

Spr 2 - ASzCz

Me

2023-04-27

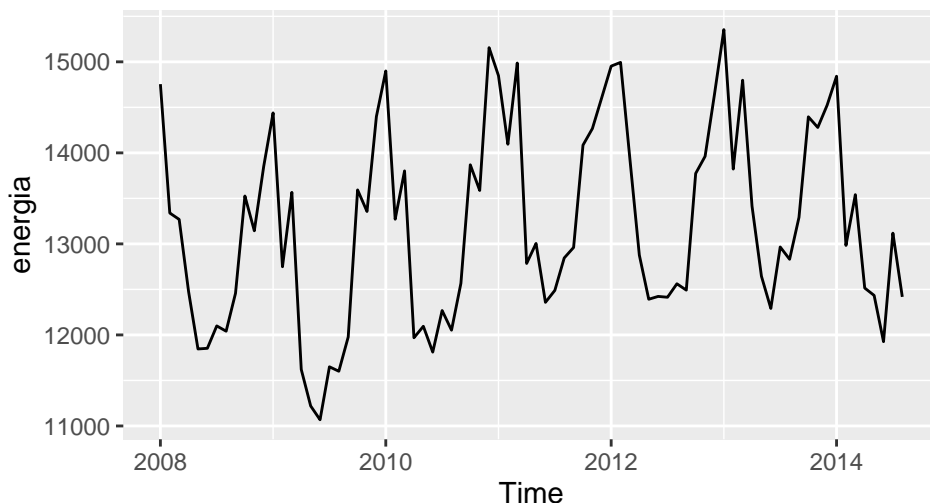
Spis treści

1	Dekompozycja szeregów czasowych – eliminacja trendu i sezonowości.	2
1.1	Wprowadzenie i analiza graficzna.	2
1.2	Dekompozycja.	3
1.3	Transformacja Boxa-Coxa.	11
1.4	Różnicowanie.	13
1.5	Podsumowanie	15
2	Zastosowanie wskaźników analizy technicznej w analizie szeregów czasowych.	15

1. Dekompozycja szeregów czasowych – eliminacja trendu i sezonowości.

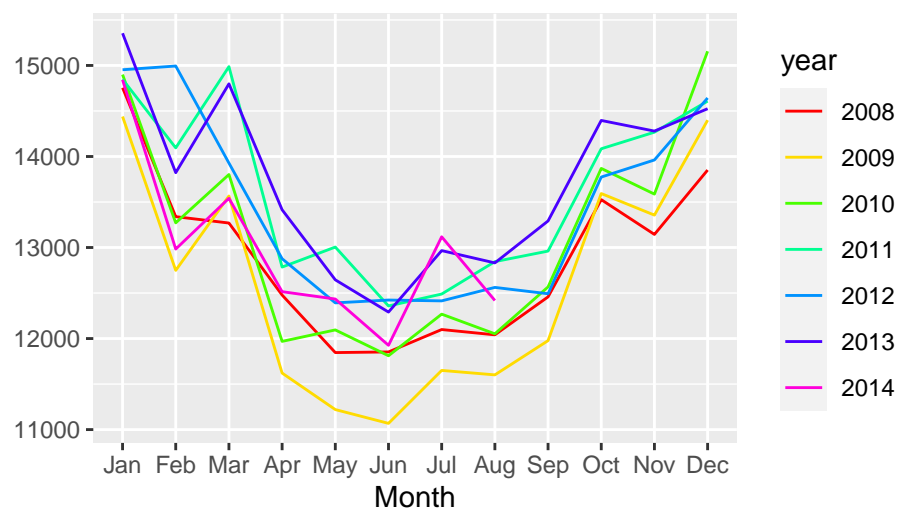
1.1. Wprowadzenie i analiza graficzna.

W pierwszej części sprawozdania zajmiemy się dekompozycją szeregu czasowego, w celu eliminacji sezonowości i trendu, oraz porównamy różne jej metody. W tym celu wykorzystaliśmy zbiór danych *energia* (w postaci zmiennej typu *ts*) z pakietu *TSAFBook*, zawierający dane o miesięcznej produkcji energii elektrycznej w Polsce. Dodatkowo, wszystkie wykonywane testy były na poziomie ufności $\alpha = 0.05$. Dane pochodzą z okresu styczeń 2008 - sierpień 2014. Na wykresie prezentują się następująco.



Wykres 1: Przedstawienie analizowanych danych na wykresie.

Na wykresie widać zachowanie sezonowe, którego okres pokrywa się z okresem podanym wraz z danymi wynoszącym 12. Przyjrzyjmy się dlatego jeszcze wykresowi sezonowemu.



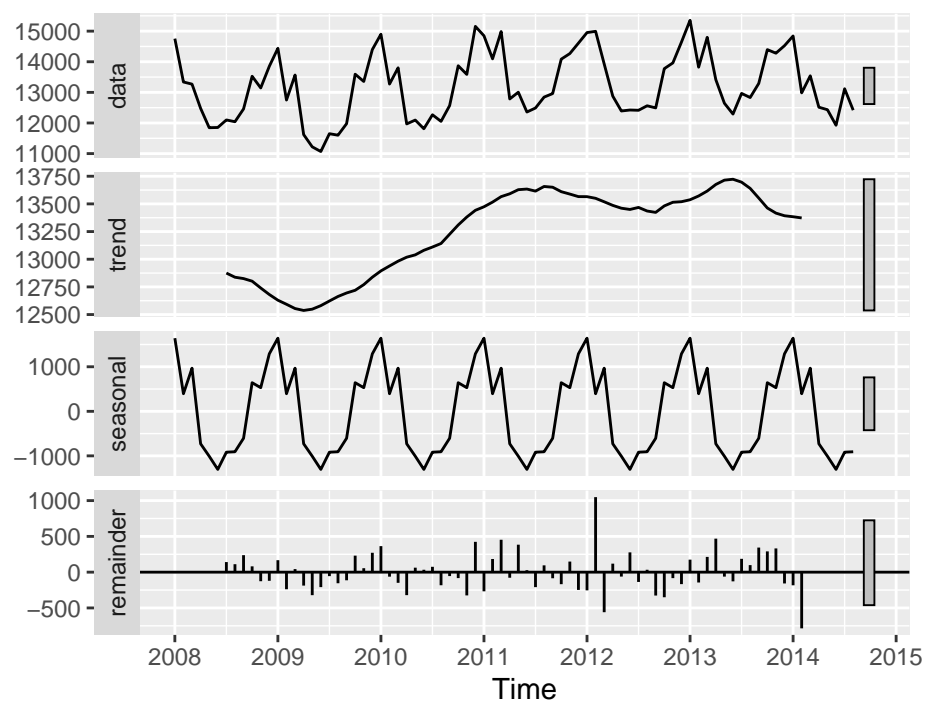
Wykres 2: Przedstawienie analizowanych danych na wykresie sezonowym.

Analizując powyższe wykresy widzimy zachowanie sezonowe, natomiast nie możemy z pewnością stwierdzić nic odnośnie istnienia trendu. Wariancja wydaje się jednorodna, jednak znajdują się dane, które odstają delikatnie od normy. Przykładowo, na wykresie sezonowym dla 2012 roku (jasnoniebieska linia), widać nietypowe zachowanie na początku roku. W lutym brakuje spadku, który zostałby skompensowany w marcu.

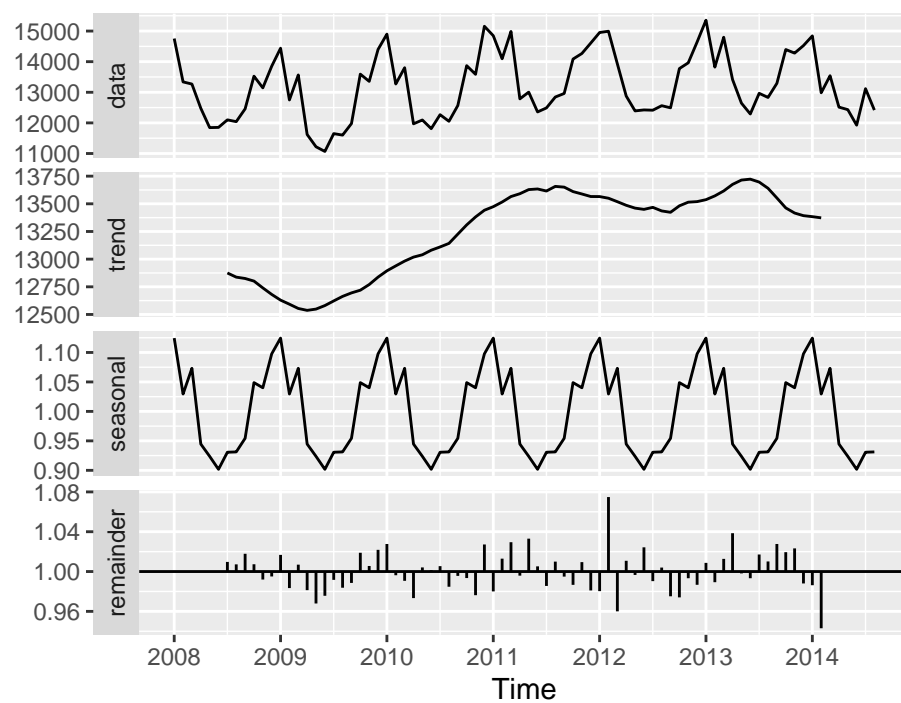
1.2. Dekompozycja.

1.2.1. Na podstawie ruchomej średniej.

Na pierwszy ogień weźmiemy dekompozycję przy użyciu funkcji `decompose` z pakietu `stats`. Dekompozycję wykonamy dwukrotnie, początkowo zakładając, że mamy model addytywny, a za drugim razem multiplikatywny.



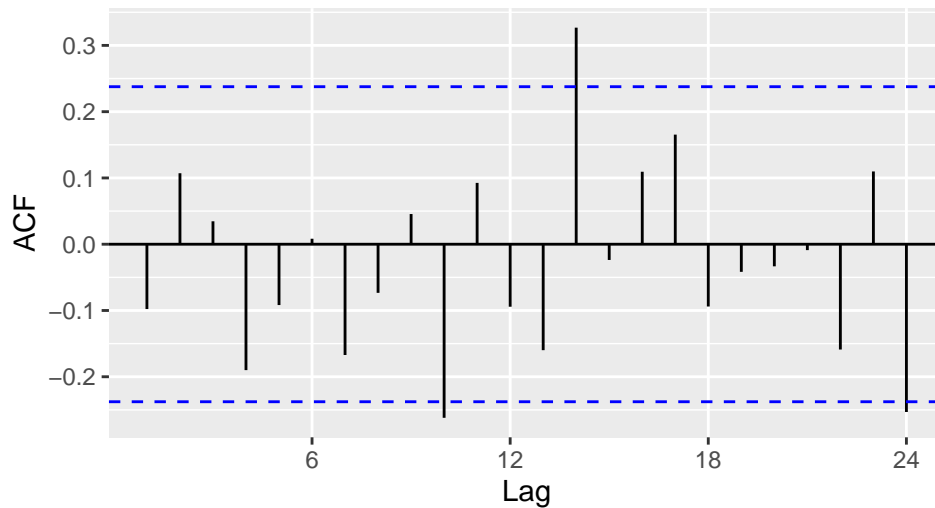
Wykres 3: Dekompozycja przy założeniu modelu addytywnego.



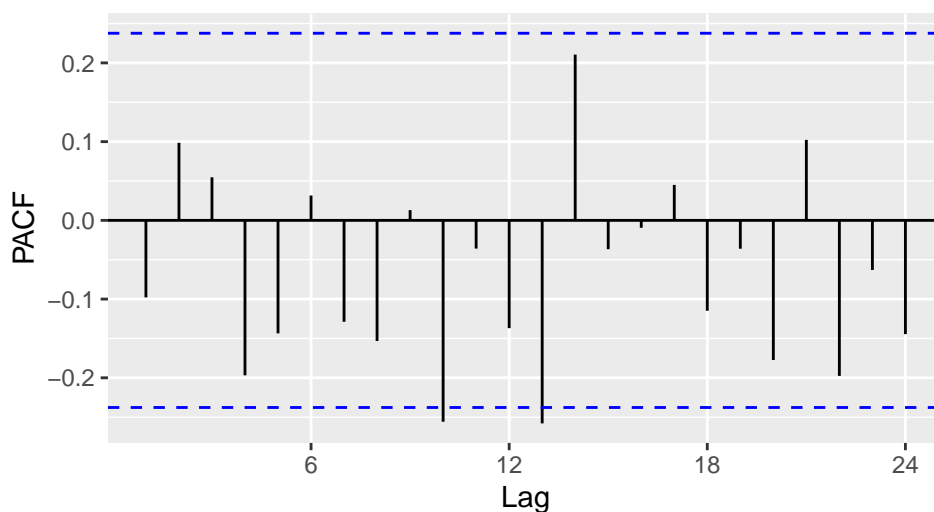
Wykres 4: Dekompozycja przy założeniu modelu multiplikatywnego.

Oba przypadki wyglądają bardzo podobnie, jeśli przyjrzymy się obserwacją odstającym w otrzymanych residuach, to zauważymy, że w przypadku addytywnym otrzymaliśmy 3 obserwacji odstających, a w przypadku multiplikatywnym 4. Za obserwacje odstające y uznaliśmy te, które spełniają $y \notin (Q_1 - 1 \cdot IQR, Q_3 + 1 \cdot IQR)$, gdzie Q_1, Q_3 są pierwszym i trzecim kwantylem, a IQR rozstępem międzykwartylowym. Jednak nie jest to dobre kryterium decyzyjne, ponieważ w przypadku addytywnym mamy mniej obserwacji odstających, jednak największa obserwacja wynosi 4.9733662 przeciętnej wysokości słupka (odległości od średniej), podczas gdy w metodzie multiplikatywnej odległość ta wynosi 4.6214001.

Sprawdźmy jeszcze czy szereg reszt jest szeregiem stacjonarnym dla modelu addytywnego. Z wcześniejszych wykresów nie mogliśmy zauważyć nic co by wskazywało na niestacjonarność reszt. Dlatego przyjrzymy się wykresowi ich funkcji autokorelacji i częściowej autokorelacji.



Wykres 5: Funkcja autokowariancji dla reszt.



Wykres 6: Funkcja częściowej autokowariancji dla reszt.

Oba wykresy szybko zanikają do wartości bliskich zeru i regularnych wzorców, zatem nie możemy odrzucić hipotezy o niestacjonarności szeregu. Dodatkowo test ADF (Augmented Dickey-Fuller test) zwrócił p-wartość na poziomie $0.0100335 < 0.05$ dla wartości opóźnienia $h=4$ (podobnie dla mniejszych wartości opóźnień). Zatem mamy podstawy do odrzucenia hipotezy zerowej, głoszącej, że istnieje pierwiastek jednostowy w wielomianie autoregresyjnym (świadczącym o niestacjonarności danych).

1.2.2. Na podstawie modelu regresji.

Drugi sposób dekompozycji będzie korzystał z funkcji `tslm` z pakietu `forecast`. Wywołanie funkcji wygląda w następujący sposób.

```
tslm(energia ~ poly(trend,4,raw=TRUE) + season)
```

W tym przypadku rozpatrujemy przypadek addytywny z: okresowością (odpowiada za to `season`) oraz trendem wielomianowym stopnia 4 (`poly(trend,4, raw=TRUE)`), opcja `raw` oznacza w jakiej postaci chcemy otrzymać współczynniki wielomianu, surowej czy ortogonalnej). Dopasowanie modelu (bez uwzględnienia residuów/losowości) wygląda w następujący sposób.

Wykres 7: Dopasowanie modelu do danych.

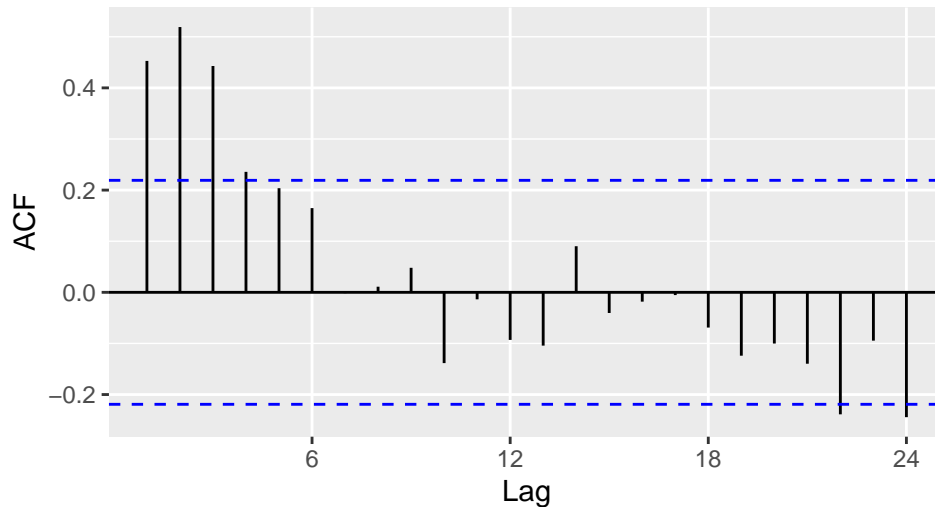
Jak można zauważyć, to dla dużych stopni wielomianu dopasowanie jest dobrze pokrywa krzywą, sprawiając tym samym, że residua są najmniejsze. Jednak przez wybranie zbyt dużego stopnia wielomianu, nie jest możliwe otrzymanie realnej predykcji.

Wykres 8: Predykcja na podstawie dopasowanego modelu.

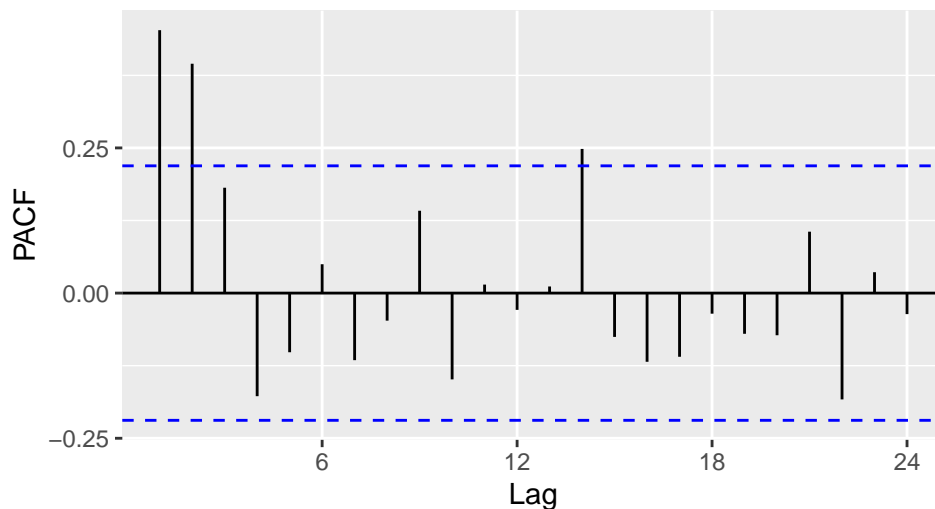
Jak możemy zauważyć na powyższych animacjach, to pomimo, że wraz z wzrostem stopnia wielomianu polepsza się jego dopasowanie do aktualnych danych, to jego predykcja staje się

nierealna. Już dla trendu kwadratowego widzimy zaskakujący spadek, który informuje, że do 2035 roku z dużym prawdopodobieństwem produkcja energii elektrycznej w Polsce spadnie do 0. Scenariusz ten jest mało realny, jedynie dla wielomianu stopnia 1 możemy rozważyć predykcję, która ma możliwość sprawdzenia się w małym okresie czasu.

Sprawdźmy jeszcze czy, przy wyborze trendu liniowego, otrzymaliśmy reszty w postaci szeregu stacjonarnego. Ponownie zacznijmy od przyjrzenia się funkcją acf i pacf.



Wykres 9: Funkcja autokowariancji dla reszt.



Wykres 10: Funkcja częściowej autokowariancji dla reszt.

Tym razem możemy zauważyć wolniejsze zanikanie powyższych funkcji oraz możliwą regularność

w pierwszej z nich. Dlatego wykorzystamy ponownie test ADF do weryfikacji stacjonarności. Tym razem test wykazał p-wartość na poziomie $0.273669 > 0.05$ (dla tego samego opóźnienia). Podobnie wygląda sytuacja dla innych opóźnień. Więc nie mamy podstawy do odrzucenia hipotezy o niestacjonarności reszt.

1.2.3. Na podstawie metody loess.

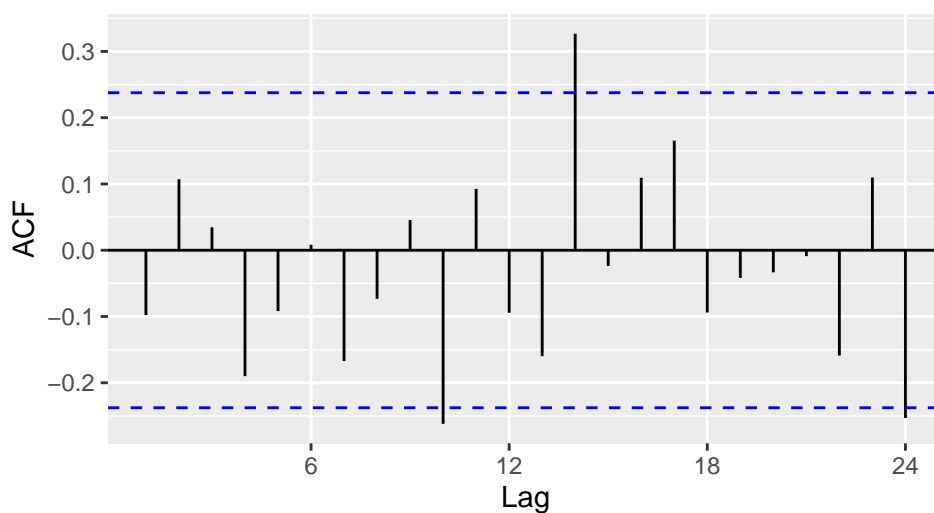
By przetestować ostatnią metodę dekompozycji posłużymy się funkcją `stl` z biblioteki `stats`. W tej metodzie będziemy zmieniać dwa parametry `s.window` oraz `t.window`, oznaczające ile obserwacji wykorzystujemy w celu dekompozycji na trend i sezonowość.

Wykres 11: Dekompozycja przy pomocy `stl` dla różnych szerokości okna dla trendu.

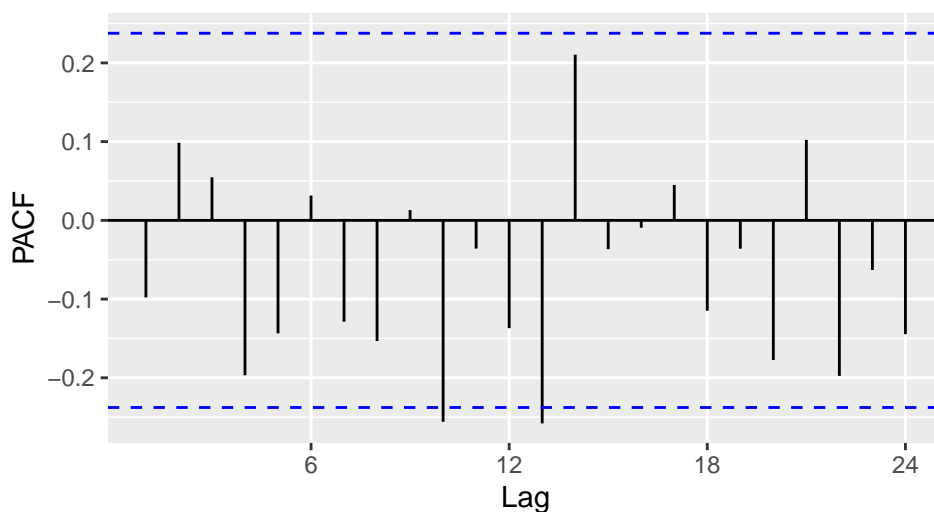
Wykres 12: Dekompozycja przy pomocy stl dla różnych szerokości okna dla sezonowości.

Jak można zauważyć na animacjach, dla małych wielkości okien otrzymujemy nie zadawalające wyniki, niezależnie czemu odpowiadała wielkość tego okna. Jednak dla dużych wartości problem ten powstaje jedynie dla trendu. W przypadku sezonowości to od wielkości okna ok. 11, zwiększanie wartości `s.window` nie polepszało, ani nie pogarszało, jakości dekompozycji.

Możemy sprawdzić, czy szereg reszt jest szeregiem stacjonarnym. Wykresy acf i pacf dla wielkości okien równych 15, prezentują się następująco.



Wykres 13: Funkcja autokowariancji dla reszt.



Wykres 14: Funkcja częściowej autokowariancji dla reszt.

W tym przypadku nie widzimy rzadnej nieregularności oraz obie funkcje szybko zanikają do zera. Sprawdźmy w jak wyglądają wyniki testu ADF dla poszczególnych wartości.

Tablica 1: Tablica p-wartości dla test ADF. Kolumny informują o wielkości rozpatrywanego okna dla trendu, podczas gdy wiersze odpowiadają za okno sezonowości.

	3	5	9	13	15	21	43
1	0.685	0.01	0.01	0.01	0.01	0.01	0.121
3	0.685	0.01	0.01	0.01	0.01	0.01	0.121
5	0.685	0.01	0.01	0.01	0.01	0.01	0.066
7	0.685	0.01	0.01	0.01	0.01	0.01	0.030
15	0.685	0.01	0.01	0.01	0.01	0.01	0.018
21	0.685	0.01	0.01	0.01	0.01	0.01	0.018

Jak możemy zauważyć, dzięki powyższej tabeli, dla większości rozmiarów okna dla średniej (przedział 5-21) mamy przesłanki do odrzucenia hipotezy o niestacjonarności. Co ciekawe, dla wartości okna $t = 43$, decydujące słowo ma również wielkość okna dla sezonowości.

1.3. Transformacja Boxa-Coxa.

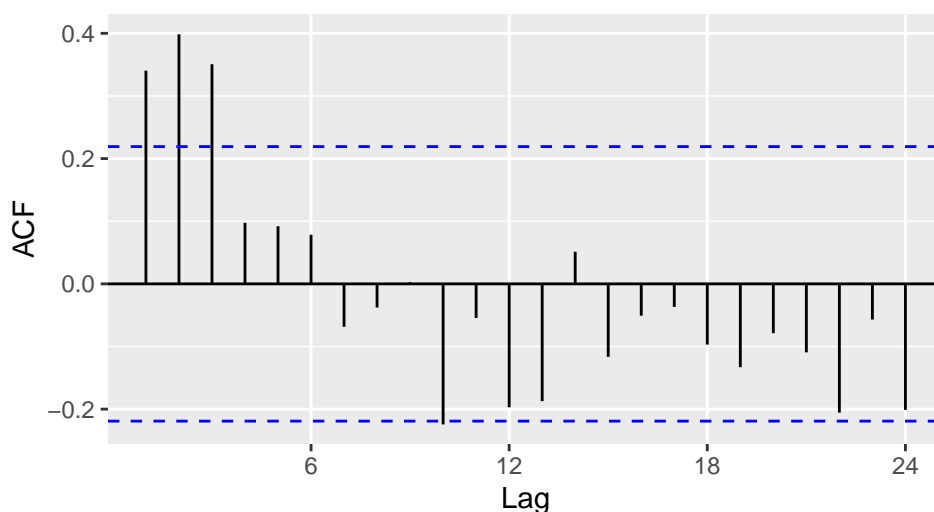
Możemy sprawdzić, jak dane będą się zachowywać po nałożeniu na nie transformacji Boxa-Coxa. Porównaliśmy transformacje dla dwóch różnych wartości λ . Pierwsza transformacja była logarytmiczna ($\lambda = 0$), natomiast dla drugiej transformacji, wartość λ została dobrana automatycznie, według algorytmu zaimplementowanego w R. Korzystając z metody Guerrero do wyboru automatycznego parametru otrzymaliśmy $\lambda = 1.9999242$, czyli górną granicę narzuconą przez funkcję. Po odblokowaniu górnego przedziału wartość ta wynosi 2.5068076.

Wcześniej, w pierwszej oraz trzeciej metodzie otrzymaliśmy z dużym prawdopodobieństwem reszty w postaci szeregu stacjonarnego, w przeciwieństwie do metody drugiej. Dlatego poniżej pokażemy jedynie wyniki połączenia transformaty B-C z metodą wykorzystującą funkcję `tslm`. Jednak w tym przypadku rozpatrywaliśmy trend kwadratowy zamiast liniowego, jak poprzednio. W tabeli poniżej zamieściliśmy wyniki testu ADF różnych transformacji oraz opóźnień. W pierwszej kolumnie znajdują się dane nieprzetransformowane.

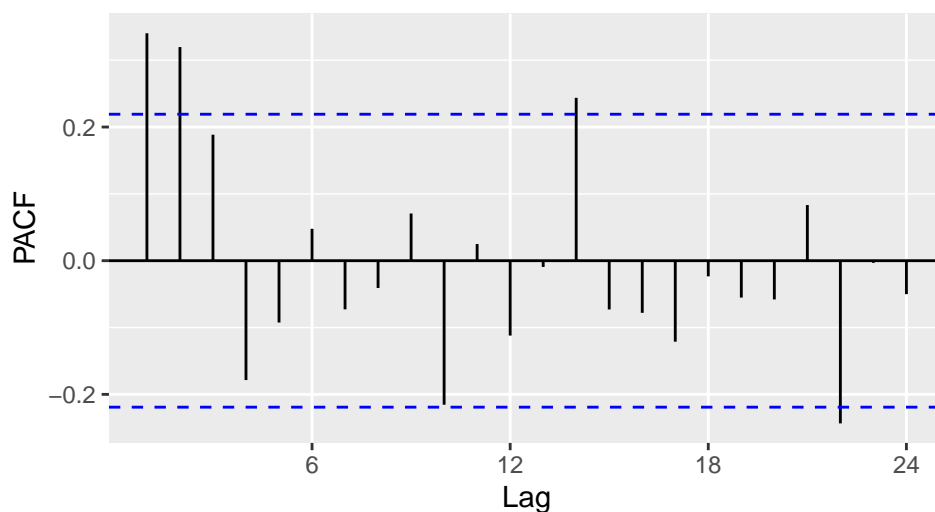
Tablica 2: Tablica p-wartości dla test ADF. Wiersze to wielkość opóźnienia, natomiast kolumny to rodzaj przekształcenia Boxa-Coxa

Dane	Log	2	0.75	2.507
0.0616119	0.0858197	0.0679544	0.0418789	0.0343040
0.3557932	0.3723882	0.3612317	0.3265077	0.3077665
0.1343561	0.1358140	0.1363791	0.1158936	0.1008736
0.0872448	0.0841430	0.0870945	0.0835158	0.0790533

Możemy zwrócić uwagę na zielone wartości w tabeli. Są to wartości uzyskane dla danych przetransformowanych przy użyciu transformacji Boxa-Coxa z parametrem λ dobranym przy pomocy jednej z zaimplementowanych metod (z odblokowanym górnym ograniczeniem). W ten sposób otrzymaliśmy model, dla którego mamy świadka postaci opóźnienia $h = 1$, który świadczy o stacjonarności otrzymanych reszt. Takich informacji nie mieliśmy dla czystych danych (bez użycia B-C). Przyjrzyjmy się jeszcze wykresom acf i pacf dla ostatniej z transformacji ($\lambda = 2.507$).



Wykres 15: Wykres acf dla danych po użyciu transformacji B-C.

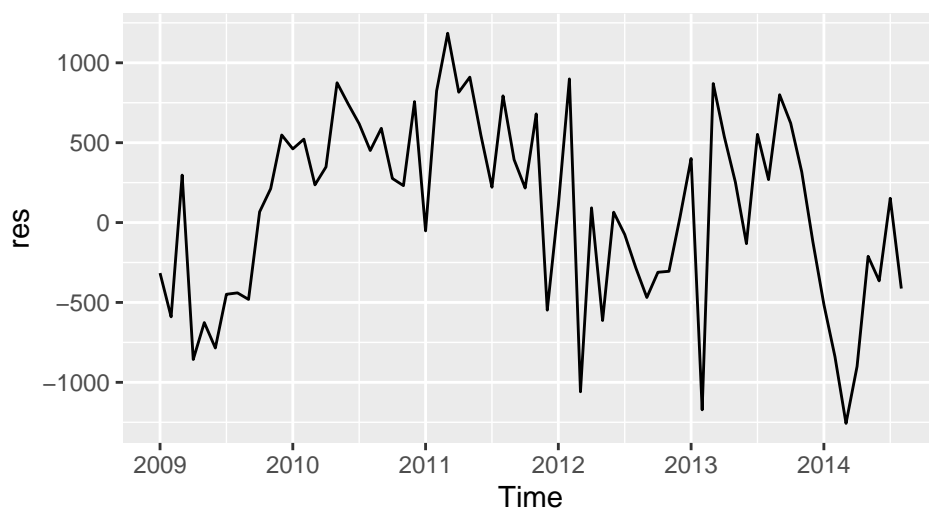


Wykres 16: Wykres pacf dla danych po użyciu transformacji B-C.

Porównując to z poprzednimi wykresami, po transformacji możliwa regularność stała się mniej zauważalna, jednak może to być spowodowane również zwiększeniem stopnia wielomianu trendu.

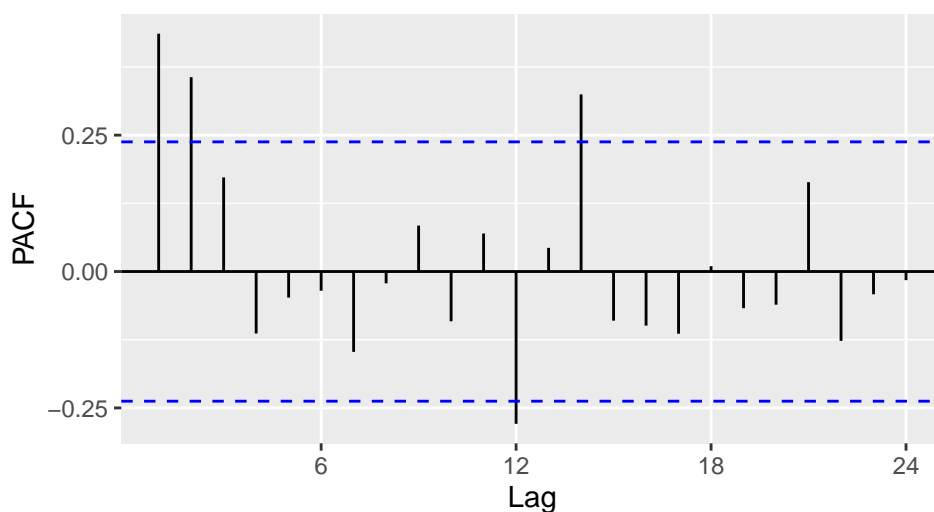
1.4. Różnicowanie.

Ostatnią metodą jaką chcielibyśmy rozpatrzeć jest różnicowanie sezonowe. Ponieważ wykres sezonowy 2 potwierdził zależność sezonową z odstępem $h = 12$ to będziemy właśnie rozpatrywać dane postaci $Z_t = X_t - X_{t-12}$, gdzie X_t to rozpatrywane dane - energia. Poniżej przedstawiliśmy otrzymane dane na wykresie.



Wykres 17: Wykres danych po zróżnicowaniu z opóźnieniem $h=12$

Na wykresie możemy zauważyć zachowanie sugerujące, że pozostał jednak jakiś trend. Możemy spojrzeć jeszcze na wykres funkcji PACF.



Wykres 18: Wykres pacf dla zróżnicowanych danych.

Możemy zauważyć znaczące wychylenie od zera dla opóźnienia $h = 14$, wskazujące, że nie mamy do czynienia z szeregiem stacjonarnym. Wykonaliśmy jeszcze test ADF, który zwrócił nam p-wartość na poziomie 0.2714654 zatem nie mamy podstaw do odrzucenia hipotezy o niestacjonarności reszt. Podobnie próbowaliśmy ponownie różnicować rozpatrywane dane oraz sprawdzaliśmy jak zachowują się, gdy rozpatrzymy iloraz w postaci Z_t/Z_{t-1} , jednak otrzymane wyniki nie były zadowalające.

1.5. Podsumowanie

Spośród rozpatrywanych metod dekompozycji najlepiej poradziły sobie związane z funkcjami `decompose` oraz `stl`. Jednak metoda drug, związana z funkcją `tslm` próbowała dopasować wzór analityczny do trendu. Na podstawie tej metody pokazaliśmy, że transformacja Boxa-Coxa może poprawić nasze dane – pozwoliła otrzymać szereg reszt, który przeszedł test na stacjonarność (ADF test) dla opóźnienia $h = 1$. Rozpatrzyliśmy jeszcze dane po zróżnicowaniu sezonowym oraz z różnymi opóźnieniami. Wyniki okazały się mniej zadowalające niż otrzymane pozostałymi metodami, ale jest to metoda najprostsza w implementacji oraz najmniej zmieniająca dane podczas transformacji (nie rozdziela je osobno na trend, sezonowość itp.).

2. Zastosowanie wskaźników analizy technicznej w analizie szeregów czasowych.

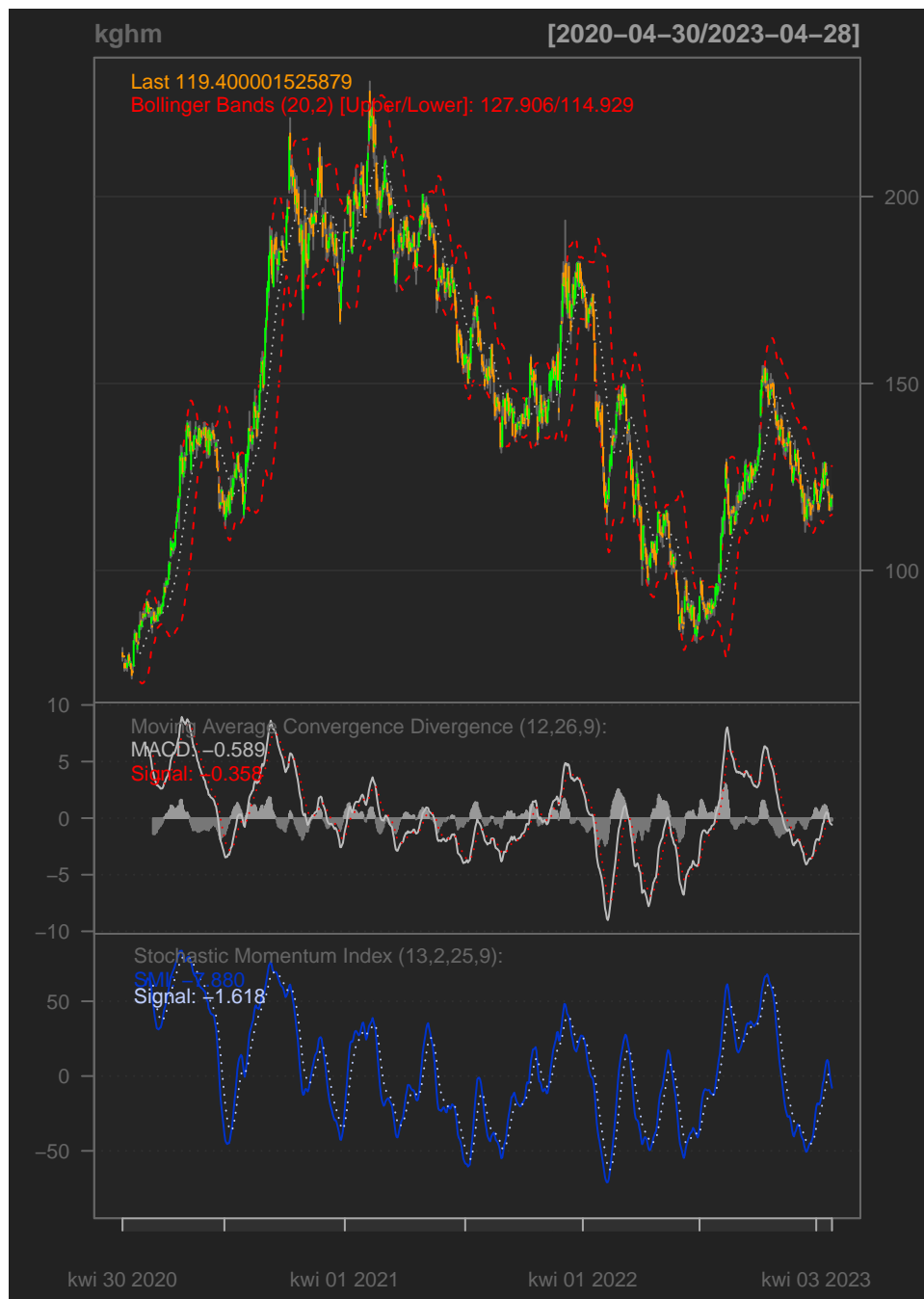
W celu analizy rozpatrujemy akcje firmy KGHM Polska Miedź z okresu od 30 kwietnia 2020 do dnia pisania tego raportu tj. 30 kwiecień 2023 roku. Dane prezentują się następująco.

```
## [1] "KGH.WA"
```



W danych (cenach zamknięcia) znaleźliśmy 15 brakujących danych, ale postanowiliśmy zignorować tę informację.

Możemy teraz przystąpić do analizy technicznej. Na poniższym wykresie zamieściliśmy wszystkie wyniki dla podstawowych parametrów (według zaimplementowanych algorytmów).



Na wykresie zamieszczone zostały wskaźniki takie jak:

- Wstęga bolingera – dwie czerwone i przerywane linie nałożone na pierwszy wykres. Jeśli cena spadnie poniżej dolnej linii jest to sygnał do kupna, jeśli przednie górną, to otrzymujemy sygnał do sprzedaży. Podczas analizy rozważymy zmianę dwóch parametrów w funkcji
 - n – odpowiadającą za liczbę okresów ruchomej średniej,
 - sd – odchylenie standardowe – odpowiadające za wielkość przedziałów.

- MACD – Moving Average Convergence Divergence – jest to wskaźnik dynamiki trendów, który pokazuje związek pomiędzy dwiema średnimi kroczącymi. Sygnał kupna jest generowany, gdy linia MACD przetnie linię sygnałną od dołu, a sygnał sprzedaży generowany jest wtedy, kiedy linia MACD przetnie linię sygnałną od góry.
- SMI – Stochastic Momentum Index – oblicza odległość między aktualną ceną zamknięcia w stosunku do mediany górnego/dolnego przedziału cenowego. Podczas analizy rozpatrzmy zmienianie parametru n odpowiedzialnego za długość rozważanego okresu.

Możemy teraz spojrzeć na ich interpretację. W naszych danych mamy duże okresy bardzo gwałtownych zmian, co jest do przewidzenia, zważając, że jest to okres pandemii. Dlatego wstęga Bollingera słabo nadaje się w tym przypadku. Każde sprzedawać, podczas gdy cena będzie dalej rosła, a kupować gdy cena ciągle spada. Oczywiście zdarzają się momenty, gdy wstęga dobrze nam doradziła, przykładowo w marcu 2021 roku. Podobnie jest z kolejnymi metodami. W wielu przypadkach źle oceniły sytuację. Jeżeli jednak ograniczymy się jedynie do ostatnich miesięcy, możemy zauważyć, że metoda MACD na przełomie stycznia i lutego nakazała nam sprzedaż, co rzeczywiście się nam opłacało w tym okresie. Podobnie z resztą metoda SMI.

2.0.1. Wstęga Bolingera

Porównajmy Wstęgę Bolingera na poniższych animacjach.

Wykres 19: Wstęga Bolingera dla różnych wartości parametrów.

Wykres 20: Wstęga Bolingera dla różnych wartości parametrów.

Możemy zauważyć, że wraz ze wzrostem parametry n zwiększa się wielkość wstęgi (odległość między obiema liniami), ale również przesuwa się delikatnie w prawo oraz staje się bardziej wygładzona. Natomiast zmiana parametru sd w głównej mierze zwiększa wielkość wstęgi.

2.0.2. MACD

Rozpatrzmy teraz jak zachowa się funkcja MACD w przypadku zmiany jej parametrów.

Wykres 21: MACD dla różnych wartości parametrów.

Wykres 22: MACD dla różnych wartości parametrów.

Jak widzimy, podobnie jak w przypadku wstęgi Bollingera i parametru n , parametry $nSlow$ i $nFast$ zwiększają wygładzenia krzywych. Jednak można zauważyć, że zwiększając wartość parametru $nSlow$ zwiększamy maksymalne wartości funkcji $macd$ oraz $signal$. Natomiast dla większych parametrów $nFast$ działanie jest odwrotne, wraz z zwiększeniem parametru maleją wartości obu funkcji.

2.0.3. SMI

Ostatnią rozpatrywaną funkcją będzie SMI. Wyniki prezentują się następująco

Wykres 23: SMI dla różnych wartości parametrów.

Również w tym przypadku widzimy, że zmiana parametru n zwiększa wygładzenie funkcji oraz zwiększenie wartości obu funkcji, analogicznie jak działa się w przypadku parametru $nSlow$.