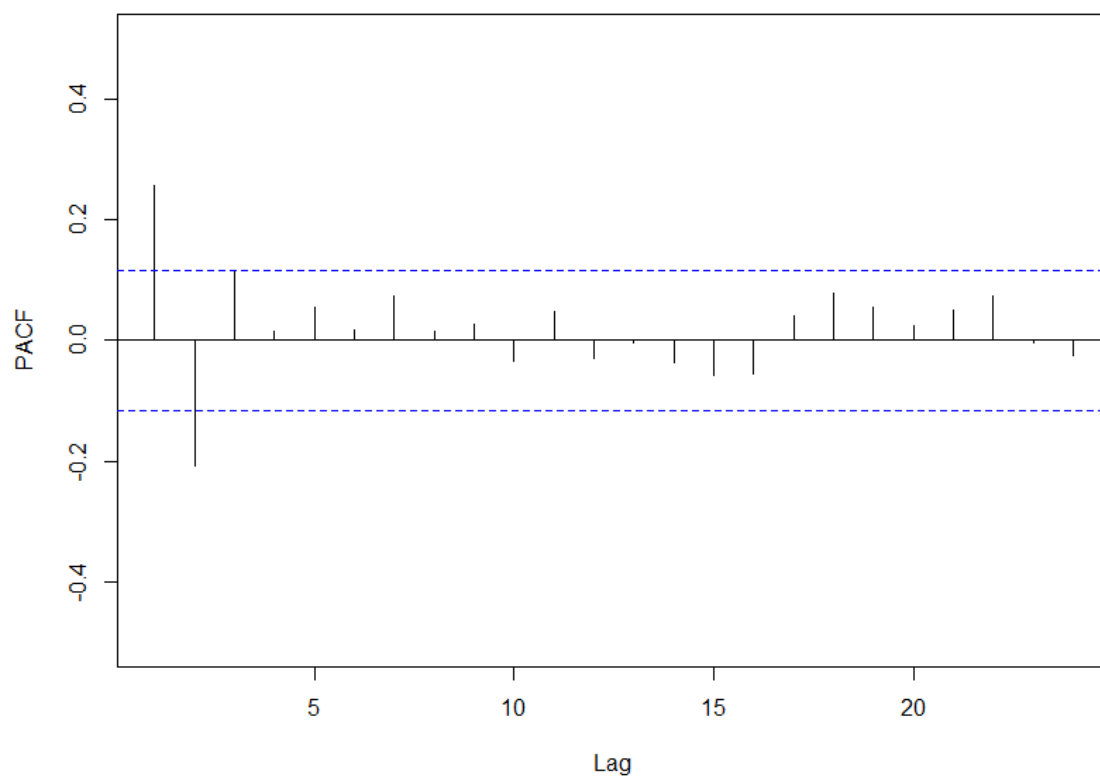
Aby zweryfikować uzyskane wyniki przy pomocy formalnych testów należy przyjrzeć się wykresom ACF i PACF dla zróżnicowanego szeregu. Na podstawie Rys. 5 i Rys. 6 można potwierdzić wyniki uzyskane na skutek przeprowadzenia testów – zarówno na wykresie ACF, jak i PACF mamy istotne wypustki dla zróżnicowanej zmiennej. Wykresy posłużą również do ustalenia parametru p , czyli rzędu procesu AR, oraz q , czyli rzędu procesu MA. Na podstawie wykresu ACF, który służy do wyznaczenia rzędu procesu MA, można zweryfikować q na poziomie 0, 1 lub 2. Z kolei do ustalenia wartości parametru p posłużono się wykresem PACF, który wskazuje, że także może on przyjąć wartość od 0 do 2.

Rys. 5. Wykres funkcji ACF



Źródło: Opracowanie własne przy pomocy RStudio.

Rys. 6. Wykres funkcji PACF



Źródło: Opracowanie własne przy pomocy RStudio.

Zatem wartości przedstawione na wykresach ACF i PACF sugerują model ARIMA(2,1,2). Występuje natomiast ryzyko, że na którymś z wykresów (PACF lub ACF) widzimy nie wynik dla danego procesu, odpowiednio autoregresji lub średniej ruchomej, tylko efekt tego drugiego procesu. Może się więc okazać, że parametr p lub q powinien wynosić 0. W celu znalezienia optymalnego modelu przetestowane zostały wszystkie kombinacje dla wartości $p = \{0,1,2\}$ oraz $q = \{0,1,2\}$. Wśród przetestowanych kombinacji istotnymi statystycznie okazały się modele ARIMA(p, d, q) takie, gdzie występuje jeden wybrany proces MA(q) lub AR(p) oraz ARIMA(1,1,1), przy czym proces MA(1) w tym przypadku ma niższe p -value ($8.58e-09$), niż AR(1) z p -value na poziomie 0,03. Aby lepiej dopasować postać modelu powyższe kombinacje zostaną porównane w wyniku dalszej analizy na podstawie kryteriów informacyjnych oraz metody od ogólnego do szczegółowego. Wykonane zostały również ponownie testy Ljunga-Boxa i Boxa-Pierce'a dla powyższych modeli. Dane dotyczące ostatecznie analizowanych modeli przedstawione zostały w Tabeli 4.

W przypadku testów Ljunga-Boxa oraz Boxa-Pierce'a mają one na celu weryfikację założenia, że reszty w analizowanych modelach są białym szumem. Jak można zauważyć, reszty w każdym z badanych modeli miały charakter białego szumu, co także potwierdziła graficzna analiza wykresów funkcji ACF i PACF dla każdego z nich. Nie możemy więc na podstawie wyników tych testów uznać któryś z przetestowanych modeli za nieodpowiedni.

Tabela 4. Wybrane modele ARIMA – porównanie

	AIC	BIC	Test Ljunga-Boxa		Test Boxa-Pierce'a		Log likelihood
			Statystyka testowa	p-value	Statystyka testowa	p-value	
ARIMA(1,1,1)	3934.942	3949.538	11.856	0.9815	11.222	0.9873	-1963.47
ARIMA(0,1,1)	3937.571	3948.518	15.058	0.9191	14.425	0.9365	-1965.79
ARIMA(0,1,2)	3934.534	3949.130	11.59	0.9841	10.956	0.9893	-1963.27
ARIMA(1,1,0)	3948.562	3959.509	27.018	0.3036	26.109	0.3477	-1971.28
ARIMA(2,1,0)	3938.197	3952.793	14.81	0.9262	14.114	0.944	-1965.1

Źródło: Opracowanie własne.

Możemy natomiast kierować się wartościami kryterium informacyjnym Akaike (AIC) oraz Bayesowskim kryterium Schwarza (BIC) dla każdego z modeli i na tej podstawie wybrać

najlepiej dopasowaną postać modelu. Zgodnie z powyższym, w oparciu o wartości AIC i BIC w Tabeli 4 można stwierdzić, że najlepsze modele to ARIMA(0,1,1), zgodnie z kryterium BIC, oraz ARIMA(0,1,2), zgodnie z kryterium BIC. Dla tych postaci kryteria informacyjne osiągają najniższe wartości, nie pozwalają natomiast jednoznacznie wskazać najlepszego modelu. Wyniki sugerują jednak, że proces można w znacznej mierze opisać przy pomocy komponentu MA(q) (gdzie $q=1$ lub $q=2$), czyli procesu średnich ruchomych. Dodatkowo, nieznacznie odróżnia się postać modelu ARIMA(1,1,1) pod względem AIC i BIC, jeżeli chodzi o kryteria informacyjne, więc należy przyrzeć się pozostałym wskaźnikom (błędem prognozy) dla tej opcji w dalszych składowych analizy porównawczej.

Aby porównać modele ARIMA(0,1,1) i ARIMA(0,1,2) warto dodatkowo zastosować metodę od ogólnego do szczegółowego, gdzie odpowiednio to są modele z restrykcjami i bez. W Tabeli 5 przedstawione są wartości potrzebne do obliczenia odpowiedniej statystyki testowej, która wynosi 5,037. Wartość krytyczna natomiast wynosi 3,841, co świadczy o tym, że odrzucamy hipotezę zerową mówiącą o tym, że prostszy model, czyli ARIMA(0,1,1) jest tak samo odpowiedni do prognozowania jak model ogólny – ARIMA(0,1,2). Wywnioskować więc można, że model ARIMA(0,1,2) jest najlepiej dopasowanym modelem do wybranego szeregu czasowego zgodnie z dotychczasową analizą.

Tabela 5. Model od ogólnego do szczegółowego

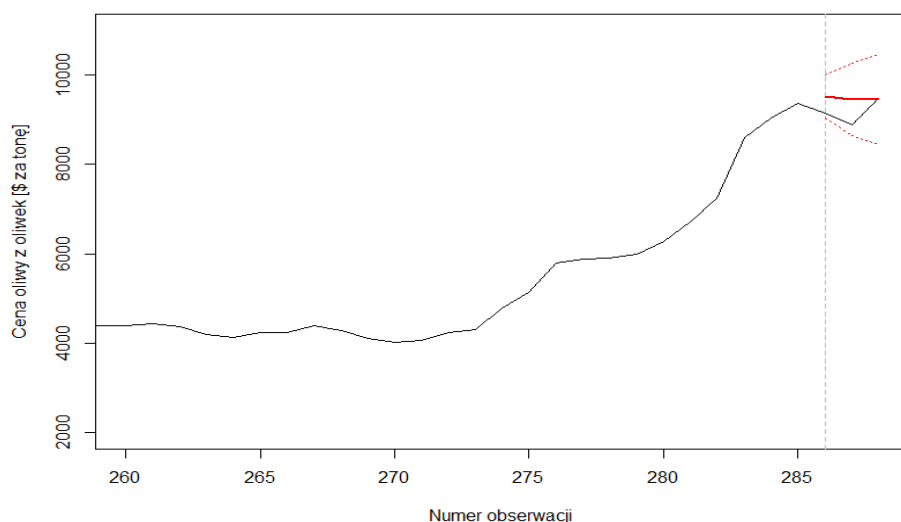
	Log likelihood	Statystyka chi-kwadrat (df=1)	Wartość krytyczna (df=1)	p-value
ARIMA(0,1,1)	-1965.79	5.037157	3.841	0.0248
ARIMA(0,1,2)	-1963.27			

Źródło: Opracowanie własne.

1.3.3. Prognozowanie

Po znalezieniu optymalnych wartości parametrów p oraz q w modelu ARIMA można przejść do etapu prognozowania szeregu. Prognozowane będą trzy obserwacje, które zostały usunięte z danych w celu uzyskania próbki in-sample. Wykres obrazujący prognozę modelu ARIMA(0,1,2) przedstawiony został na Rys. 7. Wykres przedstawia jedynie grupę najbardziej aktualnych obserwacji w celu łatwiejszego odczytania otrzymanych prognoz.

Rys. 7. Prognoza dla szeregu niesezonowego – model ARIMA(0,1,2)



Źródło: Opracowanie własne przy pomocy RStudio.

Następnie w celu zweryfikowania jakości oszacowań policzone zostały wartości MAE (ang. Mean Absolute Error), MSE (ang. Mean Square Error), MAPE (ang. Mean Absolute Percentage Error) oraz AMAPE (ang. Adjusted Mean Absolute Percentage Error), których wyniki przedstawione zostały w Tabeli 6. Zdecydowano się na weryfikację wszystkich błędów ex post, choć w przypadku analizowanego szeregu teoretycznie najlepszy byłby MAE (wartości obserwacji są co do modułu większe od 1 oraz występują obserwacje odstające). Jak widać na podstawie przedstawionej Tabeli 6, model ARIMA(0,1,2) okazał się być najlepszy pod względem jakości prognozy, choć kryterium informacyjne BIC wskazywało na to, że ARIMA(0,1,1) był lepszym modelem do prognozowania. Widzimy, że zarówno na podstawie błędu MAE, jak i pozostałych, ARIMA(0,1,2) ma najniższe wartości, jest więc najlepiej dopasowanym modelem do prognozowania.

Tabela 6. Wartości błędów prognozy ex post analizowanych modeli ARIMA

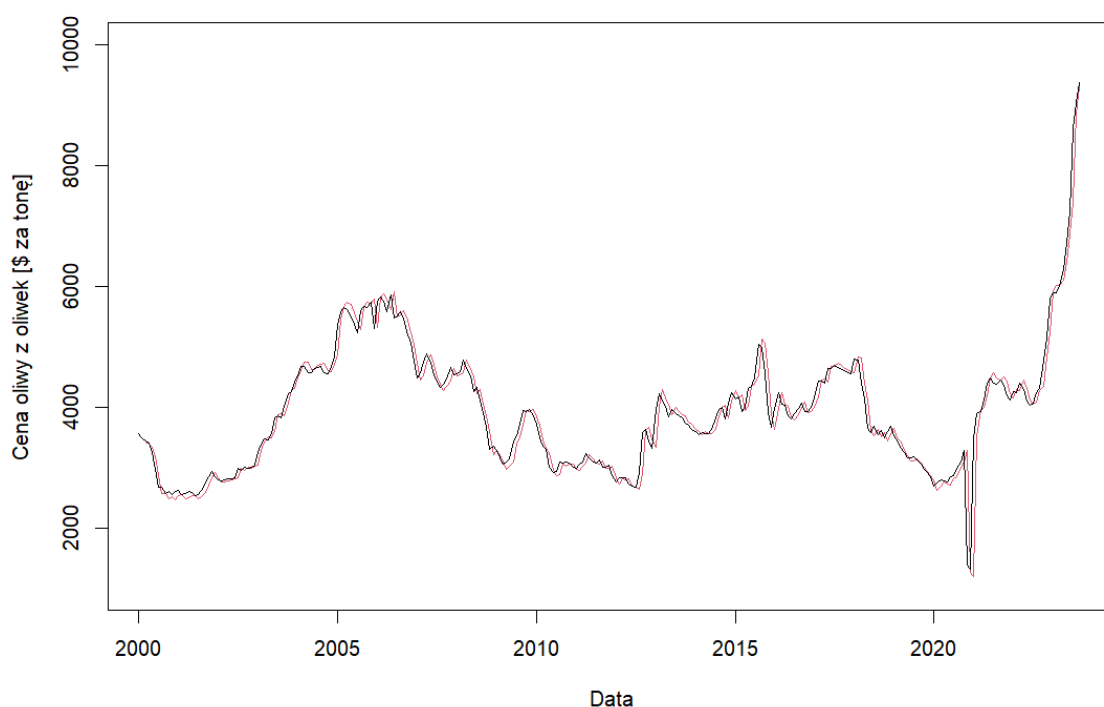
	ARIMA(0,1,2)	ARIMA(0,1,1)	ARIMA(1,1,1)
MAE	318.329	333.8532	350.8018
MSE	155818.2	167312.4	176977.9
MAPE	3.539255 %	3.710038 %	3.893433 %
AMAPE	3.445072 %	3.608817 %	3.786679 %

Źródło: Opracowanie własne.

1.4. Model ekstrapolacyjny

Jako kolejną metodę prognozowania szeregu czasowego wykorzystano model ekstrapolacyjny. W tym przypadku do analizy wykorzystany został model liniowy Holta, czyli podwójne wygładzanie wykładnicze, ze względu na potencjalne występowanie trendu. Dodatkowo, ze względu na niepewność co do występowania trendu, wykonano również prognozę metodą prostego wygładzania wykładniczego (SES), aby porównać te metody w kontekście błędów ex post dla danego szeregu.

Rys. 8. Szereg niesezonowy wygładzony metodą liniową Holta



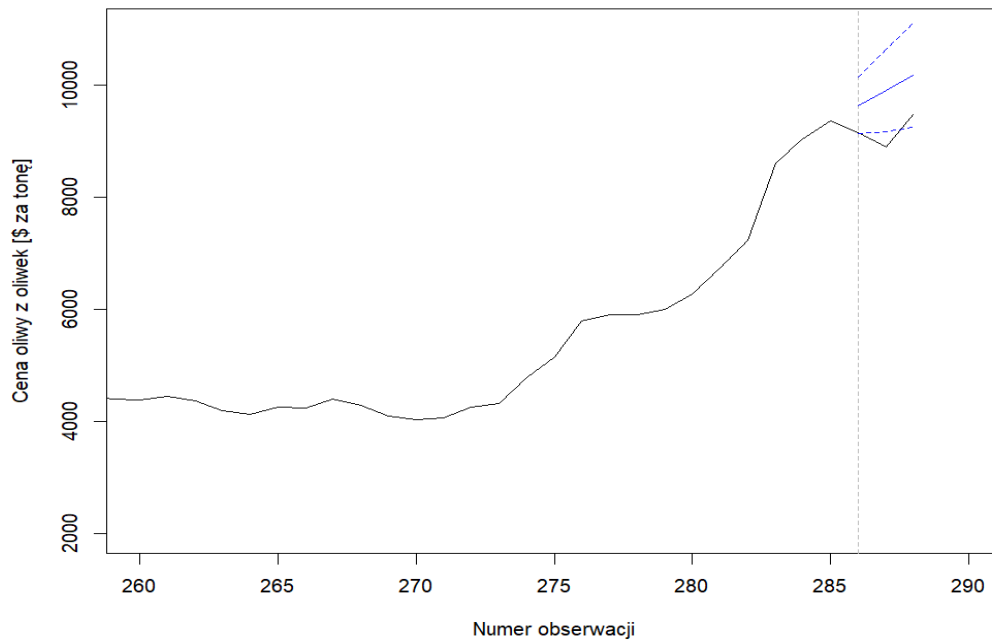
Źródło: Opracowanie własne przy pomocy RStudio.

W przypadku modelu liniowego Holta konieczne jest określenie parametrów α oraz β , które dobierane są w taki sposób, aby zminimalizować błąd średniokwadratowy. Na Rys. 8 przedstawiony został wygładzony przy użyciu liniowej metody Holta analizowany szereg z wygenerowanymi automatycznie wartościami $\alpha=1$ i $\beta=0.06901094$. Natomiast na Rys. 9 pokazana została prognoza analizowanego szeregu na 3 najbardziej aktualne okresy przy użyciu wspomnianej metody.

Rysunek 10 z kolei przedstawia prognozę przy użyciu prostego wygładzania wykładniczego (SES). Już na podstawie wykresu (Rys. 10) widać, że metoda ta jest lepiej

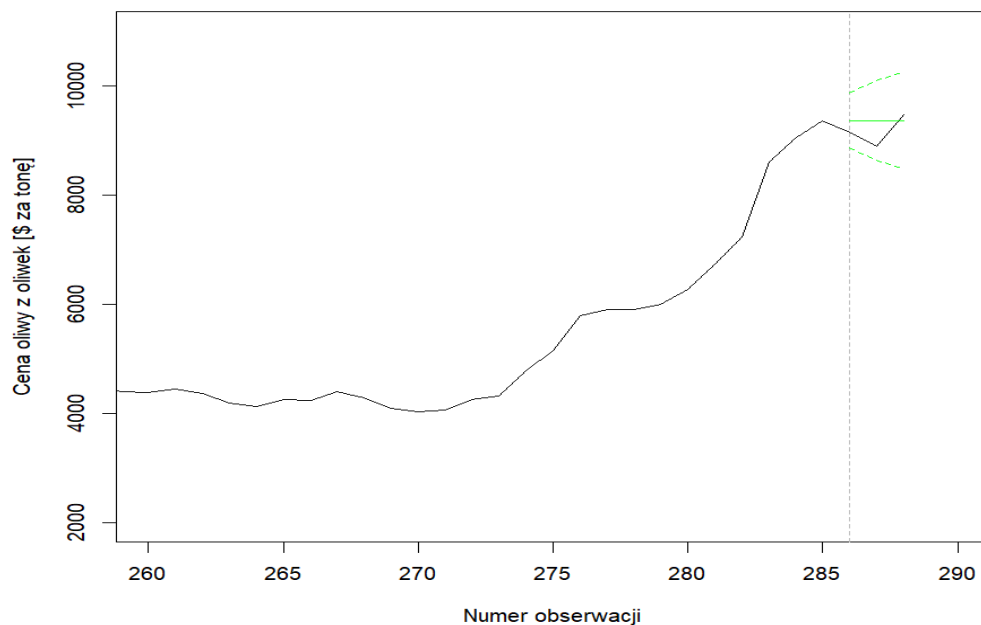
dopasowana w porównaniu do metody Holta, ponieważ wszystkie trzy faktyczne wartości znajdują się w przedziale ufności w odróżnieniu od prognozy przy użyciu metody Holta na Rys. 9.

Rys. 9. Prognoza szeregu przy użyciu metody liniowej Holta



Źródło: Opracowanie własne przy pomocy RStudio.

Rys. 10. Prognoza szeregu przy użyciu metody prostego wykładniczego (SES)



Źródło: Opracowanie własne przy pomocy RStudio.

Aby jednak formalnie porównać sprawdzone metody ekstrapolacyjne oraz model ARIMA(0,1,2) obliczono błędy prognozy ex post dla każdego modelu, które są przedstawione w Tabeli 7.

Tabela 7. Błędy prognozy ex post szeregu niesezonowego

	ARIMA(0,1,2)	Proste wygładzanie wykładnicze (SES)	Model liniowy Holta
MAE	318.329	265.7641	733.5915
MSE	155818.2	94486.56	584156.4
MAPE	3.539255 %	2.942445%	8.039456%
AMAPE	3.445072 %	2.88887%	7.701115%

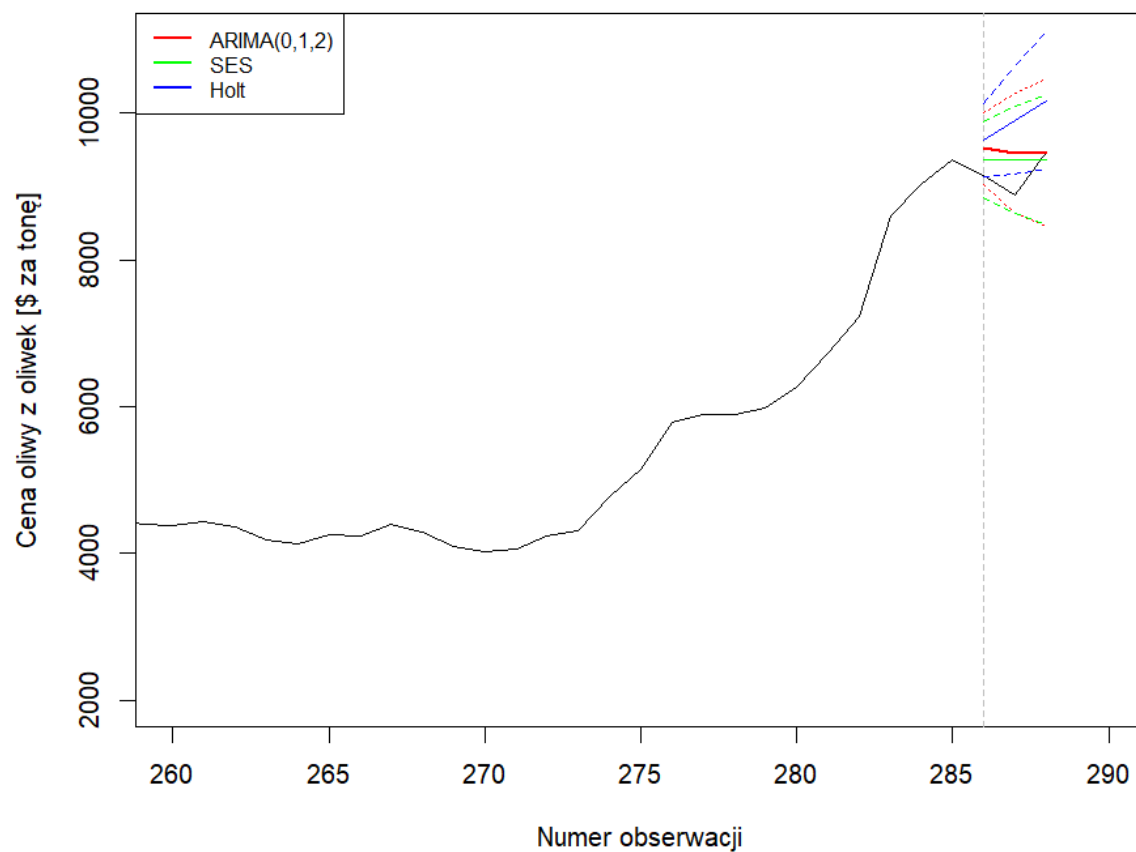
Źródło: Opracowanie własne.

Niezależnie od rodzaju błędu ex post, uzyskane wartości wskazują na to, że metoda prostego wygładzania wykładniczego (SES) jest najlepiej dopasowana do prognozowania szeregu niesezonowego. Następnym najlepszym modelem po niej jest ARIMA(0,1,2). Najgorszą z przetestowanych metod z widocznie wyższymi wartościami błędów ex post okazał się być model liniowy Holta.

1.5. Porównanie otrzymanych modeli

Jako podsumowanie przedstawionej analizy szeregu niesezonowego porównane zostały nie tylko wartości błędów ex post, ale również otrzymane prognozy modelu ARIMA, prostego wygładzania wykładniczego (SES) oraz modelu liniowego Holta w postaci wykresu. Ponieważ model ARIMA(0,1,2), bazując na wartościach błędów prognozy ex post, był dokładniejszy w swoich prognozach wśród przetestowanych modeli ARIMA, to właśnie on został porównany do wyników otrzymanych przy wykorzystaniu modeli ekstrapolacyjnych. Prognozy porównywanych modeli przedstawione zostały na Rys. 11. Jak można zauważyć na wykresie oraz wcześniejszych tabelach przedstawiających błędy prognoz ex post dla porównywanych modeli, model prostego wygładzania wykładniczego (SES) okazał się najlepszy w prognozowaniu analizowanego szeregu, gdyż znajduje się najbliżej faktycznych obserwacji.

Rys.11. Prognozy modeli ARIMA, liniowego Holta, proste wygładzanie wykładnicze



Źródło: Opracowanie własne przy pomocy RStudio.

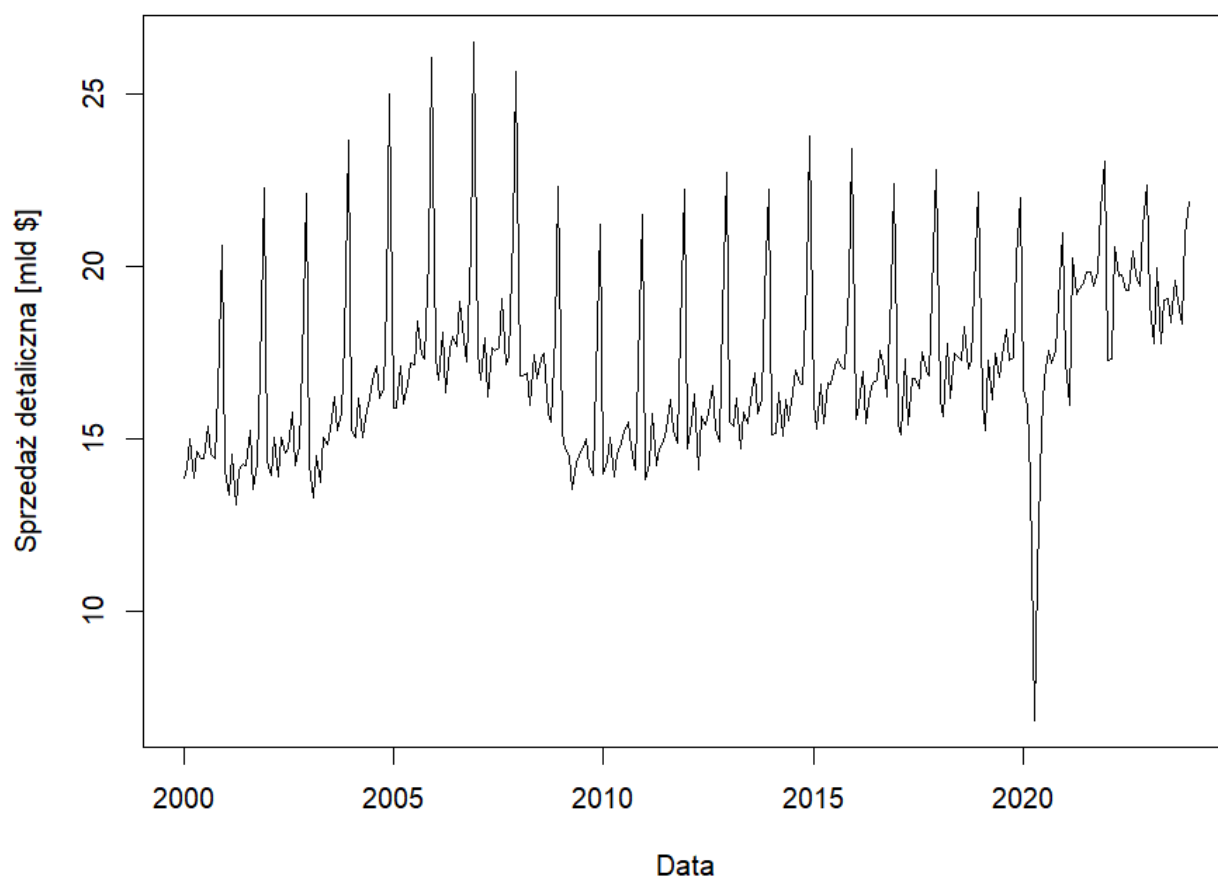
ROZDZIAŁ II

Szereg sezonowy

2.1. Opis szeregu

Analizowany szereg sezonowy przedstawia wartość sprzedaży detalicznej w kategoriach takich jak sklepy z meblami, wyposażeniem domu, elektroniką i sprzętem AGD w miliardach dolarów amerykańskich. Dane pochodzą z bazy danych Federal Reserve Economic Data (FRED). Szereg obejmuje łącznie 288 obserwacji. Pierwsza obserwacja występuje w styczniu 2000 roku, natomiast ostatnia przypada na grudzień 2023 roku. Na Rys. 12 przedstawiony został wykres obserwacji w czasie.

Rys. 12. Wykres szeregu sezonowego

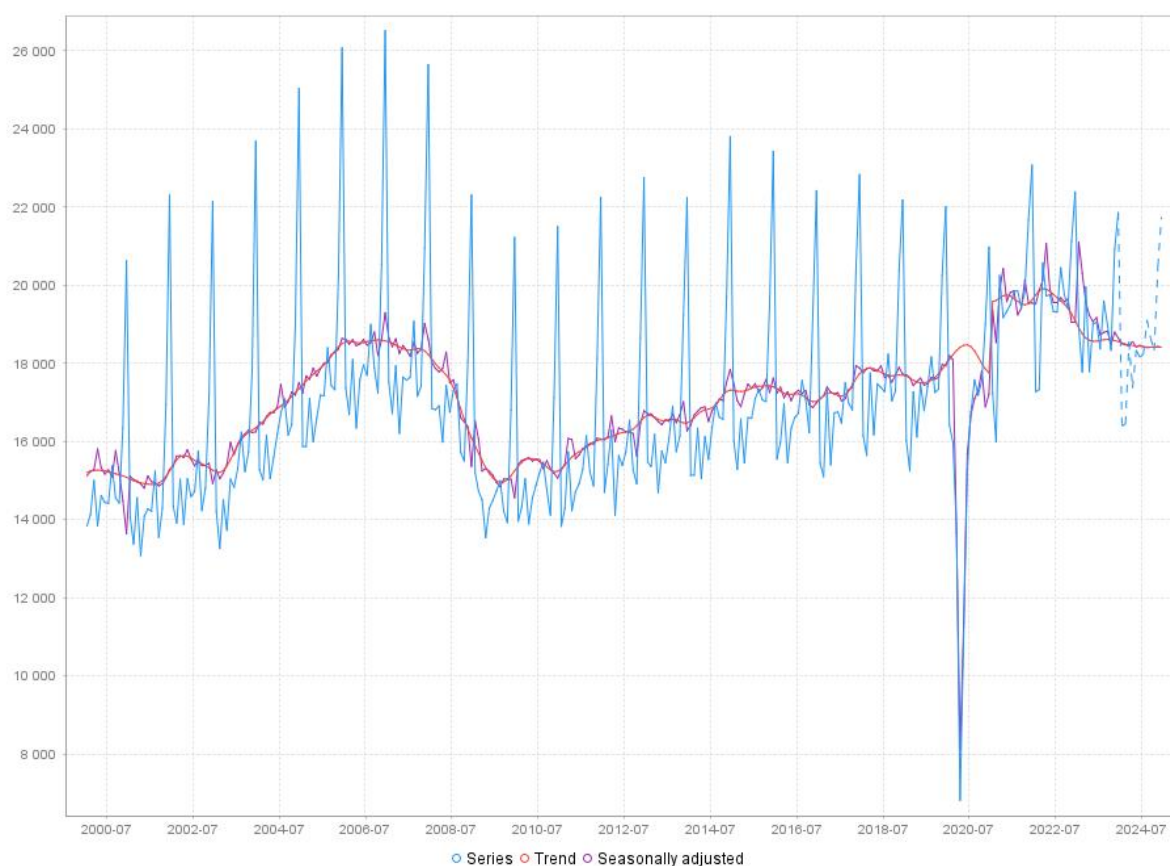


Źródło: Opracowanie własne przy pomocy RStudio.

2.2. Dekompozycja szeregu

Na podstawie Rys. 12 można stwierdzić sezonowość szeregu, co potwierdza program JDemetra+. Natomiast na bazie obserwacji sezonowość można uznać za addytywną, ze względu na brak jasno widocznej zmiany amplitudy wahań sezonowych. Dodatkowo, program JDemetra+ wskazał pięć obserwacji odstających. Obserwacje o charakterze długotrwałym (level shift) pochodzą z marca i kwietnia 2020 roku oraz stycznia i czerwca 2021 roku. Z kolei obserwacja odstająca z września 2020 roku ma charakter przejściowy (temporary change). Na Rys. 13 przedstawiona została dekompozycja analizowanego szeregu sezonowego na jego poszczególne składowe. Zauważyć można, że trend w szeregu nie jest jednoznaczny, był wzrost w latach 2000-2006 i 2009-2018 z widocznym spadkiem i następującym po nim wzroście wynikającym z pandemii COVID-19 w 2020 roku. Z kolei w latach 2006-2009 trend był malejący, podobnie jak w latach 2022-2023.

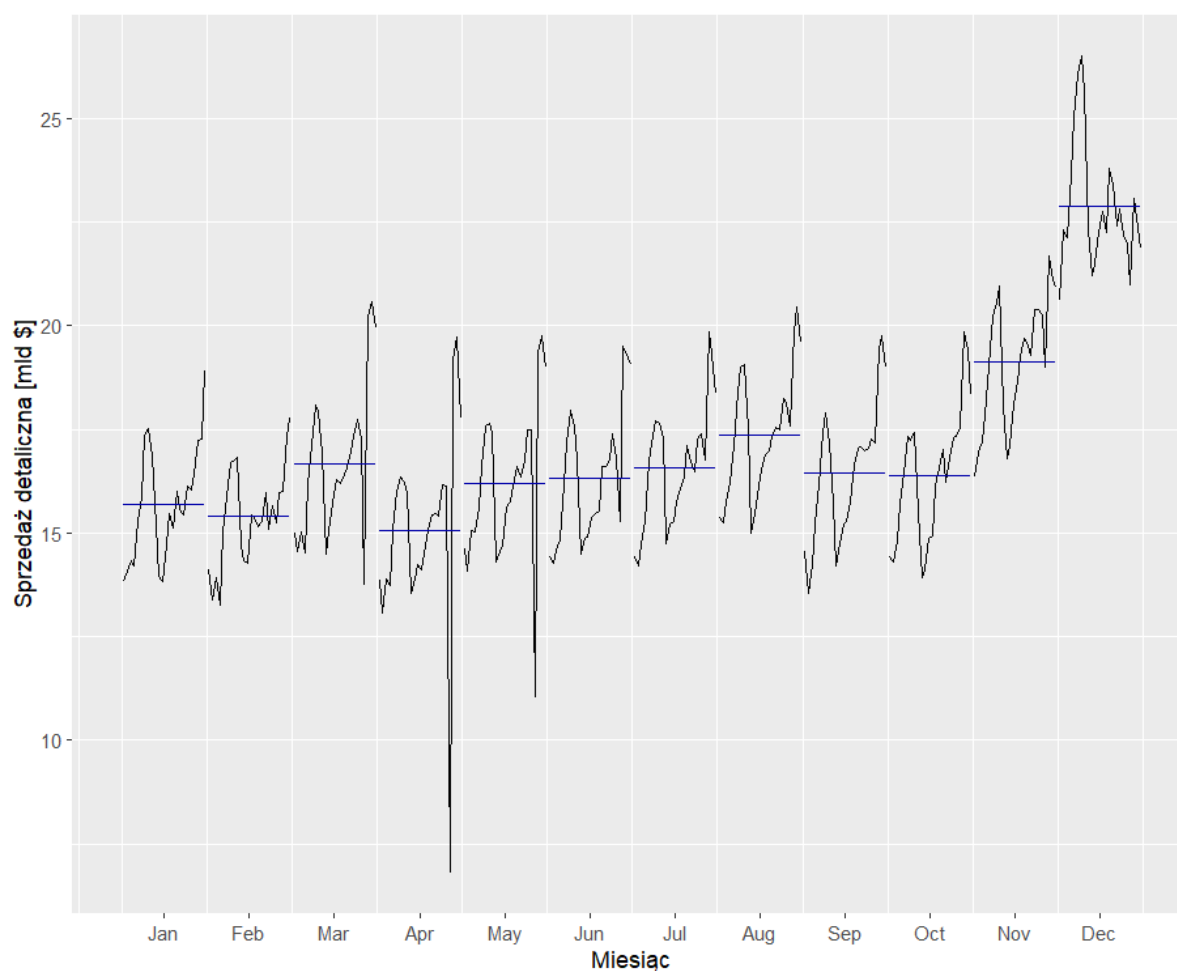
Rys. 13. Dekompozycja szeregu sezonowego (w mln \$)



Źródło: Opracowanie własne przy pomocy JDemetra+.

Dodatkowo, sezonowość badanego szeregu potwierdza dekompozycja przedstawiona na Rys. 14. Widzimy, że najwyższa sprzedaż jest osiągana w grudniu, a także jest znacząco wyższa w porównaniu do innych miesięcy w listopadzie.

Rys. 14. Dekompozycja szeregu sezonowego



Źródło: Opracowanie własne przy pomocy RStudio.

2.3. Model SARIMA

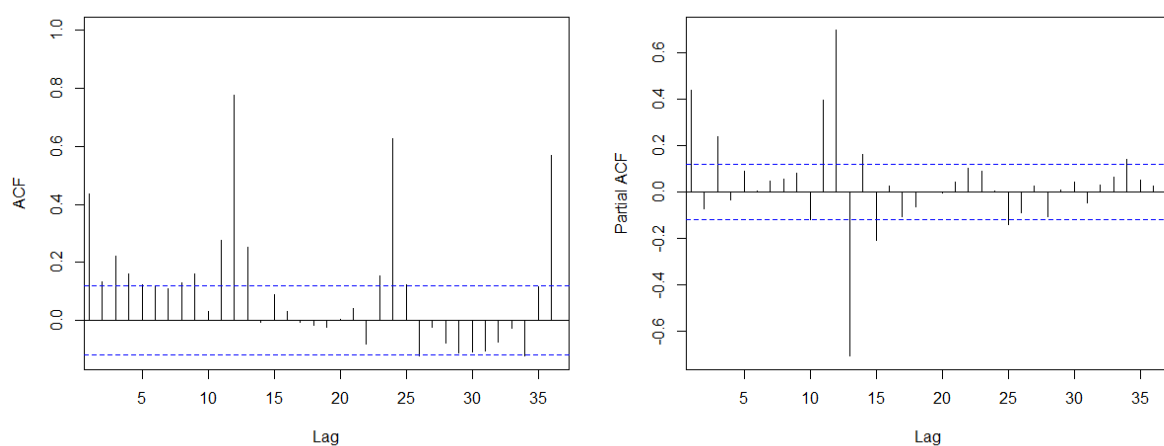
Do budowy modelu pozwalającego na prognozowanie szeregu sezonowego wykorzystano model SARIMA, czyli sezonową wersję modelu ARIMA. Ponownie do jego budowy wykorzystana została procedura Boxa-Jenkinsa. Model ten można podzielić na następujące procesy: $AR(p)$, $MA(q)$, $I(d)$ oraz sezonowy autoregresyjny $SAR(P)$, sezonowy średniej ruchomej $SMA(Q)$ i sezonowy stopnia integracji $SI(D)$. W celu budowy modelu i przeprowadzenia prognozy obserwacje zostały podzielone na okresy in-sample oraz out-of-

sample. Ponieważ analizowany szereg składa się z obserwacji miesięcznych, okres out-of-sample wyniesie 12 obserwacji, które będą prognozowane. Ponownie jak w przypadku szeregu niesezonowego, w pierwszej kolejności należy zweryfikować stacjonarność szeregu.

2.3.1. Stacjonarność szeregu

Do przeprowadzenia prognozy szeregu przy wykorzystaniu modelu SARIMA niezbędne jest, aby wykazywał on stacjonarność oraz brak sezonowości. W pierwszym kroku stacjonarność oraz sezonowość sprawdzone będą na bazie weryfikacji graficznej. W kolejnym kroku stacjonarność zostanie zweryfikowana na podstawie testów Dickey-Hasha-Fullera oraz KPSS. Na Rys. 15 przedstawione zostały wykresy ACF i PACF dla analizowanego szeregu sezonowego.

Rys. 15. Wykres ACF i PACF szeregu sezonowego

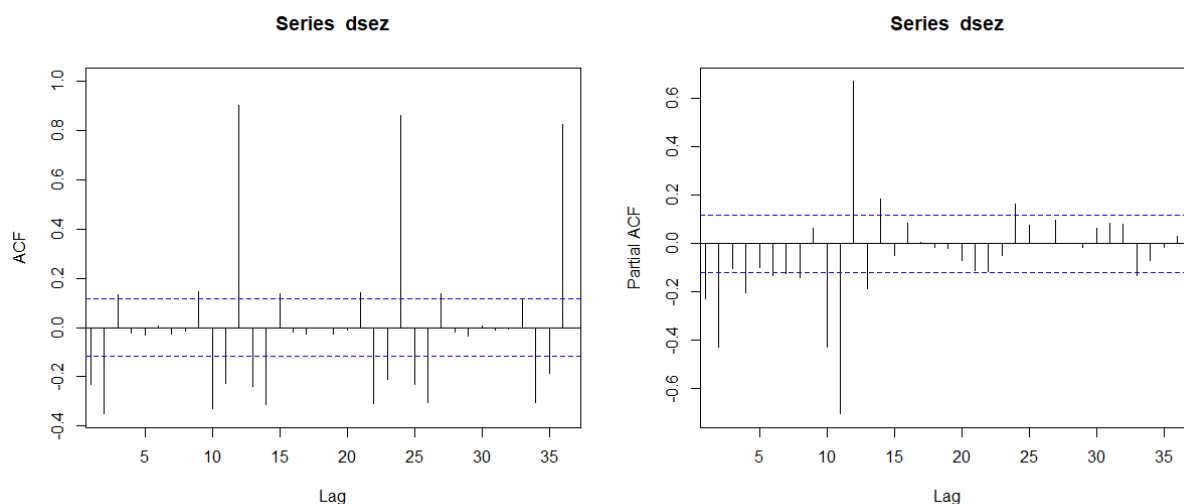


Źródło: Opracowanie własne przy pomocy RStudio.

Jak można wywnioskować z Rys. 15, szereg prawdopodobnie cechuje się niestacjonarnością oraz na wykresie ACF widać silną sezonowość poprzez istotne wypustki w pozycjach 12, 24 oraz 36. Można zatem przypuszczać, że szereg jest niestacjonarny regularnie i sezonowo. Zastosowane zostało zatem pierwsze różnice szeregu sezonowego ($d=1$). Korelogramy dla szeregu zróżnicowanego w stopniu pierwszym widoczne są na Rys. 16. Jak można zauważyć, nadal występuje silna sezonowość (wypustki statystycznie istotne dla opóźnień równych 12, 24, 36). Wymagane jest zatem wykorzystanie różnicowania sezonowego. Na Rys. 19 przedstawione zostały wykresy funkcji ACF oraz PACF dla szeregu

zróznicowanego regularnie w stopniu pierwszym oraz sezonowo w stopniu pierwszym ($d = 1$ oraz $D = 1$).

Rys. 16. Korelogram zróznicowanego szeregu (ACF, PACF)



Źródło: Opracowanie własne przy pomocy RStudio.

W celu sprawdzenia zasadności wykorzystania różnicowania sezonowego posłużono się testem Dickey-Hasha-Fullera. W przypadku zerowego opóźnienia reszty wykazywały silną autokorelację. Zastosowana została zatem rozszerzona wersja testu, czyli ADHF. W celu usunięcia autokorelacji wymagane było zastosowanie 4 opóźnień w teście. W takim przypadku autokorelacja reszt nie występowała dla badanych czterech pierwszych rzędów. Uzyskane wyniki przedstawiono w Tabeli 8. Ostateczna wartość statystyki testowej wyniosła 0,325, natomiast wartość statystyki krytycznej była na poziomie około -5.83. Nie ma zatem podstaw do odrzucenia hipotezy zerowej stwierdzającej, że występuje sezonowy pierwiastek jednostkowy. Wykorzystanie różnicowania sezonowego jest zatem zasadne zgodnie z wnioskami uzyskanymi na podstawie analizy graficznej korelogramów.

W kolejnym kroku sprawdzona została regularna stacjonarność szeregu przy wykorzystaniu testu DF ze stałą. Ze względu na występującą autokorelację reszt zastosowano rozszerzoną wersję testu, czyli ADF ze stałą. Wyniki przedstawiono na Rys. 17. Widzimy, że autokorelacja została wyeliminowana dla każdego z sześciu rzędów korelacji przy $k=3$, a wartość statystyki testowej wynosi -11,43634, co pozwala nam odrzucić hipotezę zerową dla testu ADF mówiącą o tym, że szereg jest niestacjonarny przy wartości krytycznej na poziomie -1,95. Podobny wniosek uzyskano po przeprowadzeniu testu KPSS (Rys. 18), który dla poziomu istotności 5% (wartość krytyczna równa 0.463) także stwierdzał stacjonarność

szeregu (statystyka testowa równa 0.0177). Sprawdzone zostały zatem wykresy korelogramu dla funkcji ACF i PACF, przedstawione na Rys. 19, w celu weryfikacji, czy również na nich widoczna jest stacjonarność szeregu.

Tabela 8. Wyniki testu Dickey-Hasha-Fullera dla szeregu sezonowego

Liczba opóźnień k	Test DF/ADF		p-value dla danego rzędu korelacji w teście B-G			
	Statystyka testowa	p-value	1	2	3	4
0	1.462	0.1448	2.8e-37	1.87e-37	2.18e-37	1.53e-37
1	0.16	< 2.2e-16	4.8e-5	7.57e-7	2.16e-6	1.06e-6
2	0.473	< 2.2e-16	5.3e-4	9.6e-4	2.005e-3	1.9e-3
3	0.225	< 2.2e-16	0.28612	0.03696	0.06682	0.10942
4	0.325	< 2.2e-16	0.35413	0.14144	0.14064	0.23545

Źródło: Opracowanie własne.

Rys. 17. Test ADF dla szeregu zróżnicowanego regularnie i sezonowo (d=1, D=1)

augmentations	adf	p_adf	p_bg.p.value.1	p_bg.p.value.2	p_bg.p.value.3	p_bg.p.value.4	p_bg.p.value.5	p_bg.p.value.6
0	-14.37737	0.01	0.5047497	0.0000004996515	0.000001808373	0.00000006051852	0.00000004014233	0.0000001079323
1	-15.15396	0.01	0.8839713	0.5073490128504	0.634447511504	0.02784280691008	0.01145965627327	0.0213704358532
2	-11.34459	0.01	0.9308165	0.4768916530822	0.680339038817	0.03714001781402	0.01284052402389	0.0232956856721
3	-11.43634	0.01	0.7178771	0.9272009343089	0.914323547119	0.97080530163011	0.84167610260090	0.9146869963945

Źródło: Opracowanie własne przy pomocy RStudio.

Rys. 18. Test KPSS dla szeregu zróżnicowanego regularnie i sezonowo (d=1, D=1)

```
value of test-statistic is: 0.0177
critical value for a significance level of:
      10pct  5pct  2.5pct  1pct
critical values 0.347 0.463 0.574 0.739
```

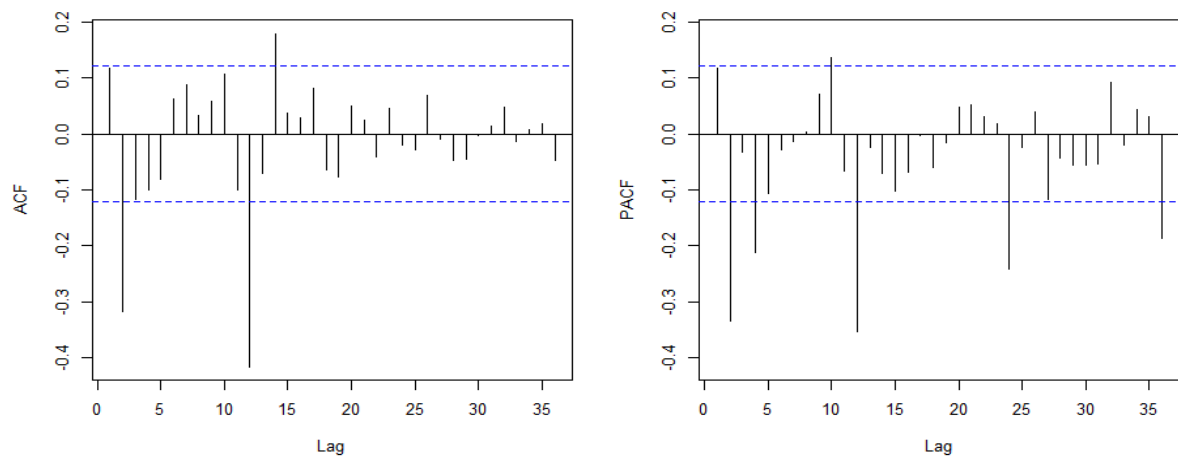
Źródło: Opracowanie własne przy pomocy RStudio.

Jak można zauważyć na Rys. 19, gdzie przedstawione zostały wykresy funkcji ACF i PACF dla szeregu zróżnicowanego regularnie w stopniu pierwszym oraz sezonowo w stopniu pierwszym (d = 1 oraz D = 1), korelogramy wskazują na występowanie sezonowości, a niestacjonarność zostaje wyeliminowana.

Aby dodatkowo upewnić się, czy zasadnie zostały wykorzystane zarówno różnice regularne i sezonowe, przeprowadzono test ADF dla szeregu zróżnicowanego jedynie regularnie (d=1) lub jedynie sezonowo (D=1). W każdym z przypadków nie udało się pozbyć

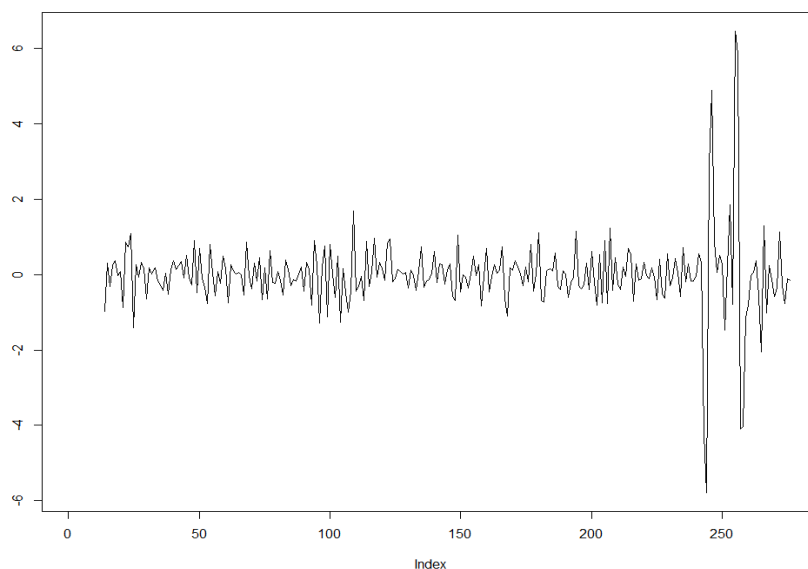
autokorelacji reszt pomimo dodania opóźnień w wersji ADF ze stałą, a wynik testu KPSS odrzuca hipotezę zerową o stacjonarności szeregu. Przyjmuje wartość 0,9732 dla szeregu zróżnicowanego regularnie i 0,6681 dla szeregu zróżnicowanego sezonowo przy wartości krytycznej 0,463.

Rys. 19. Korelogram szeregu zróżnicowanego regularnie i sezonowo (ACF, PACF)



Źródło: Opracowanie własne przy pomocy RStudio.

Rys. 20. Wykres zmiennej zróżnicowanej regularnie i sezonowo



Źródło: Opracowanie własne przy pomocy RStudio.

Aktualny szereg został pozbawiony sezonowości i zróżnicowany. Test ADF wykazał brak niestacjonarności, co potwierdził też test KPSS. Na Rys. 20 przedstawiony został wykres szeregu zróżnicowanego regularnie i sezonowo. Jak można zauważyć na wspomnianym

wykresie jasno widoczny jest wpływ obserwacji odstających wynikających z pandemii COVID-19.

Tabela 9. Test Ljunga-Boxa i Boxa-Pierce'a dla szeregu zróżnicowanego regularnie i sezonowo

Test	Statystyka testowa	p-value	Wniosek
Ljunga-Boxa	119.2	7.703e-11	Proces nie jest białym szumem.
Boxa-Pierce'a	113.87	5.166e-10	Proces nie jest białym szumem.

Źródło: Opracowanie własne.

Wyniki testów Ljunga-Boxa oraz Boxa-Pierce'a przedstawione w Tabeli 9 pozwalają odrzucić hipotezę zerową o tym, że proces jest białym szumem. Szereg sezonowy jest więc odpowiednio przygotowany i można przejść do wykonania prognozy.

2.3.2. Identyfikacja, estymacja oraz diagnostyka

Kolejnym krokiem w budowie modelu SARIMA jest identyfikacja jego parametrów (p,d,q) oraz $(P,D,Q)_s$. Parametr d przyjmuje wartość 1 ze względu na to, że wykonane zostało pierwsze różnicowanie szeregu. Także parametr D przyjmuje wartość 1 ze względu na zróżnicowanie sezonowe szeregu. Widoczna zmienna s przyjmuje wartość 12, ponieważ analizowane dane bazują na miesięcznych obserwacjach. Z kolei do identyfikacji parametrów p i q oraz P i Q należy wykorzystać wykresy funkcji ACF i PACF dla szeregu zróżnicowanego regularnie i sezonowo, widoczne na Rys. 19.

Wnioskując z wykresu funkcji ACF dla zróżnicowanego regularnie szeregu można założyć, że parametr q przyjmuje wartość od 0 do 3. W przypadku parametru Q , jak można zauważyć na wykresie ACF dla szeregu zróżnicowanego sezonowo i regularnie (Rys. 19), istotna jest jedynie wypustka na 12 pozycji, zatem przyjmie on wartość 1. Analizując wykresy funkcji ACF i PACF można zauważyć, że prawdopodobnie występuje jedynie regularny proces średnich ruchomych, który wpływa również na wykres PACF, gdzie pomimo sezonowości można zauważyć charakterystyczny kształt wypustek. Z dużym prawdopodobieństwem mamy więc do czynienia jedynie z procesem średnich ruchomych, stąd $p=0$, ale zostaną przetestowane opcje dla $p=1$ i $p=2$. Natomiast dla parametru P zauważyć można, że każda z

wypustek na wykresie PACF na pozycji 12, 24 oraz 36 jest istotna, zatem przyjmie on wartość 3.

Ostatecznie w celu znalezienia optymalnego modelu przeanalizowane zostaną modele z różnymi wartościami parametrów. Dla każdego z nich zastosowany został następujący możliwy przedział wartości: $p = \{0, 1, 2\}$, $q = \{0, 1, 2, 3\}$, $P = \{0, 1, 2, 3\}$, $Q = \{0, 1\}$. W pierwszej kolejności sprawdzone zostały wszystkie kombinacje dla parametrów sezonowych. Po sprawdzeniu kombinacji dla (P,D,Q) zaczynając od $(3,1,1)$ do $(0,1,0)$ przy pomocy procedury od ogółu do szczegółu, najprostszym modelem, który udało się osiągnąć przyjmuje wartości $P = 0$, $D = 1$, $Q = 1$. W innych przypadkach stałe przy najwyższej wartości parametrów były nieistotne zatem nie zdecydowano się na wykorzystanie do analizy modeli z innymi wartościami oszacowań parametrów sezonowych. Dla takiego modelu wypustki dla wartości 12, 24 oraz 36 w wykresach dla reszt modelu funkcji ACF i PACF okazały się być nieistotne, zatem sezonowość została wyeliminowana. Również kryterium informacyjne AIC oraz BIC wskazały model ze wspomnianymi parametrami sezonowymi jako najlepszy (Rys. 21). Dla kombinacji $(P,D,Q)=(0,1,1)$ osiągnięto najniższe wartości dla AIC (668,6556) oraz BIC (675,7999) dla wszystkich przetestowanych kombinacji.

Rys. 21. Kryteria informacyjne dla różnych kombinacji parametrów sezonowych

```
> AIC(arima010311, arima010211, arima010111, arima010011, arima010010)
      df      AIC
arima010311  5 669.6649
arima010211  4 668.9493
arima010111  3 670.4664
arima010011  2 668.6556
arima010010  1 755.3533
> BIC(arima010311, arima010211, arima010111, arima010011, arima010010)
      df      BIC
arima010311  5 687.5257
arima010211  4 683.2379
arima010111  3 681.1829
arima010011  2 675.7999
arima010010  1 758.9255
```

Źródło: Opracowanie własne przy pomocy RStudio.

W kolejnym etapie szacowane były wartości parametrów regularnych modelu SARIMA. Po usunięciu modeli, dla których zmienne najwyższej wartości parametru były nieistotne, przeprowadzono procedurę od ogólnego do szczegółowego zaczynając od regularnych komponentów $(p,d,q)=(1,1,3)$. Gdy LR (Likelihood ratio), czyli test ilorazu wiarygodności, nie dawał podstaw do odrzucenia hipotezy zerowej mówiącej, że model ograniczony jest

prawdziwy, czyli jego dopasowanie do danych nie różni się istotnie od modelu bez ograniczeń, zakończono procedurę. Przebieg procedury oraz kryteriów informacyjnych AIC i BIC zostały przedstawione w Tabeli 10.

Tabela 10. Wybrane modele SARIMA (wnioski są przedstawione dla porównania modelu w danym wierszu z przedstawionym poniżej)

	AIC	BIC	Log likelihood	Model od ogółu do szczegółu	
				p-value	Wniosek
SARIMA(1,1,3)×(0,1,1) ₁₂	644.2337	665.6666	-316.12	0.382943	Brak podstaw do odrzucenia H ₀
SARIMA(0,1,3)×(0,1,1) ₁₂	642.9949	660.8557	-316.5	0.029067	Odrzucamy H ₀ , ostateczna postać modelu
SARIMA(0,1,2)×(0,1,1) ₁₂	645.7585	660.0472	-318.88	-	Nie możemy użyć tej postaci modelu

Źródło: Opracowanie własne.

Na podstawie przeprowadzonej procedury wywnioskowano, że najlepiej dopasowaną najprostszą postacią modelu jest SARIMA(0,1,3)×(0,1,1)₁₂. Potwierdza to również najniższa wartość kryterium AIC (642,9949). Kryterium BIC natomiast sugeruje, że prostszy model byłby lepszy, jednak różni się bardzo nieznacznie od BIC dla postaci modelu, która jest wybrana na podstawie procedury od ogólnego do szczegółowego oraz AIC (o ok. 0,8).

Tabela 11. Test Ljunga-Boxa i Boxa-Pierce’a dla wybranych postaci modeli SARIMA

	Test Ljunga-Boxa		Test Boxa-Pierce’a	
	Statystyka testowa	p-value	Statystyka testowa	p-value
SARIMA(1,1,3)×(0,1,1) ₁₂	44.936	0.146	41.401	0.2467
SARIMA(0,1,3)×(0,1,1)₁₂	46.373	0.1154	42.789	0.2026
SARIMA(0,1,2)×(0,1,1) ₁₂	51.617	0.0443	48.029	0.0866

Źródło: Opracowanie własne.

Kolejno w Tabeli 11 przedstawione zostały wyniki testów Ljunga-Boxa i Boxa-Pierce’a dla wybranych postaci modeli SARIMA. W przypadku najlepszego dotychczas

modelu mamy brak podstaw do odrzucenia hipotezy zerowej mówiącej o tym, że reszty są białym szumem. Warto zauważyć, że dla prostszej postaci modelu reszty nie byłyby białym szumem na podstawie testu Ljunga-Boxa, choć kryterium BIC wskazywało na to, że jest to najlepsza postać modelu.

Kierując się kryteriami informacyjnymi AIC oraz BIC, najlepszym modelem okazały się modele odpowiednio $SARIMA(0,1,3) \times (0,1,1)_{12}$ oraz $SARIMA(0,1,2) \times (0,1,1)_{12}$. Jednak procedura modelu od ogółu do szczegółu sugeruje, że prosty model nie jest odpowiedni do prognozowania pomimo minimalnie niższego kryterium BIC. Nie różni się to kryterium znacznie od bardziej ogólnego modelu, a dodatkowo przy prostszym modelu testy formalne wskazują na to, że reszty nie są białym szumem. Nie powinno się więc w tym przypadku używać prostszego modelu do prognozowania szeregu, co oznacza, że do dalszej analizy posłuży nam postać modelu $SARIMA(0,1,3) \times (0,1,1)_{12}$.

2.3.3. Prognozowanie

Mając już gotowe wartości zmiennych do modeli SARIMA możliwe jest przejście do prognozowania szeregu. Tak jak zostało wcześniej wspomniane, prognozowane będą 12 najnowszych obserwacji w szeregu. Na Rys. 22 przedstawiona została prognoza ustalonego dotychczas za najlepszy ze wszystkich postaci modelu, czyli $SARIMA(0,1,3) \times (0,1,1)_{12}$, natomiast w Tabeli 12 przedstawione zostały wartości błędów prognozy dla analizowanych modeli. Również w przypadku szeregu sezonowego zdecydowano się przeanalizować wartości wszystkich błędów prognozy ex post.

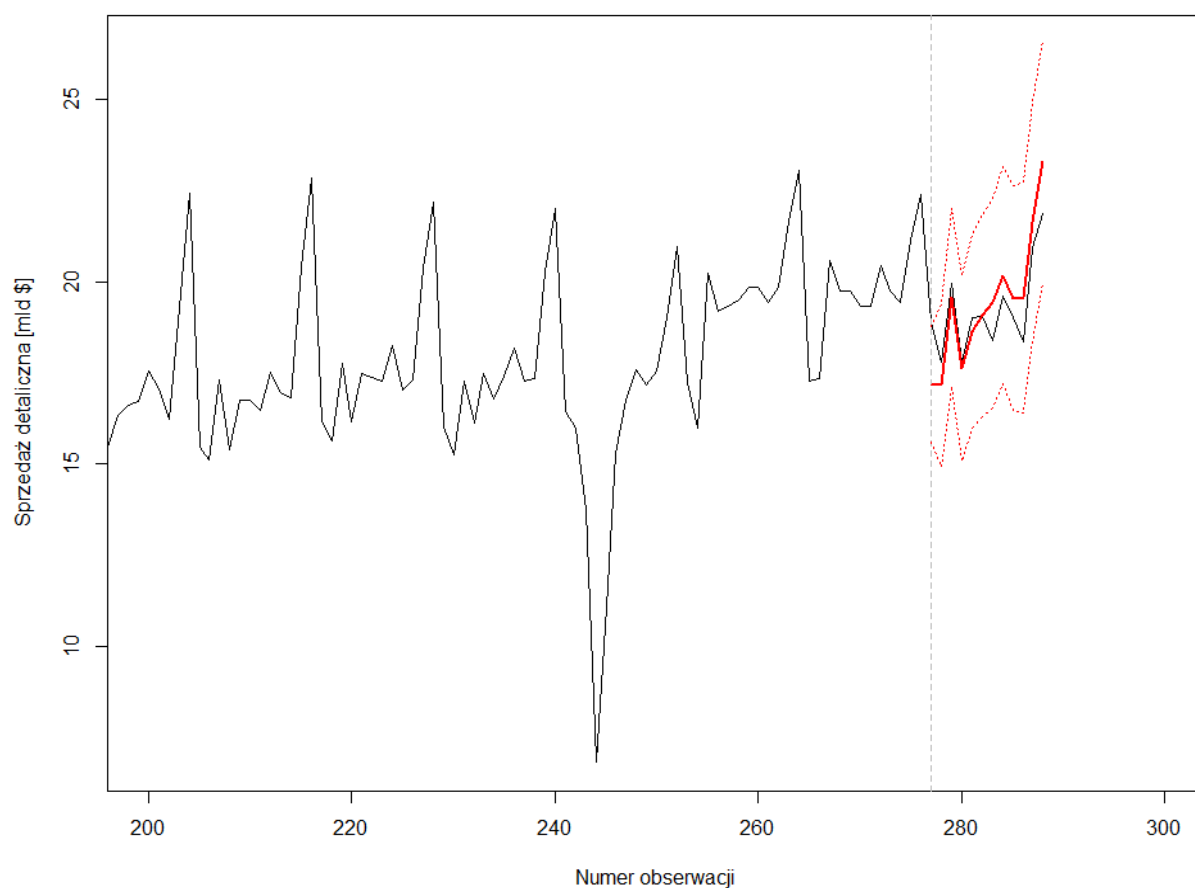
Tabela 12. Wartości błędów prognozy analizowanych modeli SARIMA

	SARIMA (0,1,3)×(0,1,1)₁₂	SARIMA (0,1,2)×(0,1,1)₁₂
MAE	0.7367557	0.7045229
MSE	0.7943053	0.6728376
MAPE	3.813208%	3.668992%
AMAPE	3.797638%	3.689144%

Źródło: Opracowanie własne.

Do błędów ex post obliczono również wartości dla prostszego modelu SARIMA, choć zgodnie z dotychczasowymi wnioskami, nie powinno się go używać do prognozowania. Okazało się, że błędy w przypadku prostszego modelu są nieco mniejsze, natomiast różnice te nie są znaczące (poniżej 0,15% w przypadku MAPE i AMAPE).

Rys. 22. Wykres prognozy modelu SARIMA(0,1,3)×(0,1,1)₁₂



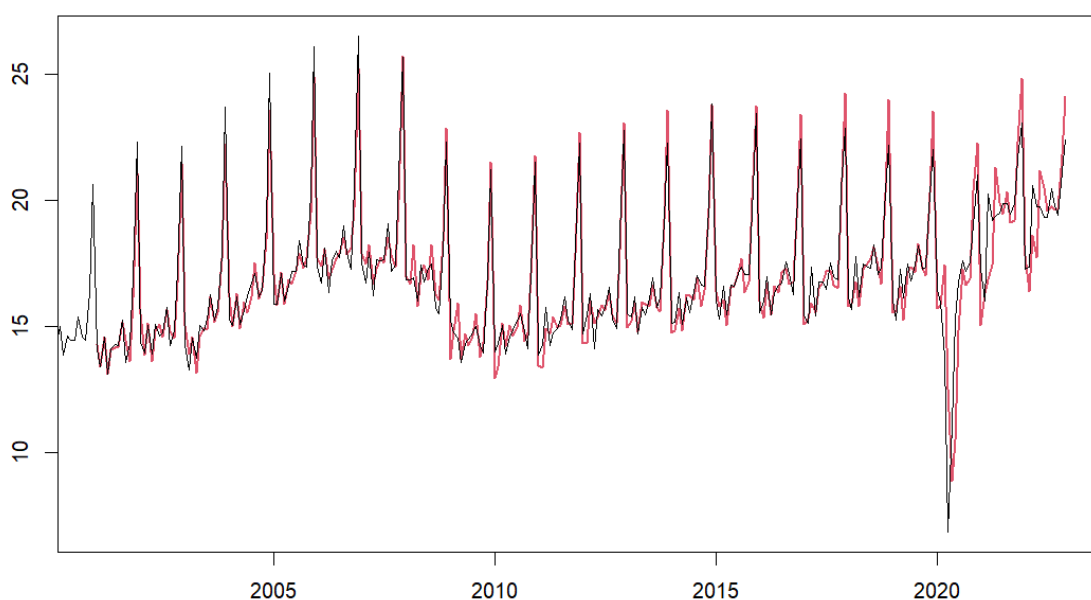
Źródło: Opracowanie własne przy pomocy RStudio.

Sprzecznie z analizą błędów prognozy ex post, ale zgodnie z wcześniejszym wnioskowaniem na podstawie kryteriów informacyjnych, procedury od ogólnego do szczegółowego i testów Ljunga-Boxa i Boxa-Pierce'a, do porównania jakości prognozy z modelem ekstrapolacyjnym wybrano bardziej ogólny model SARIMA(0,1,3)×(0,1,1)₁₂. Wynika to ze spełnienia formalnych warunków i wniosków z użytych metod. Analiza porównawcza zostanie przeprowadzona w kolejnym podrozdziale po zbudowaniu modelu ekstrapolacyjnego.

2.4. Model ekstrapolacyjny

Inną opcją przeprowadzania prognoz szeregów czasowych jest wykorzystanie modeli ekstrapolacyjnych. Ponieważ analizowana zmienna zawiera wahania sezonowe, wykorzystany został model Holta-Wintersa. Jak zostało wcześniej ustalone efekty sezonowe w analizowanym szeregu wydają się mieć charakter addytywny, ponieważ są stałe w czasie. W związku z tym wykorzystany zostanie addytywny wariant modelu Holta-Wintersa. Parametry modelu dobrane w celu minimalizacji błędu średniokwadratowego przyjęły następujące wartości: $\alpha=1$, $\beta=0$ oraz $\gamma=1.149e-13$. Na Rys. 23 przedstawiony został szereg z wygładzonymi wartościami analizowanego modelu. Natomiast na Rys. 24 przedstawiona została prognoza szeregu przy wykorzystaniu metody Holta-Wintersa. W Tabeli 13 pokazane zostały wartości błędów predykcji modelu.

Rys. 23. Szereg wygładzony przy wykorzystaniu addytywnej postaci modelu Holta-Wintersa

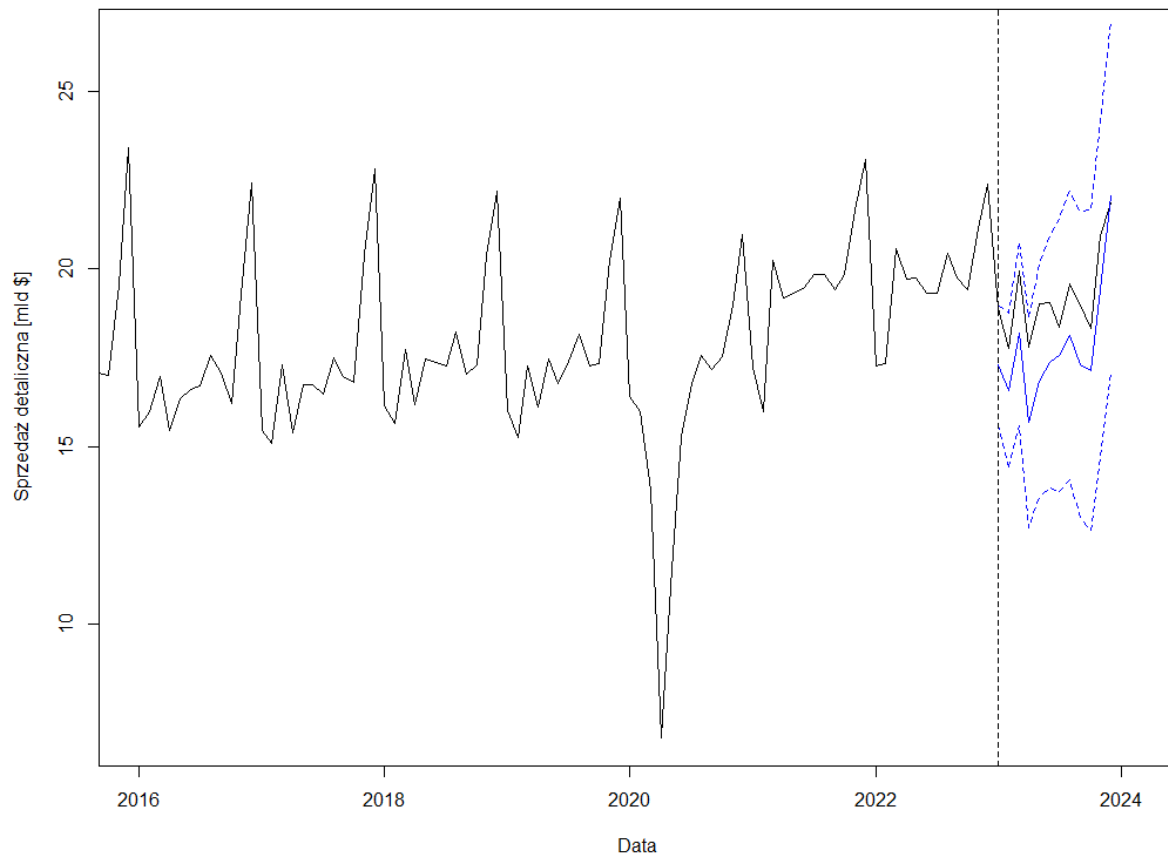


Źródło: Opracowanie własne przy pomocy RStudio.

Aby sprawdzić, czy faktycznie addytywna metoda jest odpowiednią do prognozowania, postanowiono zweryfikować, jak by wyglądała prognoza dla szeregu sezonowego przy założeniu, że prognoza jest multiplikatywna (Rys. 25). Na podstawie Rys. 25 można zauważyć, że rzeczywiste dane nie należą do przedziału ufności dla początkowych wartości, a zgodnie z

Tabela 13 błędy tej wersji prognozy ex post są nawet dwukrotnie wyższe. Słusznie więc zastosowano addytywny model Holta-Wintersa (Rys. 24).

Rys. 24. Prognoza wykonana przy użyciu addytywnej metody Holta-Wintersa



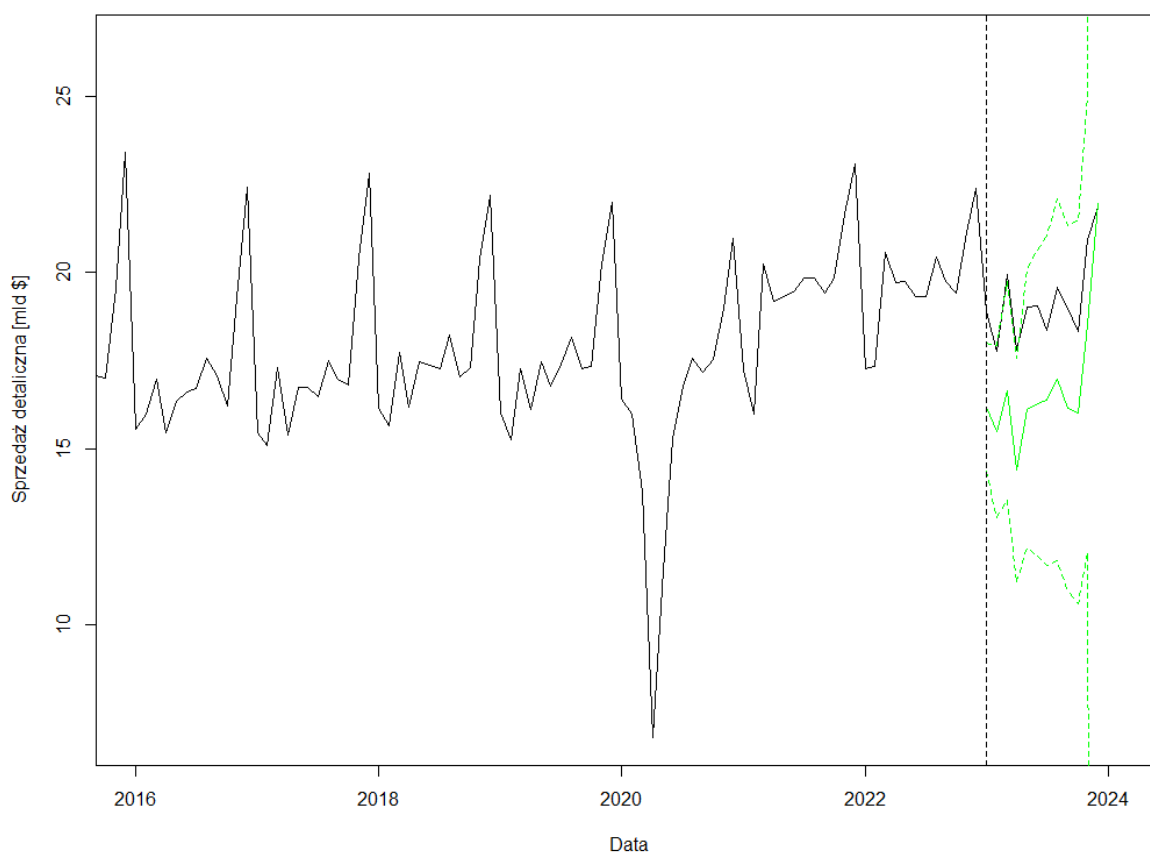
Źródło: Opracowanie własne przy pomocy RStudio.

Tabela 13. Błędy prognozy dla addytywnego i multiplikatywnego modelu Holta-Wintersa

	Addytywny model HW	Multiplikatywny model HW
MAE	1.449008	2.469427
MSE	2.372147	6.761477
MAPE	7.628869%	13.03085%
AMAPE	7.974455%	14.05065%

Źródło: Opracowanie własne.

Rys. 25. Prognoza wykonana przy użyciu multiplikatywnej metody Holta-Wintersa



Źródło: Opracowanie własne przy pomocy RStudio.

2.5. Porównanie modeli

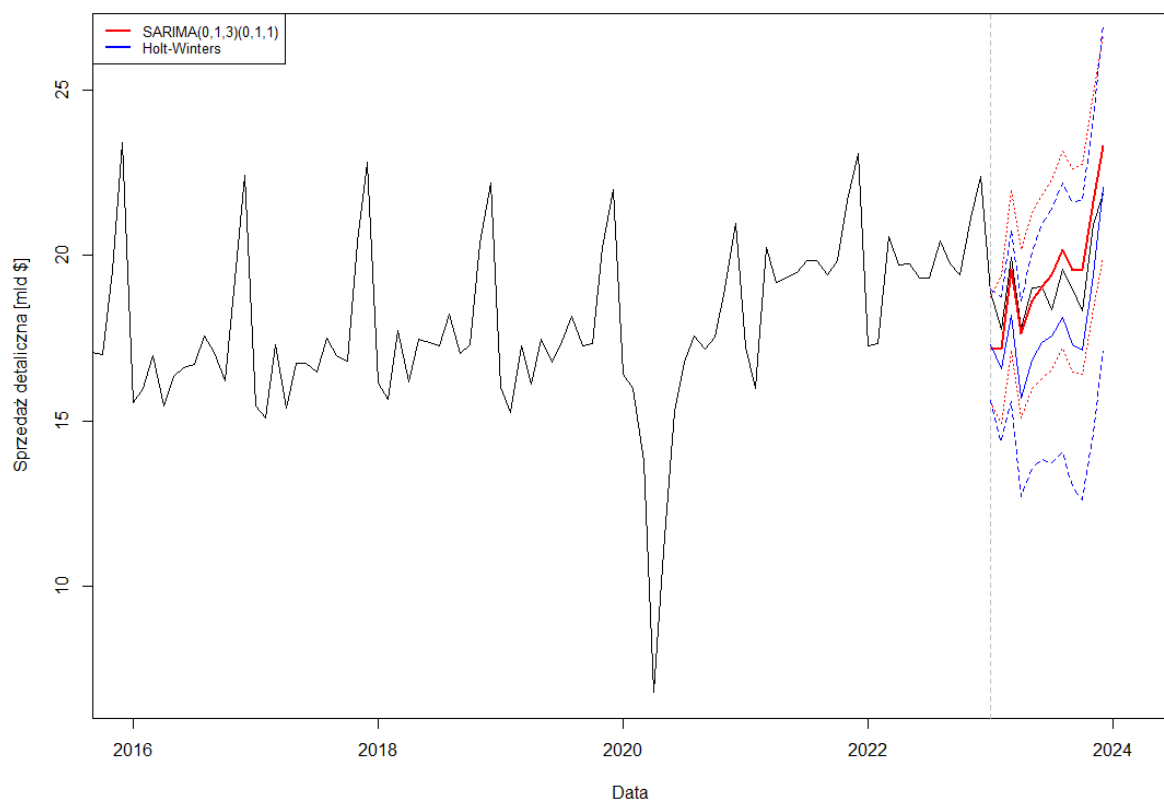
Jak można zauważyć na podstawie wartości błędów prognozy w Tabeli 14 w przypadku szeregu sezonowego model sezonowy ARIMA okazał się lepszym predyktorem zachowań analizowanego szeregu niż model ekstrapolacyjny. Model SARIMA(0,1,3)×(0,1,1)₁₂ osiągnął znacząco lepsze wyniki zarówno graficznie (Rys. 26), jak i na podstawie błędów prognozy ex post (Tabela 14) w porównaniu addytywnego modelu Holta-Wintersa. Kierując się błędami prognozy ex post, względne odchylenie prognozy dla SARIMA są prawie dwukrotnie mniejsze, co pozwala dużo trafniej prognozować wartości dla badanego szeregu. Graficzne porównanie prognoz wspomnianych modeli przedstawione zostało na Rys. 26 i prowadzi do tych samych wniosków, gdyż prognoza Holta-Wintersa wydaje się być niedoszacowana – znajduje się poniżej faktycznych wartości zmiennej.

Tabela 14. Błędy prognozy dla addytywnego modelu Holta-Wintersa i $\text{ARIMA}(0,1,3) \times (0,1,1)_{12}$

	Addytywny model HW	SARIMA $(0,1,3) \times (0,1,1)_{12}$
MAE	1.449008	0.7367557
MSE	2.372147	0.7943053
MAPE	7.628869%	3.813208%
MAPE	7.974455%	3.797638%

Źródło: Opracowanie własne.

Rys. 26. Porównanie prognoz dla szeregu sezonowego metodą ARIMA i Holta-Wintersa



Źródło: Opracowanie własne przy pomocy RStudio.

PODSUMOWANIE

Przedstawione w pracy badanie miało na celu pokazanie procesu analizy i prognozowania dwóch szeregów czasowych – nieposiadającego efektów sezonowych oraz posiadającego efekty sezonowe. Szereg niesezonowy został poddany krótkookresowej prognozie, wynoszącej 3 najnowsze obserwacje (3 miesiące). W tym przypadku najlepszy okazał się model prostego wygładzania wykładniczego (SES) w porównaniu do ARIMA(0,1,2) i modelu liniowego Holta. Z kolei szereg sezonowy poddany został prognozie wynoszącej 12 miesięcy. W jego przypadku model ekstrapolacyjny okazał się gorszym predyktorem. Model SARIMA uzyskał znacznie lepsze wyniki prognozowania w porównaniu do modelu Holta-Wintersa, zarówno w wersji addytywnej, jak i multiplikatywnej.

BIBLIOGRAFIA

U.S. Census Bureau, Retail Sales: Furniture, Home Furnishings, Electronics, and Appliance Stores [MRTAM4423USN], retrieved from FRED, Federal Reserve Bank of St. Louis; Dostęp 8 maja 2024. <https://fred.stlouisfed.org/series/MRTSSM4423XUSN>

International Monetary Fund, Global price of Olive Oil [POLVOILUSDM], retrieved from FRED, Federal Reserve Bank of St. Louis; Dostęp 8 maja 2024. <https://fred.stlouisfed.org/series/POLVOILUSDM>