

19.06.2022, Rzeszów

Szeregi czasowe - projekt

Jonasz Koczynasz, Kierunek: FS-DI, Nr indeksu: 166658



SPIS TREŚCI

| | |
|--|----|
| Wprowadzenie do dziedziny analizowanych problemów | 1 |
| Omówienie głównych cech analizowanych szeregów na podstawie wykresów | 2 |
| Wygładzenie danych poprzez użycie średniej ruchomej | 8 |
| Dekompozycja na podstawie modelu regresji | 9 |
| Dekompozycja addytywna i multiplikatywna | 12 |
| Eliminacja sezonowości oraz trendu, stacjonarność szeregu | 14 |
| Wyznaczenie rzędu p oraz q modeli $AR(p)$ oraz $MA(q)$ | 17 |
| Wyznaczenie współczynników modelu AR , porównanie dopasowania z wykorzystaniem różnych metod estymacji, automatycznie dobrana wartość rzędu. | 19 |
| Wyznaczenie współczynników dla modelu $MA(q)$. | 21 |
| Wyznaczenie optymalnych modeli z wykorzystaniem funkcji <code>auto.arima()</code> oraz wyznaczenie ich współczynników, Porównanie analizowanych modeli, wybór najlepszego. | 22 |
| Prognozowanie z wykorzystaniem metod naiwnych, dobór najlepszej metody dla danego szeregu. | 25 |
| Wprowadzenie do analizowanych danych | 29 |
| Utworzenie szeregów czasowych z średnimi: dziennymi, miesięcznymi, kwartalowymi (Po odpowiednim „wycięciu” potrzebnych danych, interpolacji wartości brakujących) | 29 |
| Reprezentacja szeregu czasowego na różnego rodzaju wykresach | 30 |
| Dekompozycje na podstawie modelu regresji: trend liniowy/wielomianowy, transformacje Boxa-Coxa | 34 |
| Uczynienie szeregu stacjonarnym – eliminacja trendu, sezonowości poprzez różnicowanie i stabilizacja wariancji poprzez transformację Boxa-Coxa | 36 |
| Sprawdzenie stacjonarności otrzymanego szeregu | 41 |
| Wyznaczenie współczynników modelu autoregresji i porównanie dopasowania różnymi metodami estymacji | 41 |
| Wyznaczenie współczynników modelu ruchomej średniej z użyciem funkcji <code>Arima()</code> | 42 |
| Porównanie modelu ruchomej średniej z modelem autoregresji, analiza dobroci dopasowania | 43 |
| Wyznaczenie optymalnych modeli z wykorzystaniem funkcji <code>auto.arima()</code> , oraz ich porównanie na podstawie kryteriów dopasowania, wybór najlepszego | 44 |
| Test stacjonarności automatycznie wybranego modelu | 45 |
| Prognozowanie z wykorzystaniem metod naiwnych | 46 |

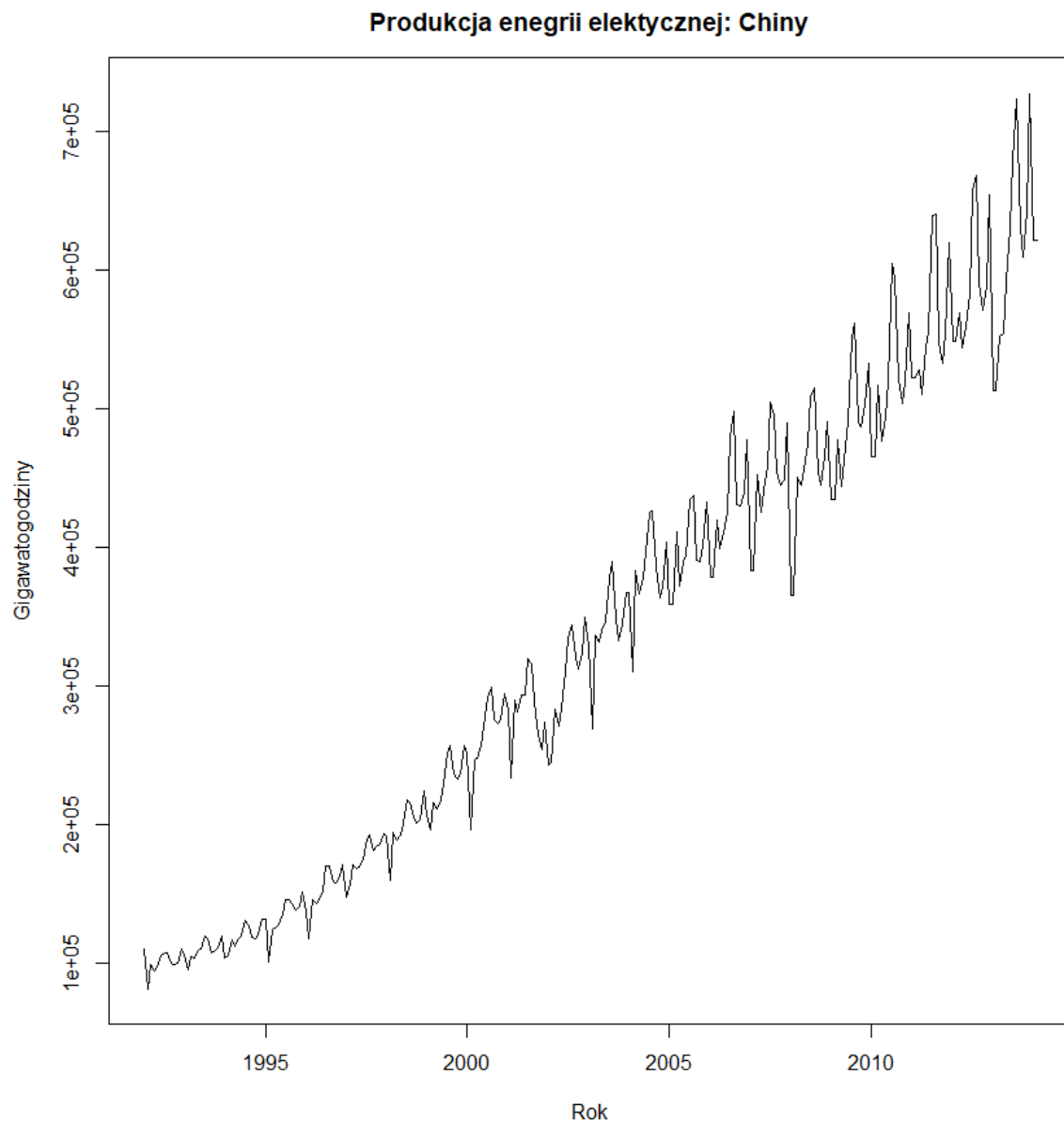
WPROWADZENIE DO DZIEDZINY ANALIZOWANYCH PROBLEMÓW

Pierwszym szeregiem, którym będę się zajmować jest szereg dotyczący produkcji energii elektrycznej w Chinach.

Dane pochodzą ze strony <https://fred.stlouisfed.org/>. Obejmują okres od 01.01.1999 do 01.03.2022. Dane są o częstotliwości miesięcznej. Jednostka: Gigawatogodziny. Chiny to kraj z największą ilością wytwarzanej energii elektrycznej, a tym samym kraj o największej ilości emisji CO₂ do atmosfery. Odpowiadają one za ponad połowę światowej energii z węgla. Mimo to Chiny mają duży potencjał do stania się liderem w dziedzinie energetyki odnawialnej i innych zielonych rozwiązań. Mam nadzieję, że podczas analizy tego szeregu będę mógł zobaczyć, w których miesiącach zużycie energii było na najwyższym poziomie, a w którym na najniższym oraz wysnuć potencjalne przyczyny takich spadków/wzrostów.

OMÓWIENIE GŁÓWNYCH CECH ANALIZOWANYCH SZEREGÓW NA PODSTAWIE WYKRESÓW

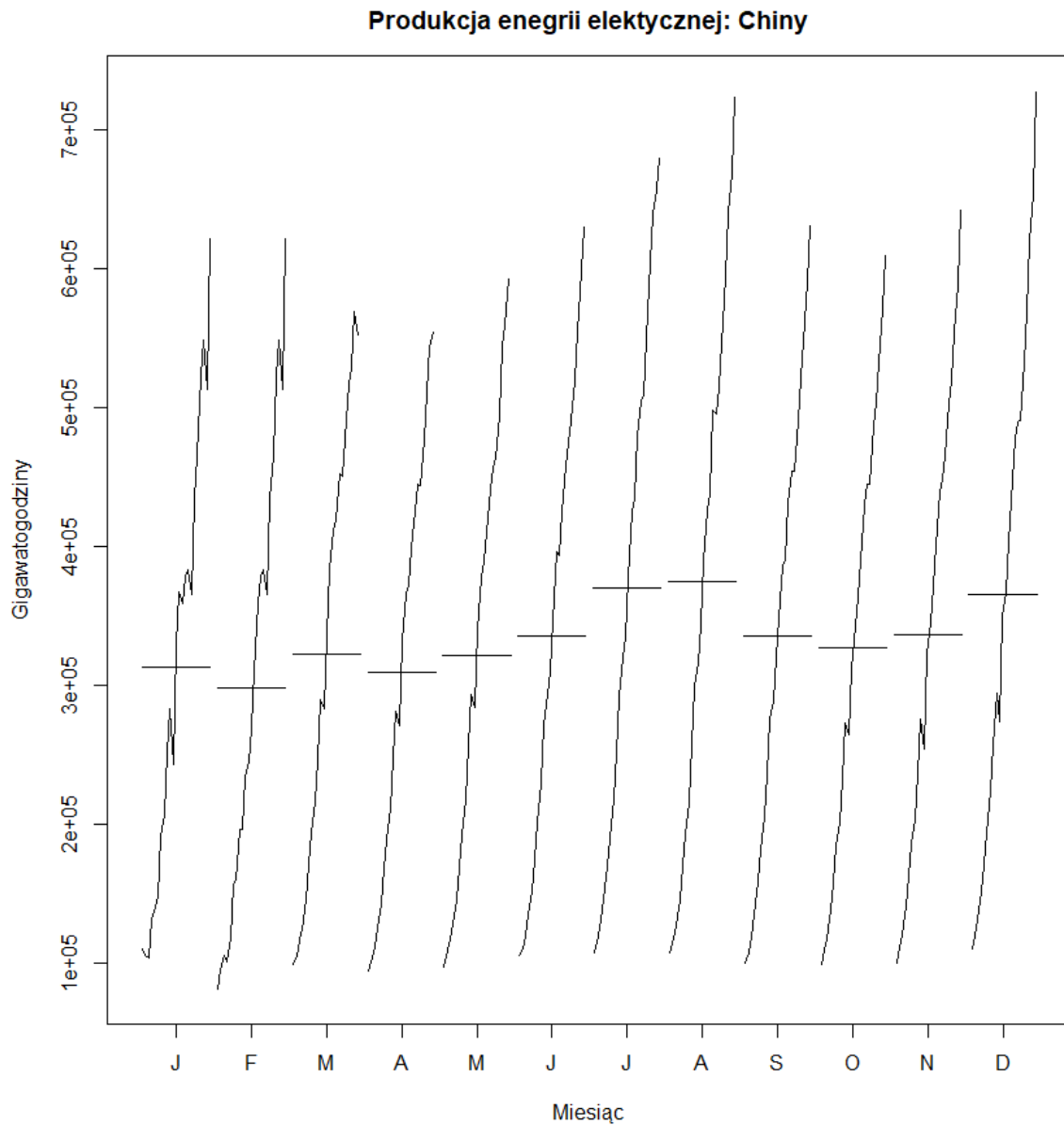
Wykres szeregu ilości produkowanej energii elektrycznej w Chinach:



Wnioski:

- Możemy zobaczyć że wykres rośnie (ma trend wzrostowy), jednak ciężko na razie mówić o sezonowości. Nie ma brakujących danych.

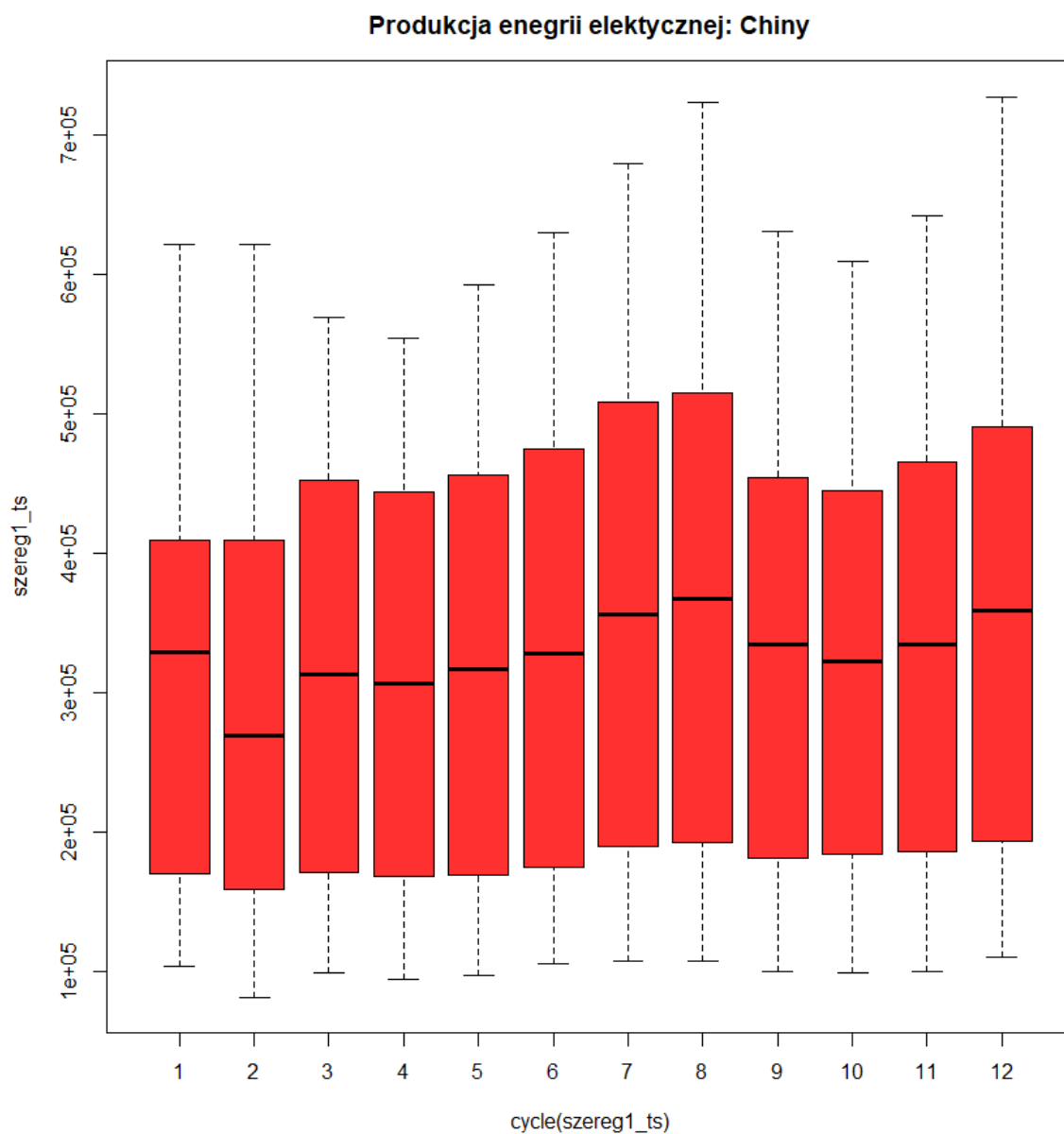
Wykres miesięcznej ilości produkowanej energii elektrycznej w Chinach:



Wnioski:

- Widzimy że w każdym miesiącu jest trend rosnący. Średnio najwięcej energii elektrycznej było produkowane w sierpniu, a najmniej w lutym. Wysokie zużycie energii elektrycznej w sierpniu może być spowodowane używaniem klimatyzatorów biorąc pod uwagę że jest to najcieplejszy miesiąc w Chinach.

Wykres pudełkowy ilości produkowanej energii elektrycznej w Chinach:

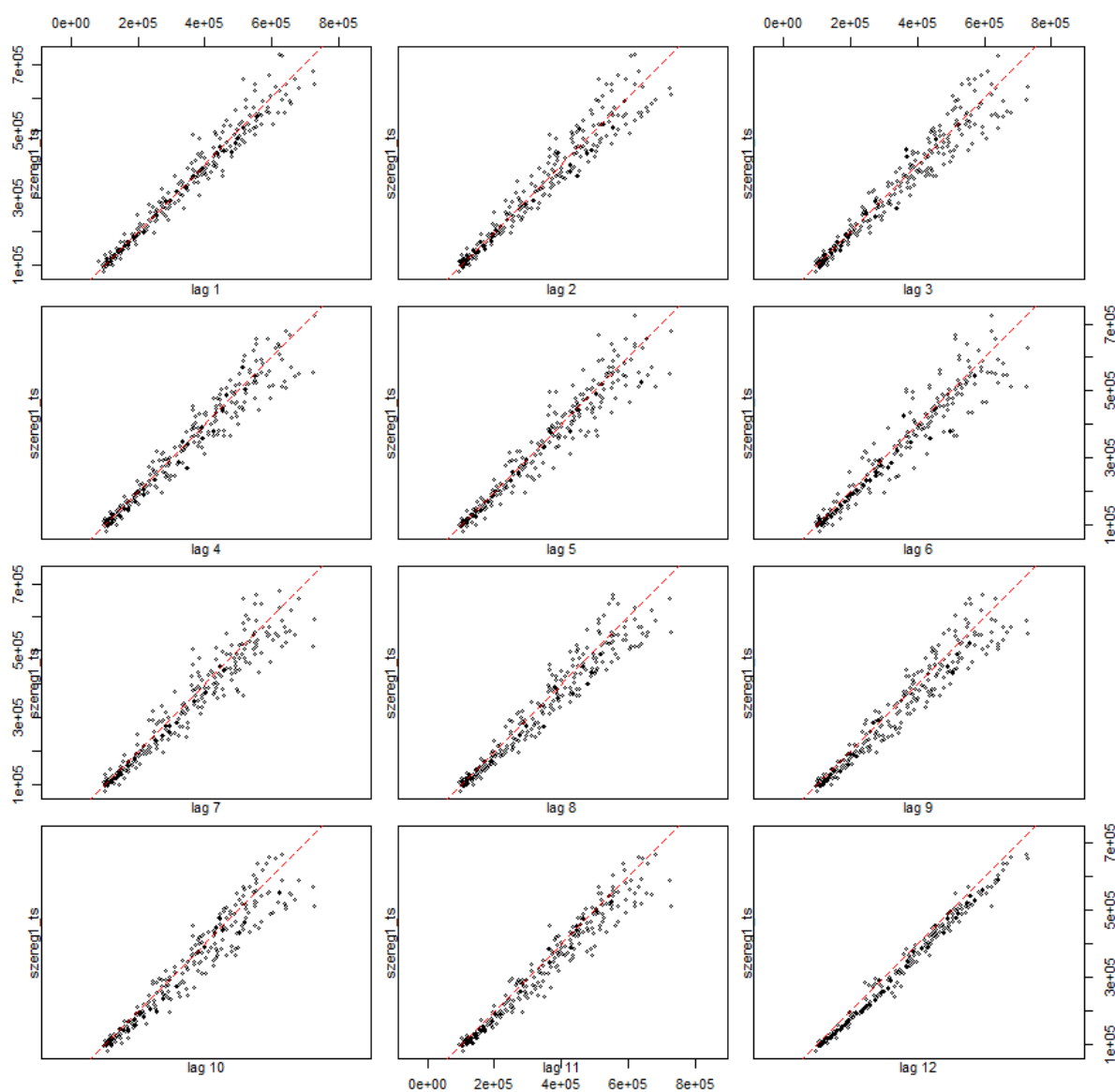


Wnioski:

- Nie mamy wartości odstających.
- Wartość mediany jest największa dla stycznia, a najmniejsza dla lutego.
- Wysokość pudełek jest relatywnie duża, co świadczy o tym że dane są rozrzucone.

Wykresy rozrzutu ilości produkowanej energii elektrycznej w Chinach:

Produkcja energii elektrycznej: Chiny

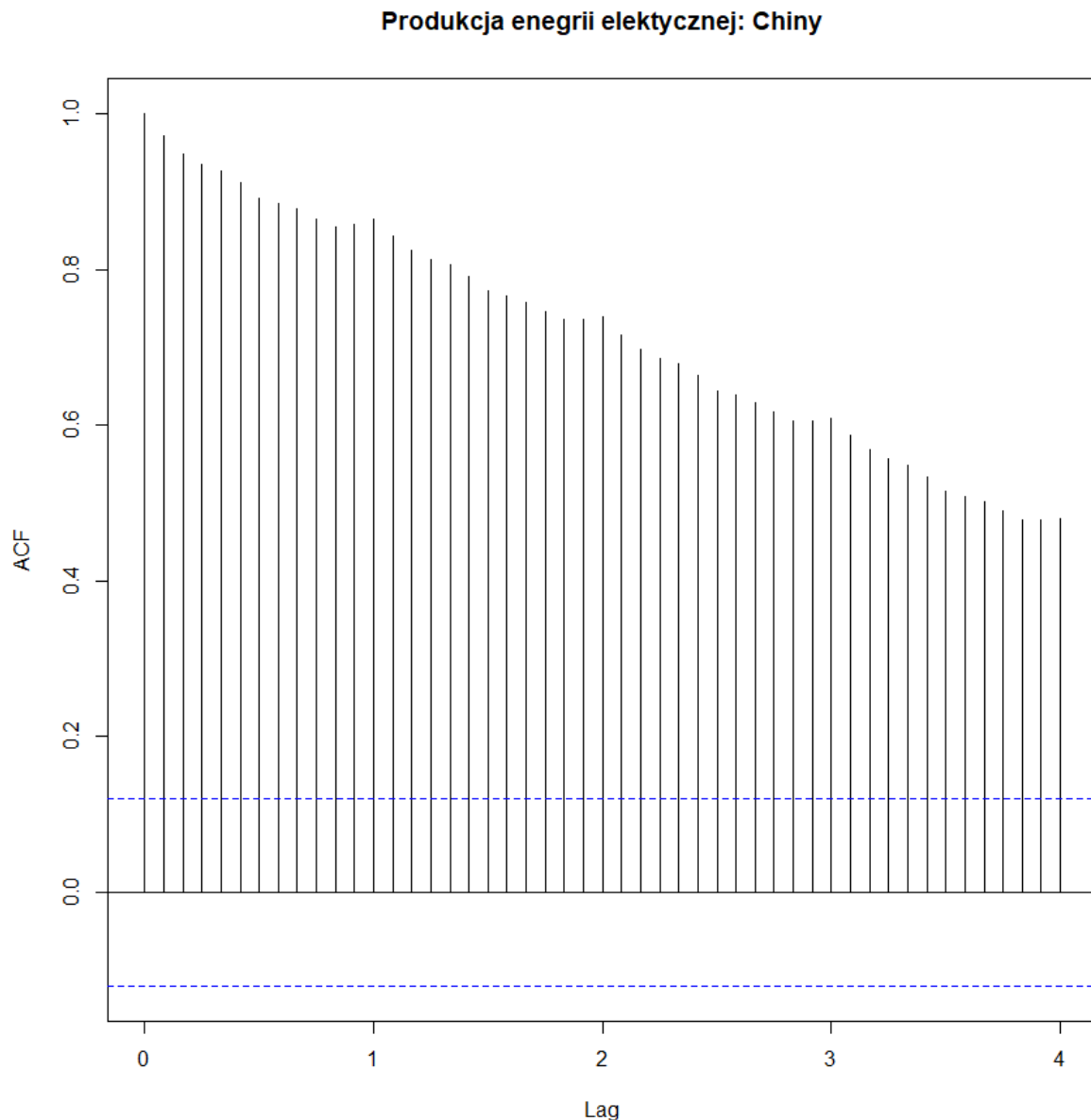


Wnioski:

- Na wykresie lag 1 widzimy jednorodną chmurę punktów co świadczy o tym że szereg ma trend.

- Dla opóźnienia 12 możemy zauważyć najsilniejszą zależność co sugeruje sezonowość roczną sezonowość roczna.

Wykres autokorelacji ilości produkowanej energii elektrycznej w Chinach:

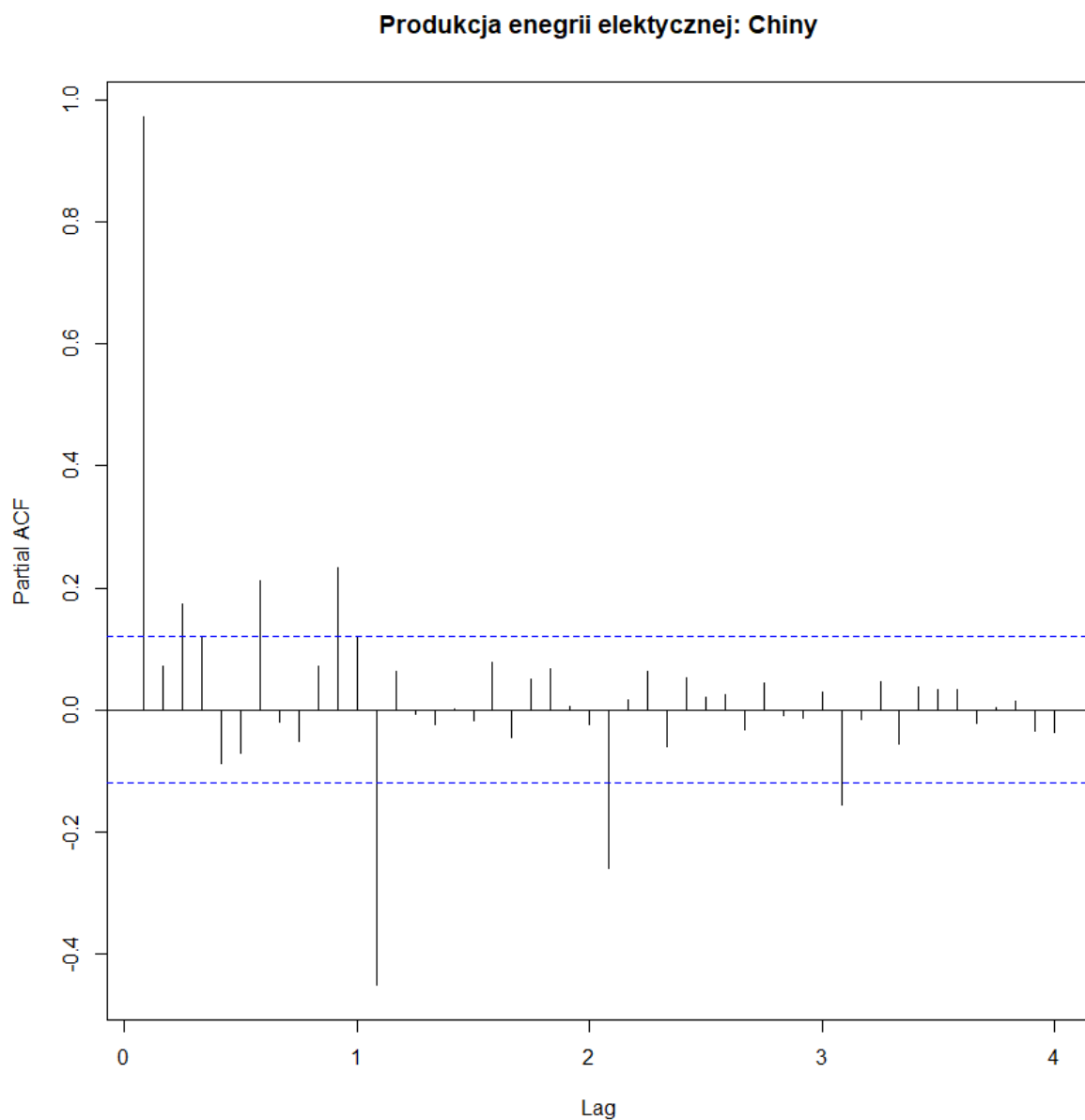


Wnioski:

- Na wykresie możemy zobaczyć dodatnie i powoli zanikające wartości funkcji ACF, które sugerują, że dane zawierają deterministyczną składową trendu.

- Możemy również zauważyć że wartości ACF zanikają bardzo powoli i cyklicznie, co sugeruje na obecność sezonowości - trendu sezonowego.
- Na wykresie znajdują się „delikatne góry”, które świadczą o sezonowości.

Wykres częściowej autokorelacji ilości produkowanej energii elektrycznej w Chinach:



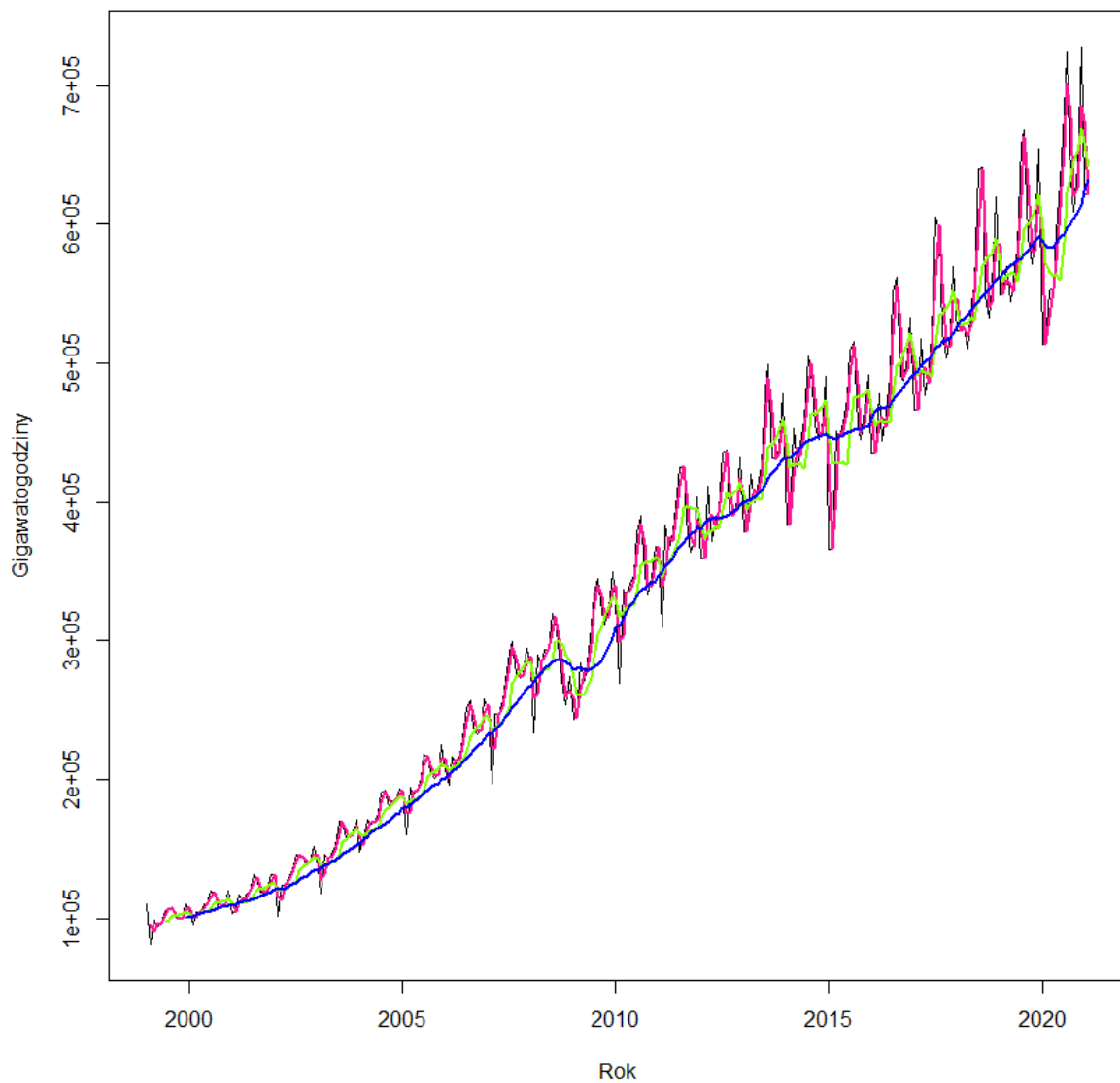
Wnioski:

-Duża wartość dla początkowego opóźnienia świadczy o obecności silnego trendu wzrostowego.

WYGŁADZENIE DANYCH POPRZECZ UŻYCIEM ŚREDNIEJ RUCHOMEJ

Zastosowałem technikę wygładzania za pomocą ruchomej średniej.

Produkcja energii elektrycznej: Chiny

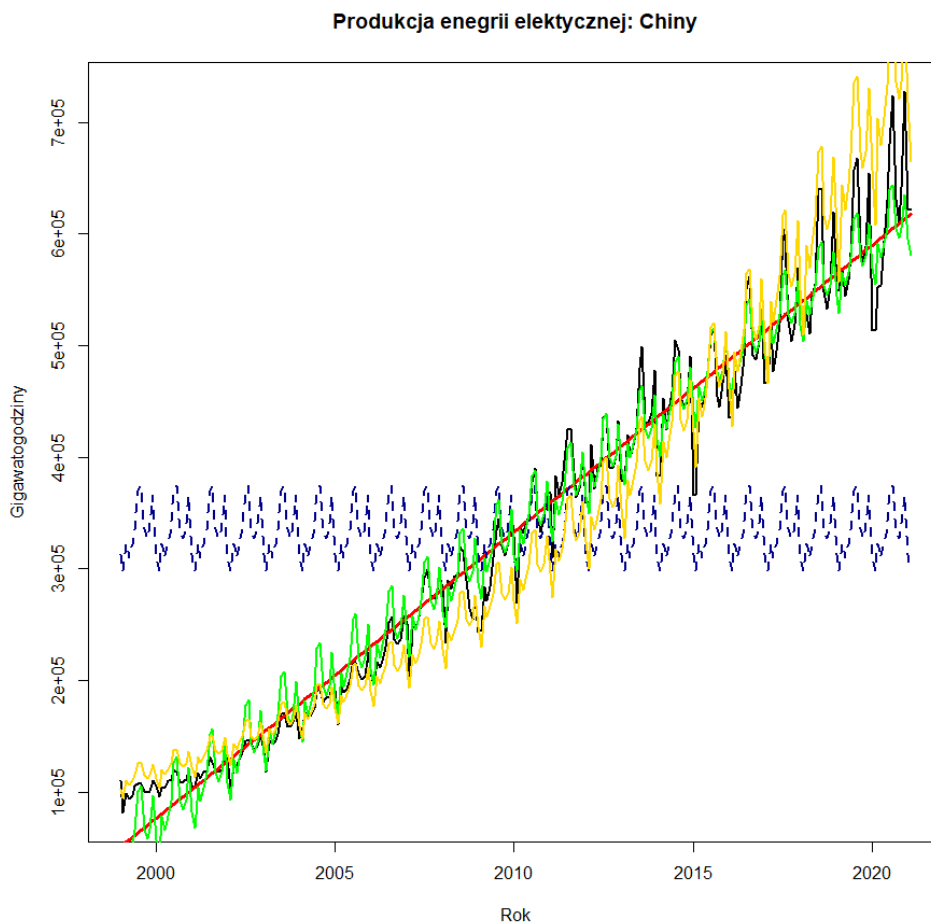


Wnioski:

- Linia różowa reprezentuje średnią 2-miesięczną i jest ona bardzo zbliżona do podstawowego wykresu funkcji.
- Linia zielona reprezentuje średnią 6-miesięczną i bardzo dobrze wygładza podstawowy wykres.
- Linia niebieska reprezentuje średnią 12-miesięczną i wskazuje na tendencję wzrostową. Dzięki zastosowaniu średniej 12-miesięcznej najbardziej wygładziliśmy wykres.

DEKOMPOZYCJA NA PODSTAWIE MODELU REGRESJI

Wykres utworzony na podstawie modelu regresji – trend liniowy, sezonowość:

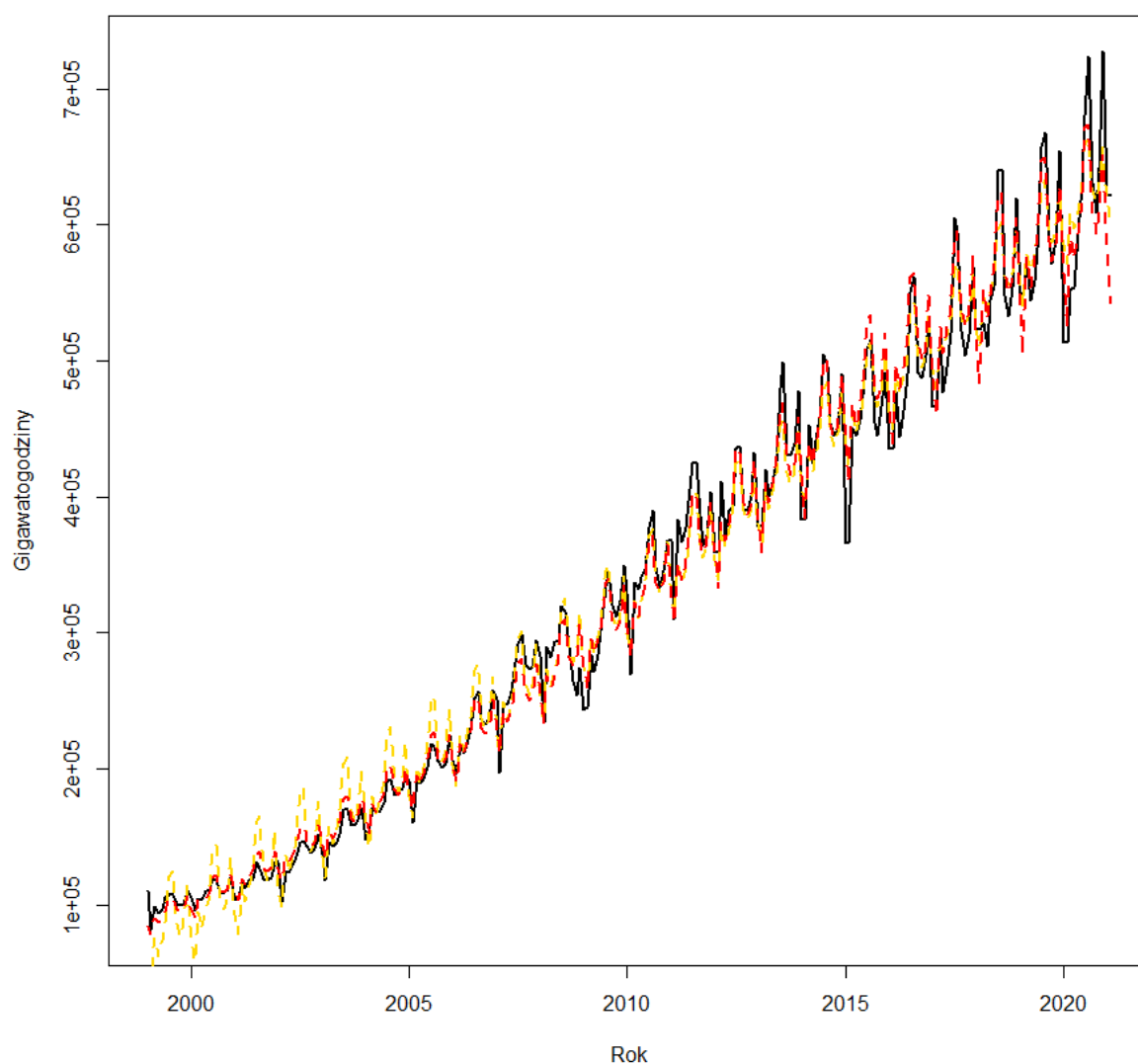


Wnioski:

- Kolorem czerwonym zaznaczona jest linia trendu. Widzimy że trend jest silnie rosnący.
- Linia niebieska przerywana wizualizuje model regresji liniowej sezonowości szeregu czasowego.
- Linia zielona wizualizuje model regresji liniowej trendu oraz sezonowości szeregu czasowego.
- Linia żółta wizualizuje model regresji transformacji Box-Coxa
- Myślę że warto rozważyć trend stopnia drugiego.

Następnie dokonałem dekompozycji na podstawie modelu regresji - trend wielomianowy:

Produkcja energii elektrycznej: Chiny



Wnioski:

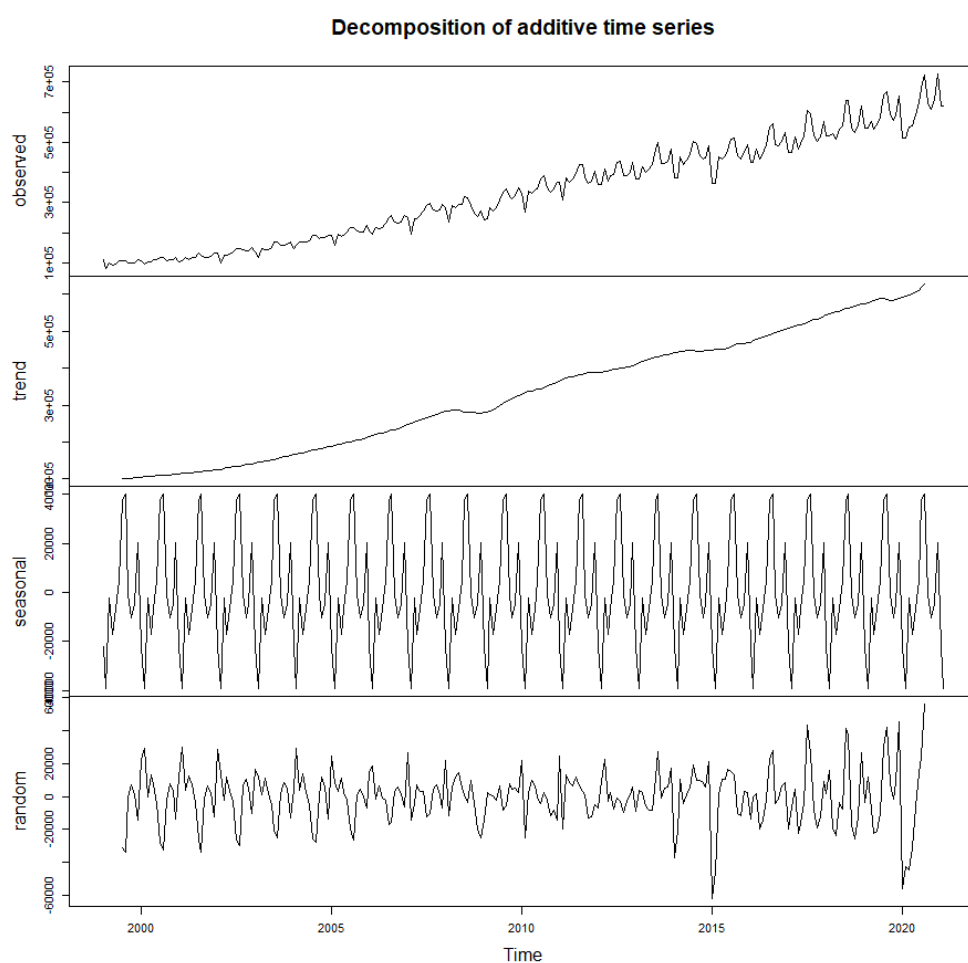
- Linia czerwona jest wizualizacją modelu regresji kwadratowej trendu oraz sezonowości szeregu czasowego.

- Linia czerwona jest wizualizacją modelu kwadratowej funkcji trendu po transformacji Box-Coxa.

- Widać że najlepszym dopasowaniem jest linia czerwona czyli model kwadratowej funkcji trendu po transformacji Box-Coxa.

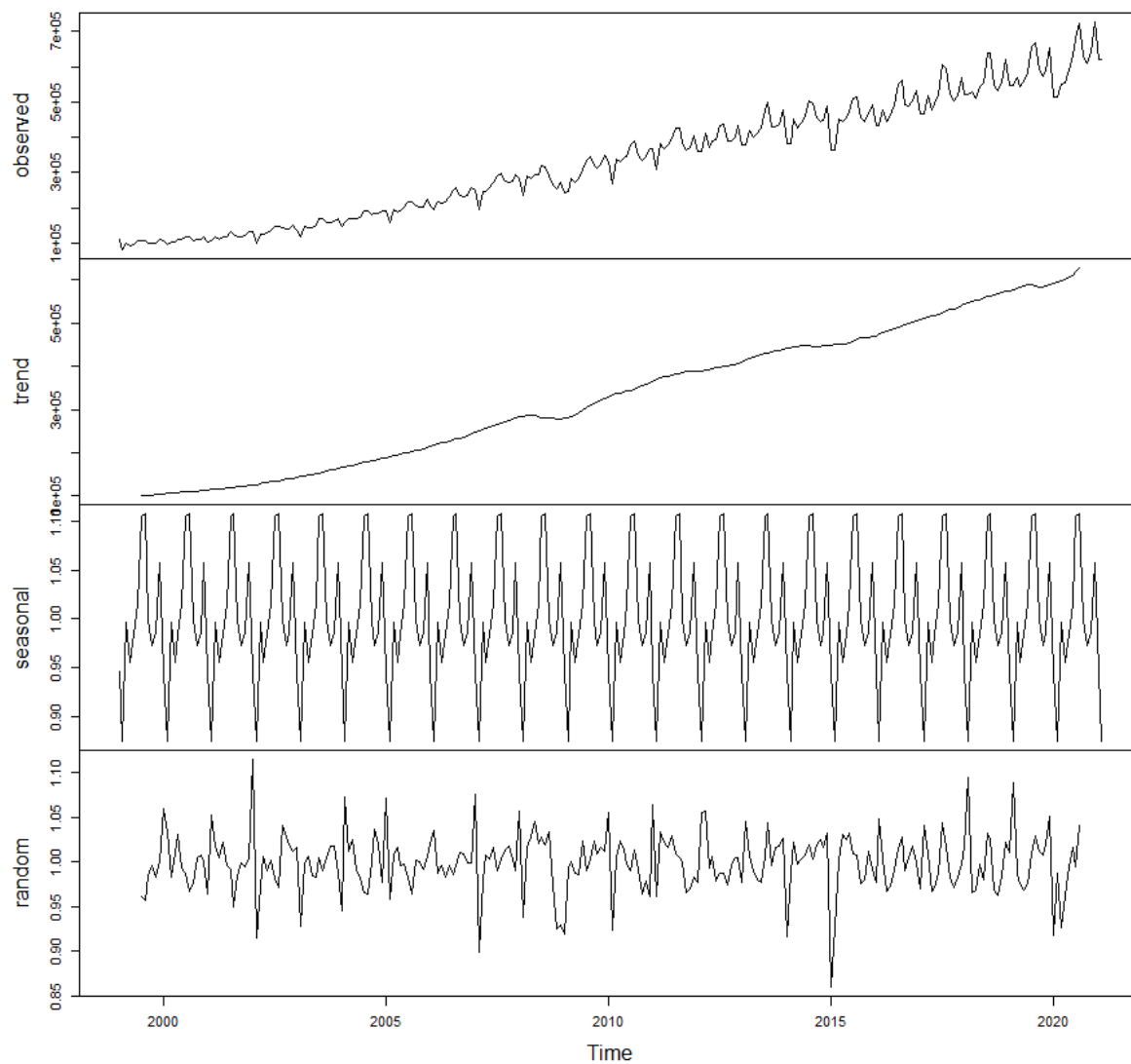
DEKOMPOZYCJA ADDYTYWNA I MULTIPLIKATYWNA

Wykres dekompozycji addytywnej:



Wykres dekompozycji multiplikatywnej:

Decomposition of multiplicative time series



Wnioski:

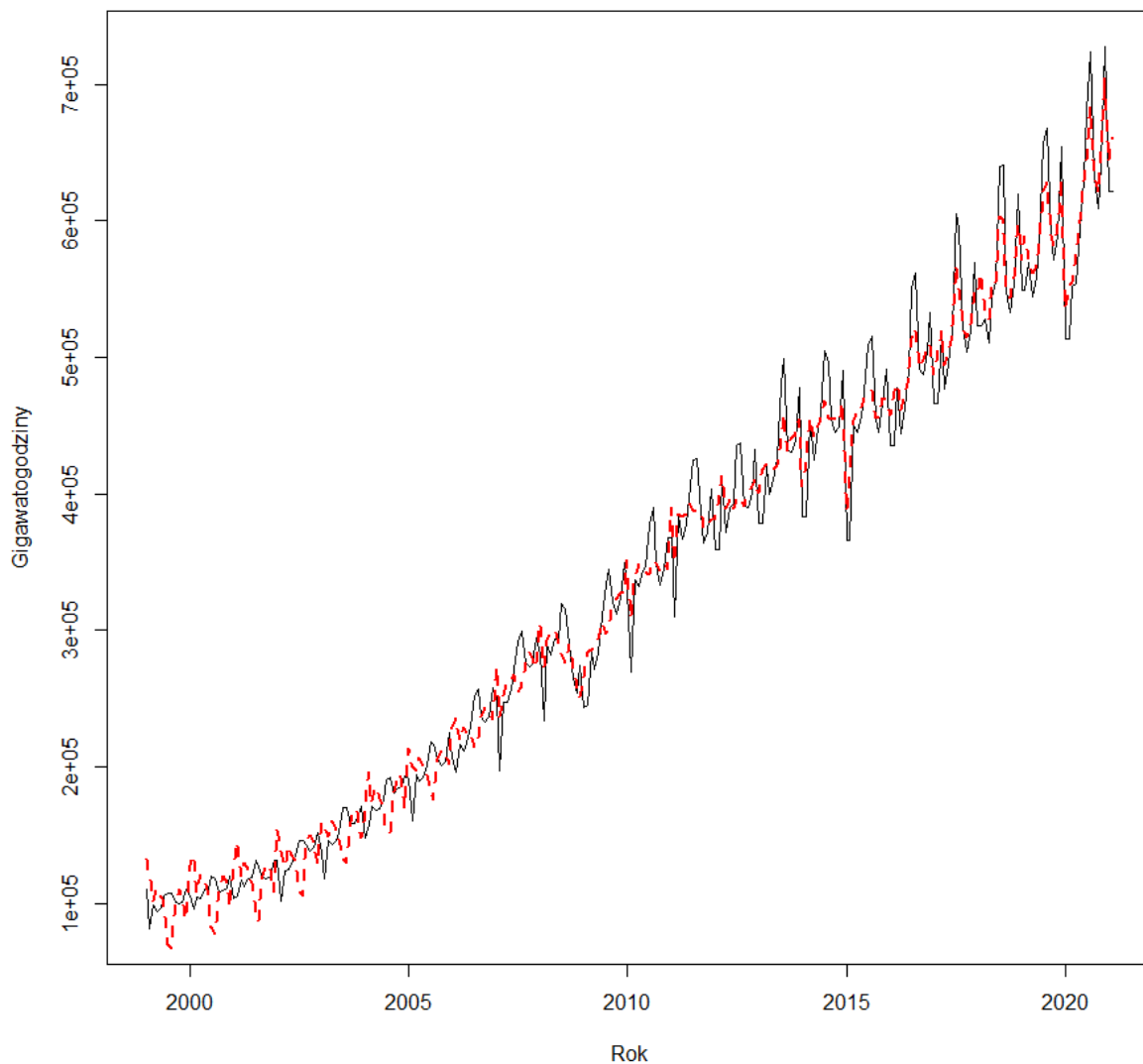
- Myślę że lepszym wyborem jest wybranie dekompozycji multiplikatywnej.



ELIMINACJA SEZONOWOŚCI ORAZ TRENDU, STACJONARNOŚĆ SZEREGU

Najpierw wyeliminowałem sezonowość przed identyfikację trendu. Zrobiłem to dzięki funkcji seasadj.

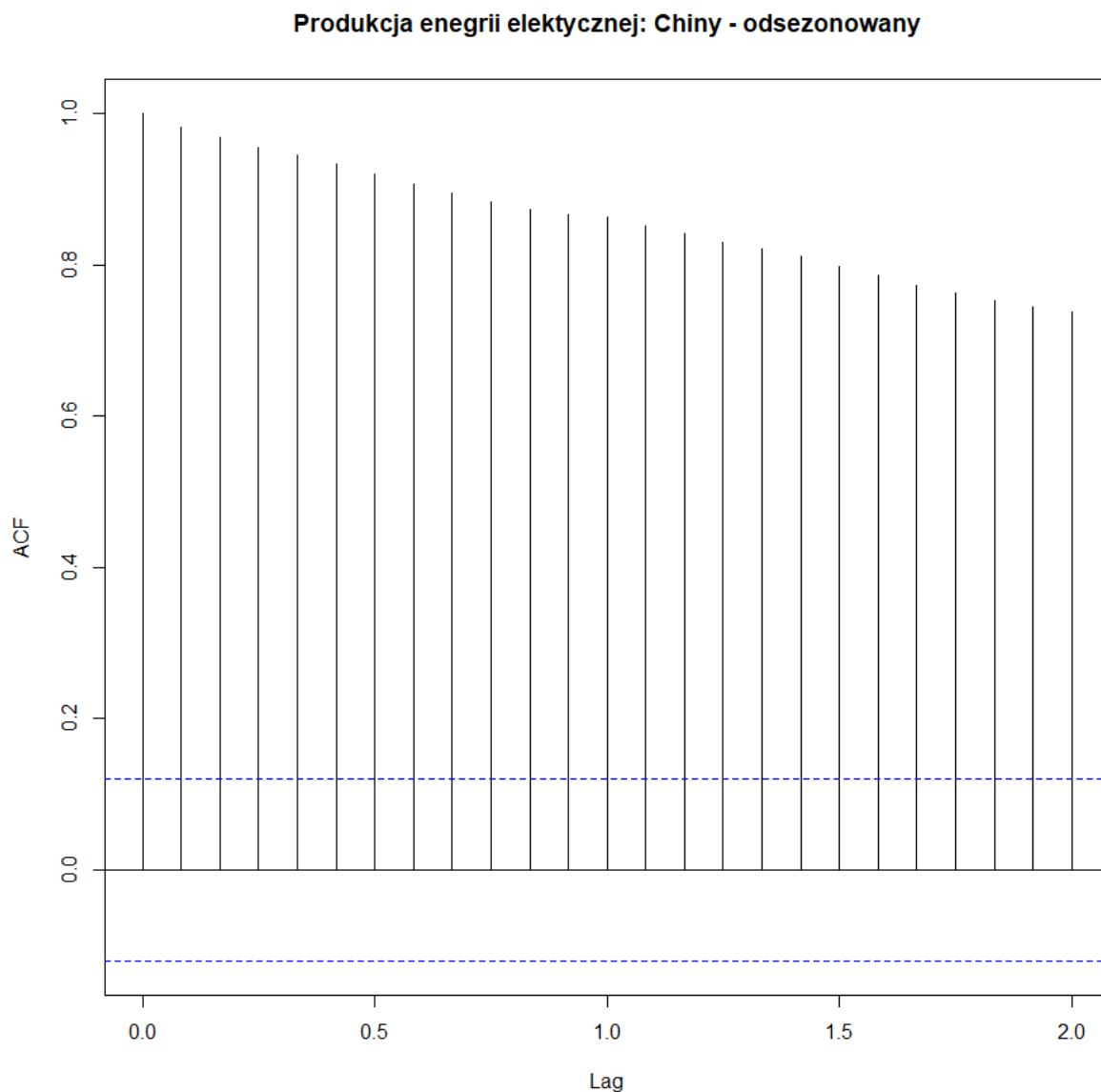
Produkcja energii elektrycznej: Chiny



Wnioski:

- Patrząc tylko na ten wykres możemy stwierdzić że został on wygładzony jednak czy sezonowość została usunięta na pewno przekonamy się za chwilę.

Tak wygląda wykres autokorelacji:



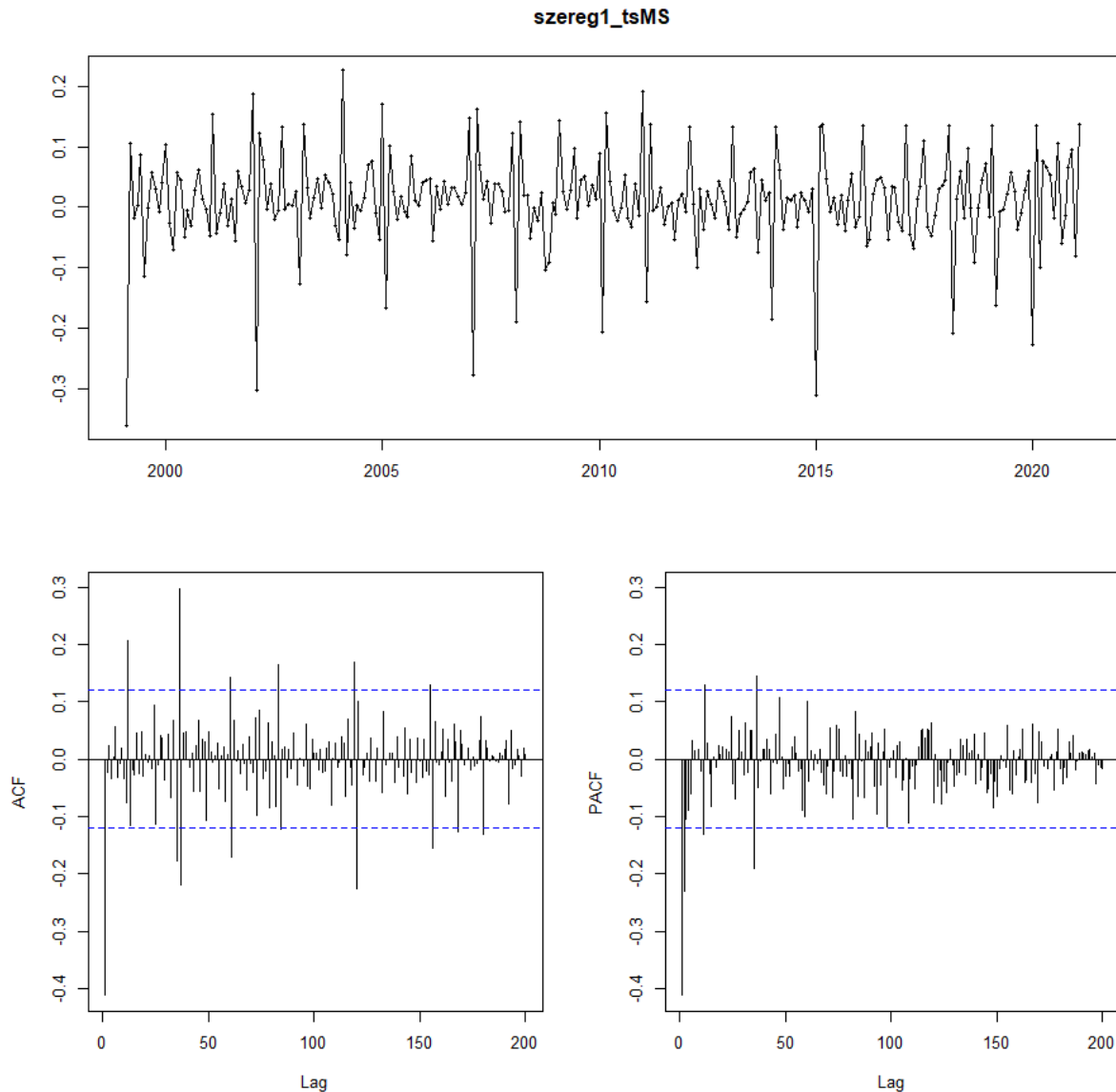
Wnioski:

- Widac, że sezonowość faktycznie się zmniejszyła.

Następnie przystąpiłem do usuwania trendu.

Pierwsze co zrobiłem to zastosowałem transformację Boxa-Coxa z automatycznie ustawionym parametrem λ , a następnie zróżnicowałem utworzony szereg czasowy z lagiem równym 1.

Wykres tsdisplay po przekształceniach:



Wnioski:

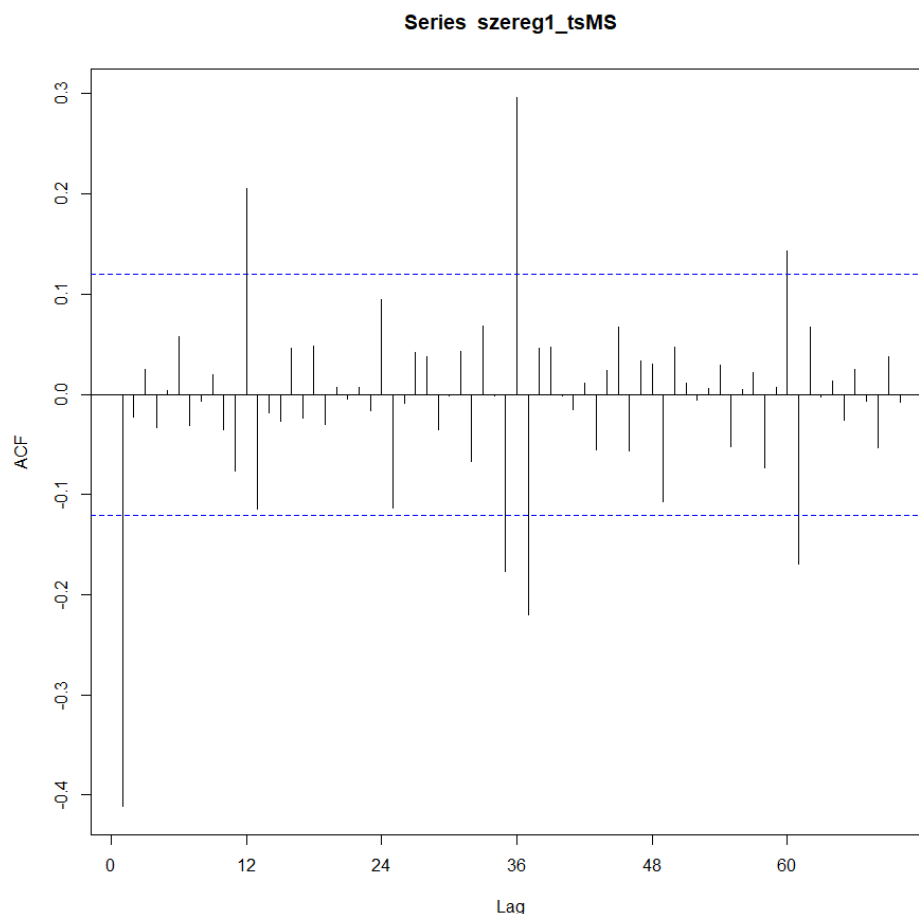
- Jak widać bardzo ciężko będzie dopasować do tego szeregu model ruchomej średniej albo model autoregresji. Rząd takiego modelu musiałby być bardzo wysoki.

Sprawdzałem również czy szereg jest stacjonarny. Przeprowadziłem test adf. Okazało się że nie ma podstaw do odrzucenia hipotezy zerowej na poziomie istotności 0.01, gdzie hipoteza zerowa: szereg nie jest stacjonarny.

WYZNACZENIE RZĘDU P ORAZ Q MODELI $AR(P)$ ORAZ $MA(Q)$

Wykorzystując szereg z usuniętą sezonowością oraz trendem wyznaczyłem rzędy p oraz q modeli $AR(p)$ oraz $MA(q)$

Najpierw wyświetliłem wykres autokorelacji:

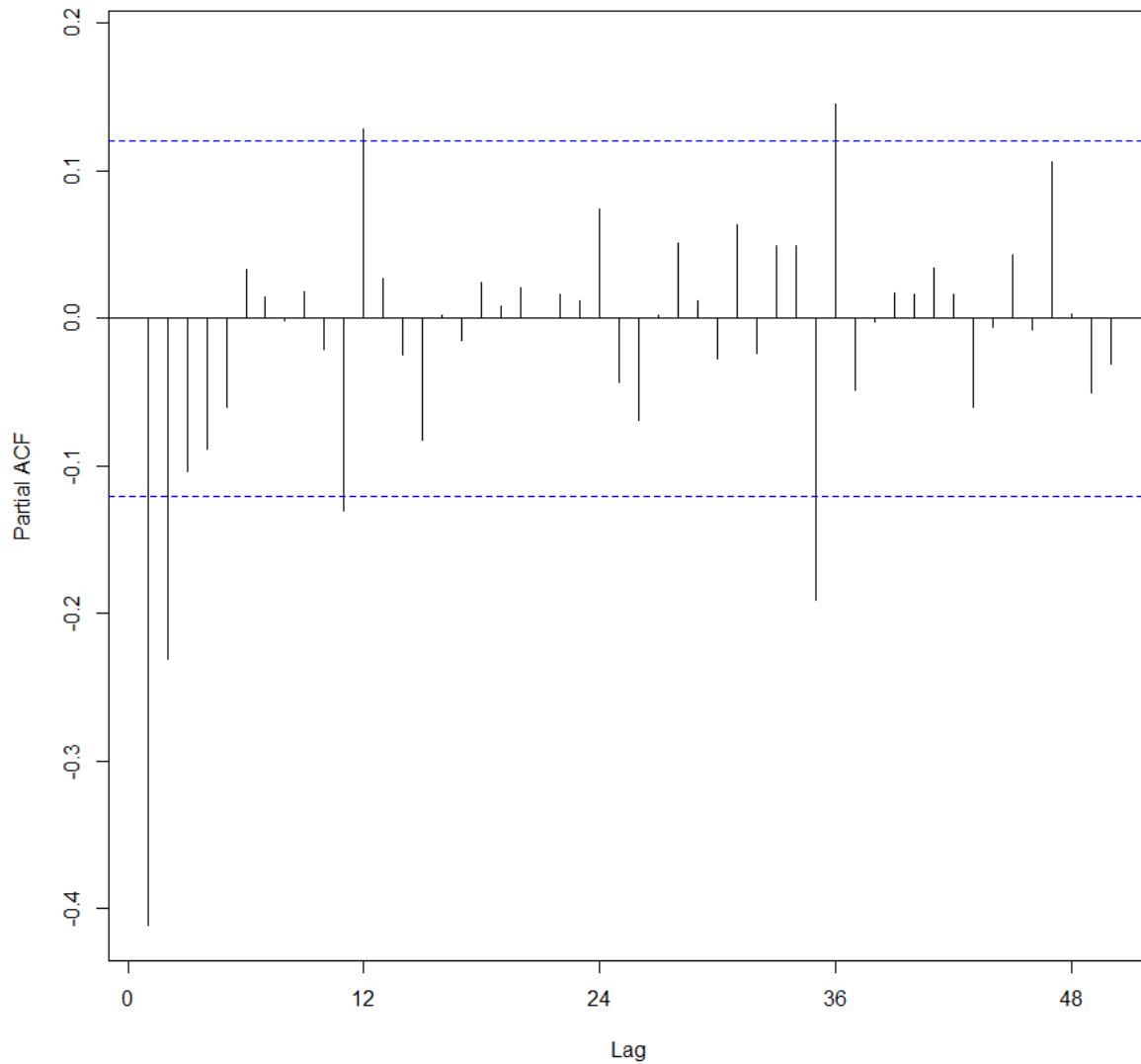


Wnioski:

-Warto sprawdzić $MA(61)$, $MA(37)$, $MA(35)$.

Następnie wyświetliłem wykres częściowej autokorelacji

Series szereg1_tsMS



Wnioski:

- Warto sprawdzić AR(35), AR(2), AR(11).

WYZNACZENIE WSPÓŁCZYNNIKÓW MODELU AR, PORÓWNANIE
DOPASOWANIA Z WYKORZYSTANIEM RÓŻNYCH METOD ESTYMACJI,
AUTOMATYCZNIE DOBRANA WARTOŚĆ RZĘDU.

- Do wyznaczenia współczynników modelu AR wykorzystałem funkcję `ar`, aby dopasowywać modele autoregresji do szeregów czasowych. Wybrałem sobie do dalszej analizy rząd 11.

Wybierając metodę yule-walker dostawiliśmy współczynnik:

```
Coefficients:
      1      2      3      4      5      6      7      8      9     10     11
-0.5448 -0.3115 -0.1622 -0.1048 -0.0315  0.0359  0.0015 -0.0197 -0.0344 -0.0920 -0.1303
```

Wybierając metodę ols dostawiliśmy współczynnik:

```
Coefficients:
      1      2      3      4      5      6      7      8      9     10     11
-0.6133 -0.3972 -0.2547 -0.1783 -0.1177 -0.0319 -0.0479 -0.0520 -0.0686 -0.1242 -0.1530
```

Wybierając metodę mle dostawiliśmy współczynnik:

```
Coefficients:
      1      2      3      4      5      6      7      8      9     10     11
-0.6336 -0.4033 -0.2518 -0.1850 -0.0932 -0.0230 -0.0530 -0.0625 -0.0722 -0.1334 -0.1644
```

Wybierając metodę yw dostawiliśmy współczynnik:

```
Coefficients:
      1      2      3      4      5      6      7      8      9     10     11
-0.5448 -0.3115 -0.1622 -0.1048 -0.0315  0.0359  0.0015 -0.0197 -0.0344 -0.0920 -0.1303
```

Wybierając metodę burg dostawiliśmy współczynnik:

```
Coefficients:
      1      2      3      4      5      6      7      8      9     10     11
-0.6027 -0.3798 -0.2345 -0.1706 -0.0801 -0.0133 -0.0444 -0.0527 -0.0630 -0.1248 -0.1577
```

Wnioski:

- Widzimy że współczynniki nie zawsze są takie same, jednak oscylują wokół tych samych wartości.

-Współczynniki wyznaczone metodami yw oraz yule-walker są takie same.

- Następnie zastosowałem funkcję `ar` z automatycznie dobraną wartością rzędu dla metody yule-walker. Dostaliśmy współczynnik:

```
Coefficients:
      1      2      3      4
-0.5387 -0.3079 -0.1507 -0.0887
```

Wnioski:

- Okazało się że rzędem najlepiej dopasowanym jest rząd 4.

- Spróbowałem też zastosować funkcję `ar` z automatycznie dobraną wartością rzędu dla innych metod tj. `burg`, `yw`, `mle`.

Wnioski:

- Współczynnik w każdej z tych metod różniły się od siebie.
- Okazało się dla metody `ols` otrzymujemy rząd 11.
- Dla metod `burg` oraz `yw` otrzymujemy rząd 4.

WYZNACZENIE WSPÓŁCZYNNIKÓW DLA MODELU MA(Q).

Przy użyciu funkcji `Arima` wyznaczyłem współczynniki modelu ruchomej średniej:

```

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8      ar9      ar10     ar11     mean
-0.6336 -0.4033 -0.2518 -0.1850 -0.0931 -0.0230 -0.0529 -0.0625 -0.0722 -0.1334 -0.1644 0.0120
s.e.    0.0638  0.0746  0.0790  0.0807  0.0815  0.0819  0.0817  0.0808  0.0792  0.0751  0.0642 0.0014

```

Wnioski:

- Współczynniki są takie same jak przy metodzie `mle`.

**WYZNACZENIE OPTYMALNYCH MODELI Z WYKORZYSTANIEM FUNKCJI
AUTO.ARIMA() ORAZ WYZNACZENIE ICH WSPÓŁCZYNNIKÓW,
PORÓWNANIE ANALIZOWANYCH MODELI, WYBÓR NAJLEPSZEGO.**

- Dla kryterium aicc:

Funkcja `auto.arima` dopasowała model ruchomej średniej rzędu dwa z sezonowością (ARIMA(1,0,2)(0,0,1)[12] z zerową średnią).

Wyznaczone współczynniki:

```
Coefficients:
      ar1      ma1      sma1      mean
    -0.0073  -0.6064   0.2052   0.0121
s.e.    0.1080    0.0877   0.0700   0.0019
```

- Dla kryterium aic dostajemy te same wyniki.
- Dla kryterium bic:

Otrzymujemy inny wynik. Funkcja `auto.arima` dopasowała model ruchomej średniej rzędu jeden z sezonowością (ARIMA(1,0,1)(0,0,1)[12] z zerową średnią).

Współczynniki:

```
Coefficients:
      ar1      ma1      sma1      mean
    -0.0073  -0.6064   0.2052   0.0121
s.e.    0.1080    0.0877   0.0700   0.0019
```

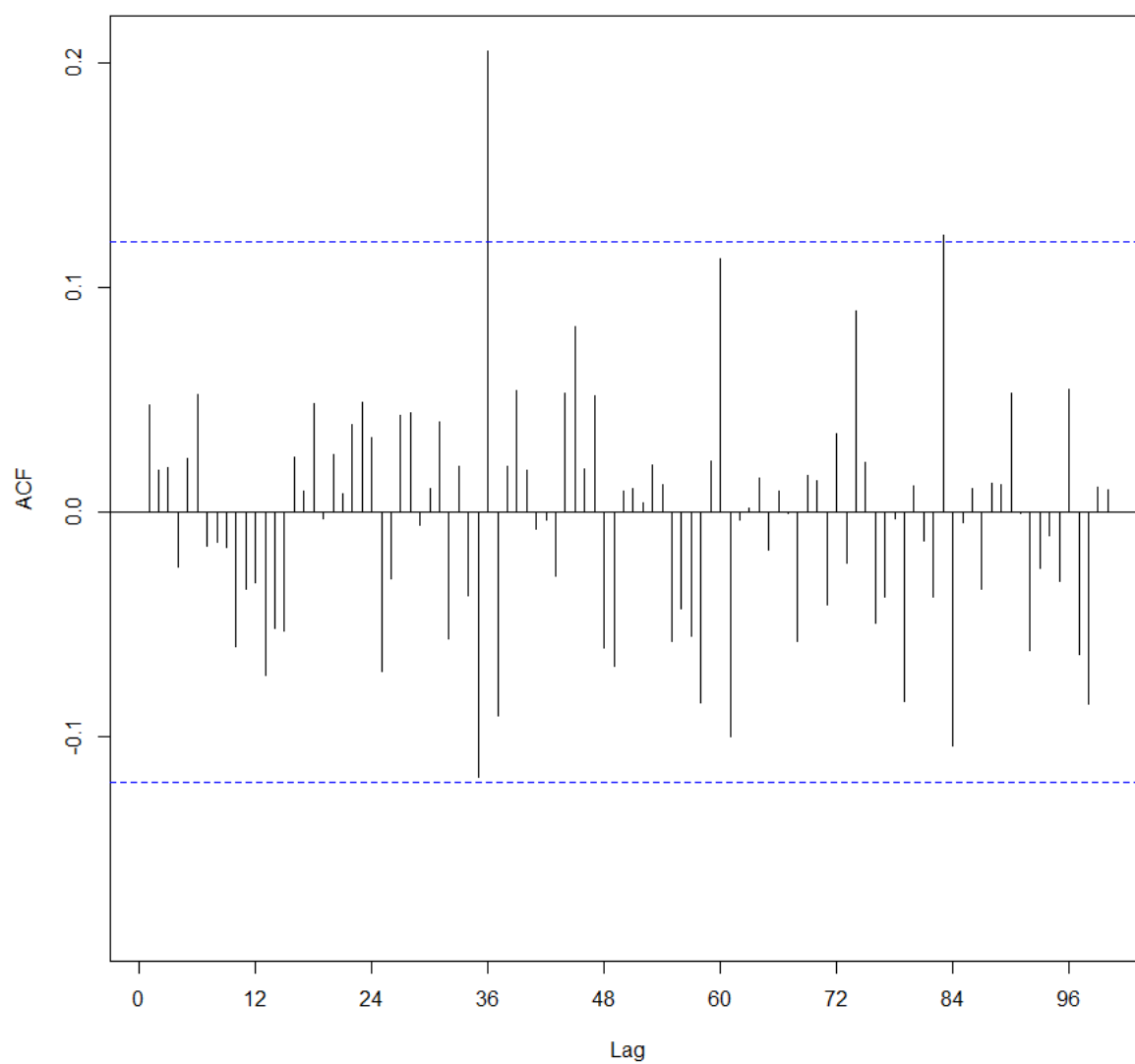
Wnioski:

- Patrząc na wartości AIC, AICc, BIC możemy stwierdzić że model ruchomej średniej rzędu jeden z sezonowością (ARIMA(1,0,1)(0,0,1)[12] z zerową średnią) jest to najbardziej optymalny model, ponieważ wartości tych parametrów są bliższe zera.

Następnie sprawdziłem czy model można uznać za realizacją szumu białego.

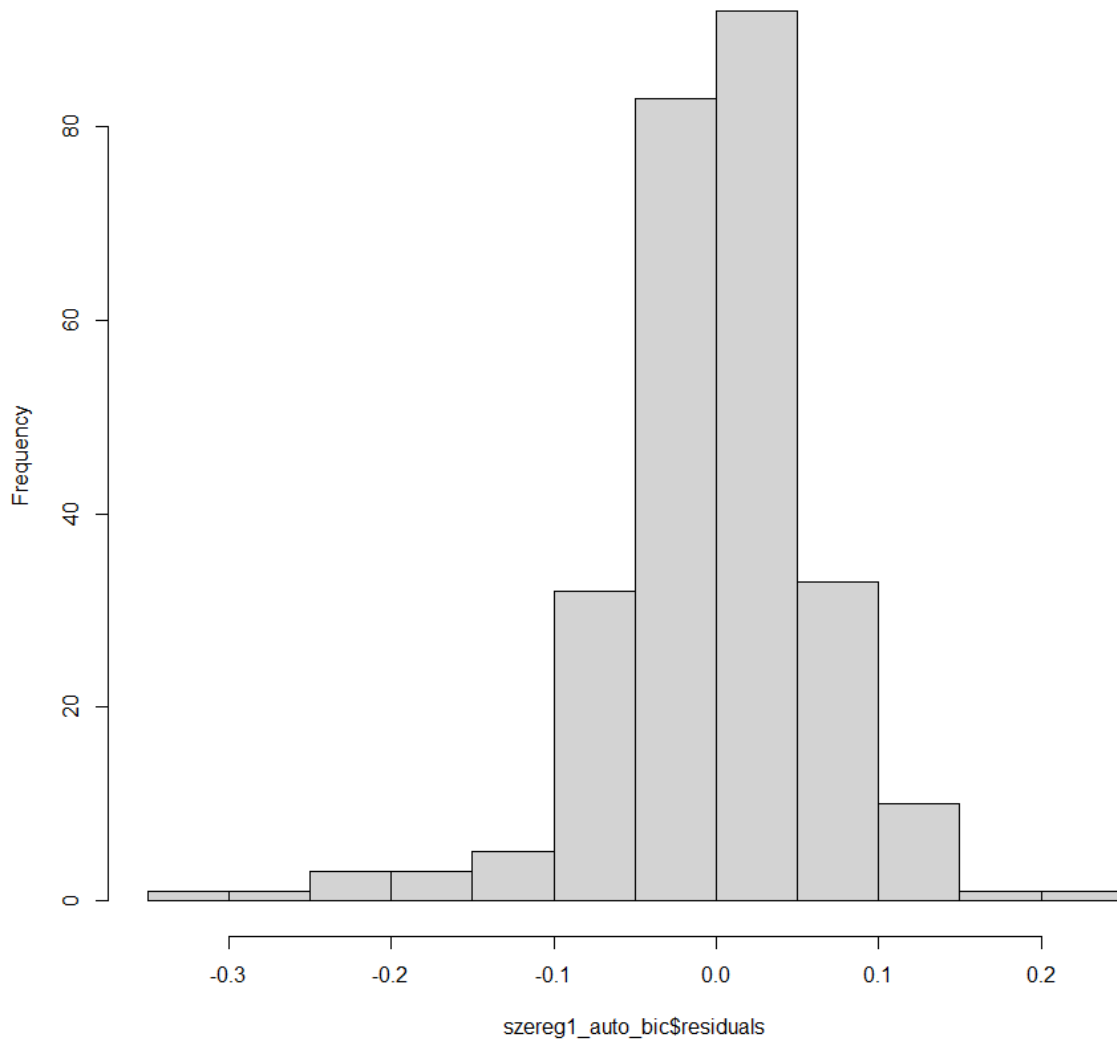
Wykres autokorelacji:

Series szereg1_auto_bic\$residuals



Histogram:

Histogram of szereg1_auto_bic\$residuals



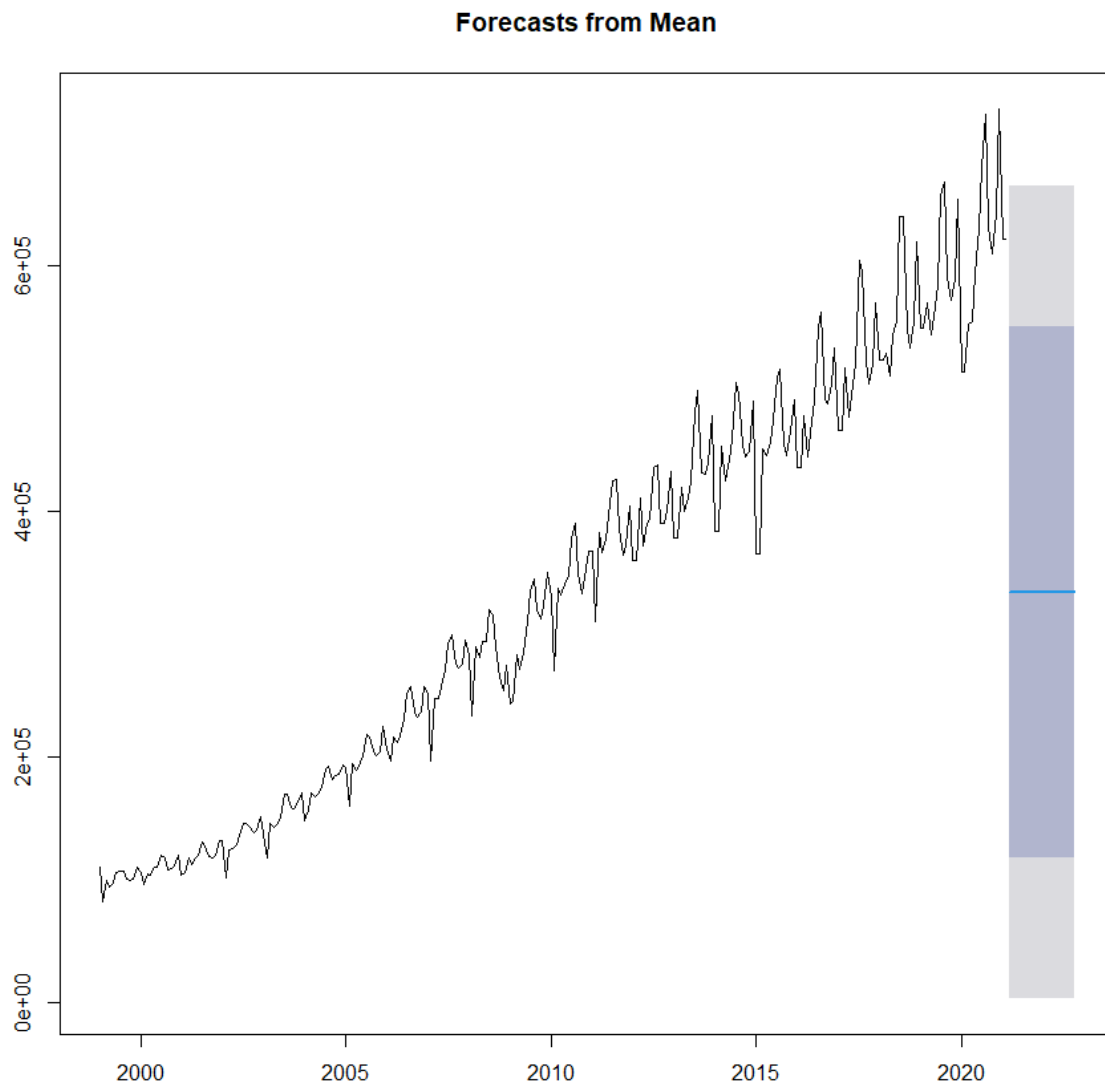
Wnioski:

-Po wykresie autokorelacji oraz histogramie reszt można wysnuć wnioski że model można uznać za realizacją szumu białego.

Za pomocą testu Shapiro-Wilka odrzuciłem hipotezę zerową na poziomie istotności 0.05 o rozkładzie normalnym reszt.

PROGNOZOWANIE Z WYKORZYSTANIEM METOD NAIWNYCH, DOBÓR NAJLEPSZEJ METODY DLA DANEGO SZEREGU.

Prognoza oparta na średniej:

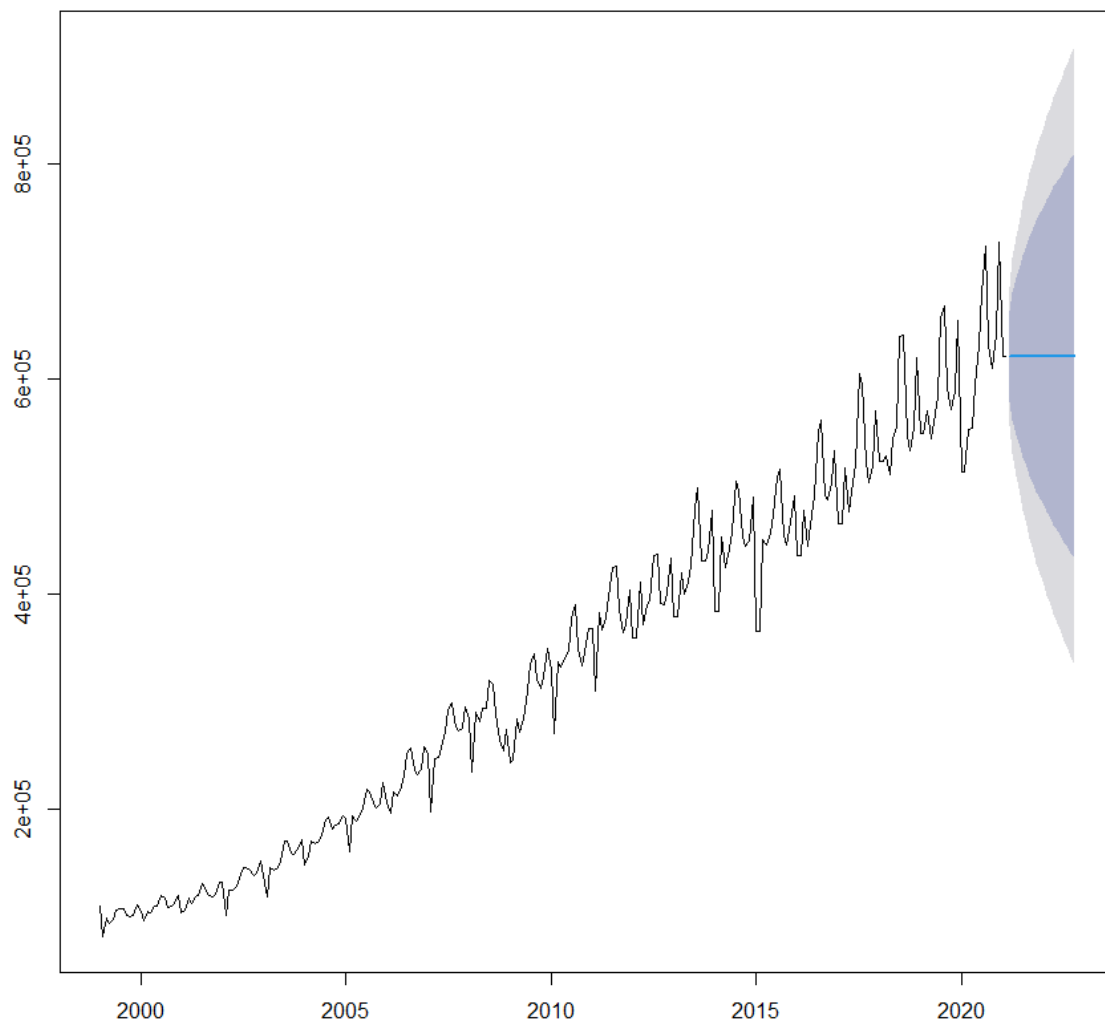


Wnioski:

- Prognozowanie za pomocą tej metody nie daje nam satysfakcjonujących wyników. Było to do przewidzenia, gdyż nasz szereg ma zarówno sezonowość jak i trend.

Naiwne prognozowanie:

Forecasts from Naive method

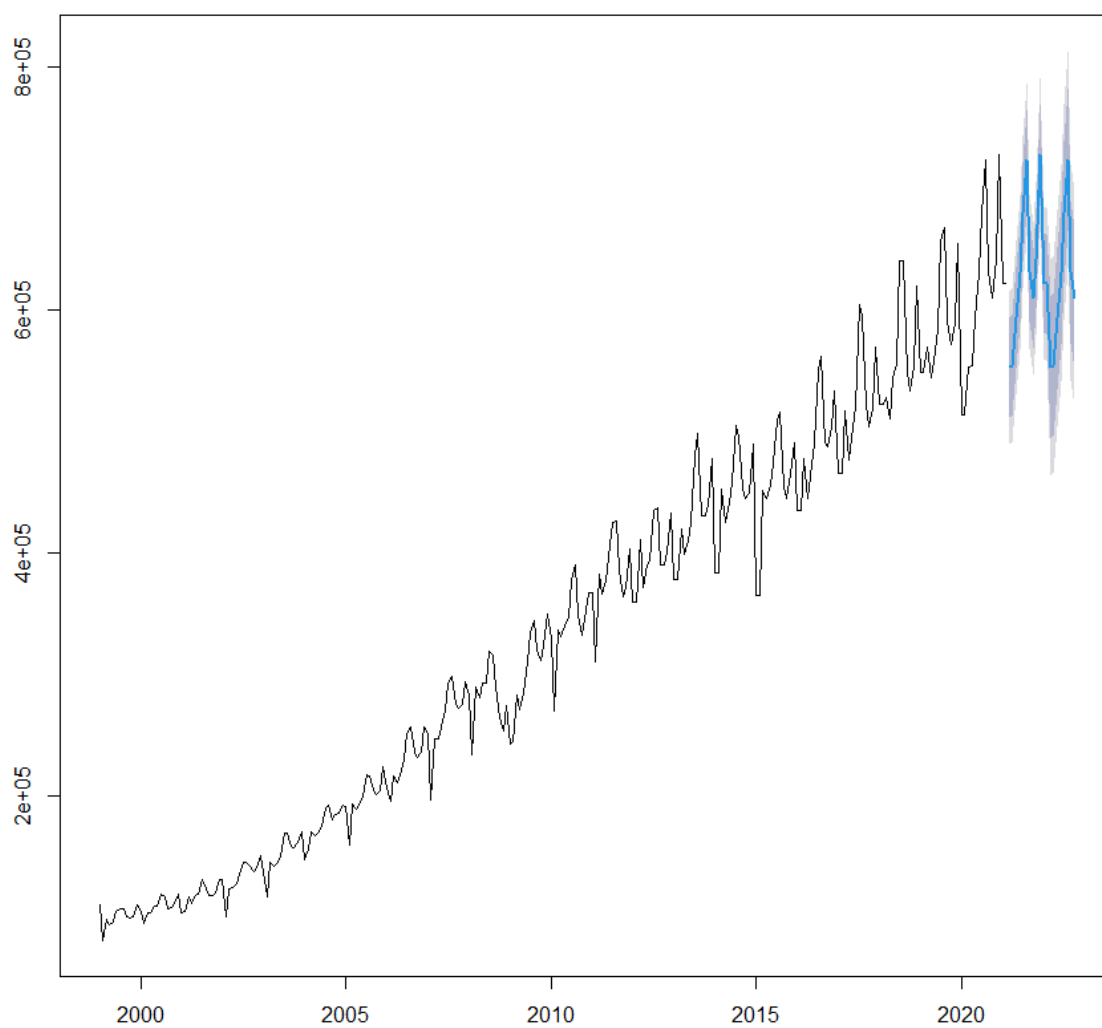


Wnioski:

- Wynik prognozowania jest lepszy niż poprzednio.

Sezonowe prognozowanie naiwne

Forecasts from Seasonal naive method

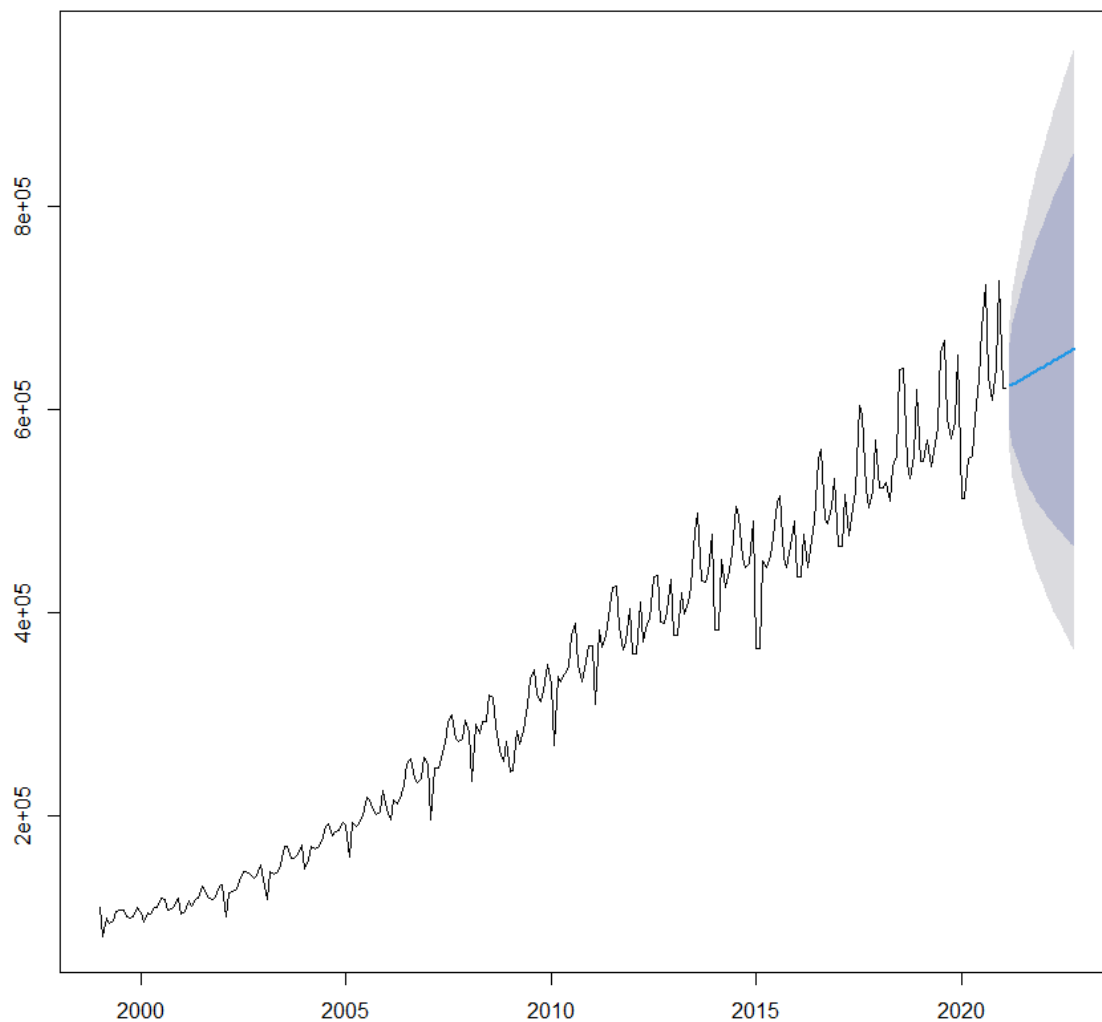


Wnioski:

- Jest to najbardziej satysfakcjonująca prognoza.
- Przedziały ufności rosną.

Prognozowanie z dryfem:

Forecasts from Random walk with drift



Wnioski:

- Prognoza jest lepsza niż w przypadku naiwnego prognozowania, jednak gorsza niż sezonowe prognozowanie naiwne.

Sprawdziłem jeszcze oceny dokładności za pomocą funkcji `accuracy`, żeby przekonać się czy moje przewidywania są prawdziwe. Okazało się, że nie ponieważ, kryteria wskazują na to, że prognozowanie z dryfem jest prognozowaniem z najlepszym dopasowaniem.

WPROWADZENIE DO ANALIZOWANYCH DANYCH

Dane z szeregu z trendem dotyczą liczby szczepień na COVID-19 w państwach danego roku. Analizowane będą dane wyłącznie dla Polski od marca danego roku. Zostały one pobrane ze strony:

<https://www.kaggle.com/gpreda/covid-world-vaccination-progress>

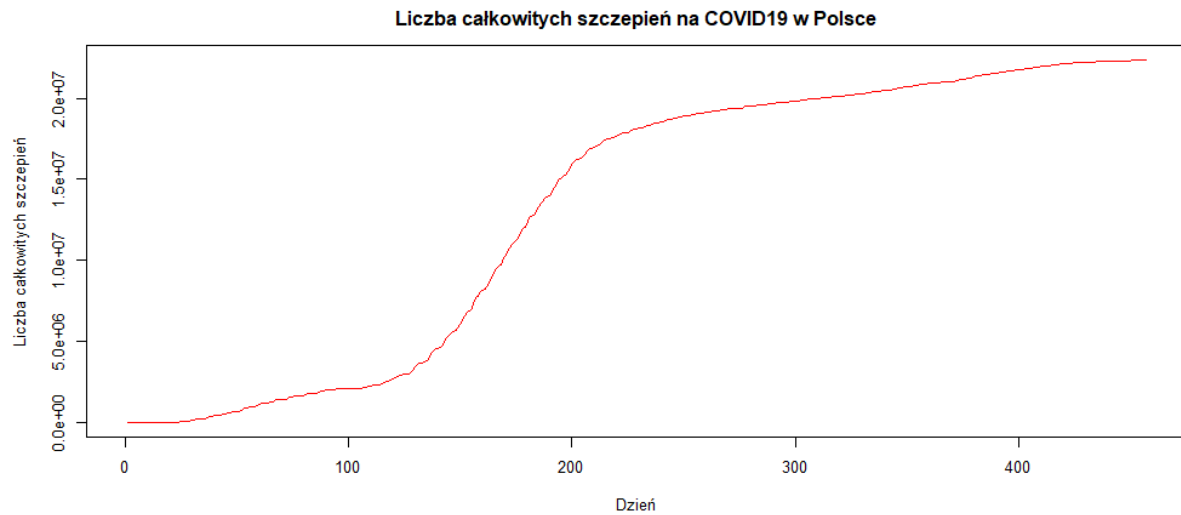
UTWORZENIE SZEREGÓW CZASOWYCH Z ŚREDNIMI: DZIENNYMI, MIESIĘCZNYMI, KWARTAŁOWYMI (PO ODPOWIEDNIM „WYCIĘCIU” POTRZEBNYCH DANYCH, INTERPOLACJI WARTOŚCI BRAKUJĄCYCH)

wybór kraju dla analizy danych (np. "Poland", "Palestine")

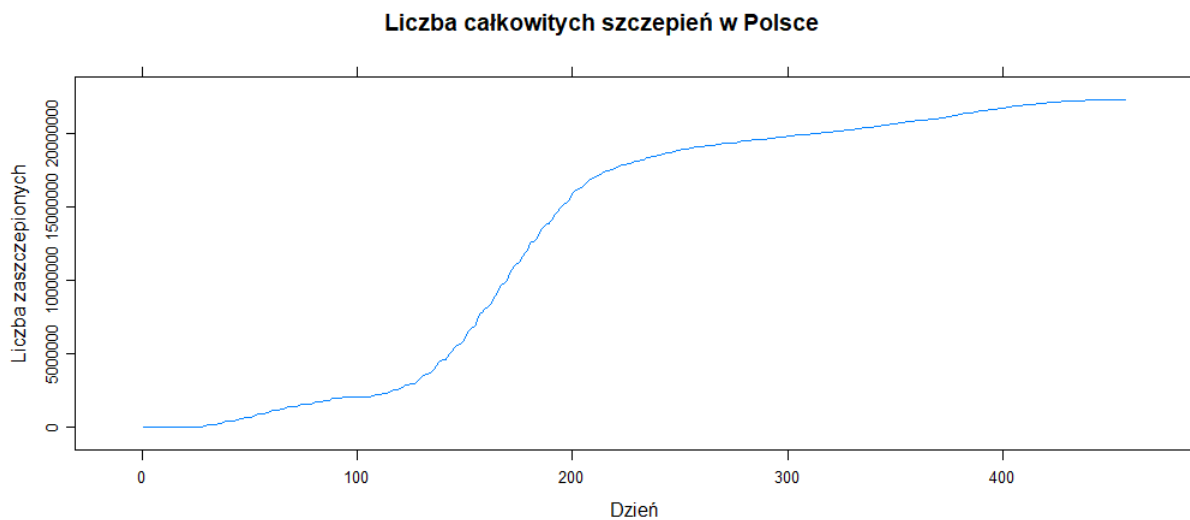
```
> sort(unique(my_data$country))
[1] "Afghanistan"
[5] "Angola"
[9] "Armenia"
[13] "Azerbaijan"
[17] "Barbados"
[21] "Benin"
[25] "Bonaire Sint Eustatius and Saba"
[29] "British Virgin Islands"
[33] "Cambodia"
[37] "Cayman Islands"
[41] "Colombia"
[45] "Costa Rica"
[49] "Curacao"
[53] "Denmark"
[57] "Ecuador"
[61] "Equatorial Guinea"
"Albania"
"Anguilla"
"Aruba"
"Bahamas"
"Belarus"
"Bermuda"
"Bosnia and Herzegovina"
"Brunei"
"Cameroon"
"Central African Republic"
"Comoros"
"Cote d'Ivoire"
"Cyprus"
"Djibouti"
"Egypt"
"Estonia"
"Algeria"
"Antigua and Barbuda"
"Australia"
"Bahrain"
"Belgium"
"Bhutan"
"Botswana"
"Bulgaria"
"Canada"
"Chile"
"Congo"
"Croatia"
"Democratic Republic of Congo"
"Dominica"
"El Salvador"
"Eswatini"
"Ethiopia"
"England"
"France"
"Germany"
"Guatemala"
"Hong Kong"
"India"
"Indonesia"
"Israel"
"Italy"
"Japan"
"Kazakhstan"
"Kenya"
"Kuwait"
"Kyrgyzstan"
"Laos"
"Latvia"
"Lebanon"
"Lesotho"
"Lithuania"
"Madagascar"
"Malawi"
"Malaysia"
"Maldives"
"Marshall Islands"
"Mexico"
"Monaco"
"Morocco"
"Namibia"
"Nepal"
"Nicaragua"
"Niger"
"Nigeria"
"North Macedonia"
"Oman"
"Pakistan"
"Panama"
"Paraguay"
"Peru"
"Philippines"
"Poland"
"Portugal"
"Romania"
"Russia"
"San Marino"
"Senegal"
"Serbia"
"Sierra Leone"
"Singapore"
"Slovakia"
"South Africa"
"South Korea"
"Spain"
"Sri Lanka"
"Sweden"
"Switzerland"
"Taiwan"
"Thailand"
"Timor-Leste"
"Trinidad and Tobago"
"Turkey"
"Ukraine"
"United Kingdom"
"United States"
"Uruguay"
"Uzbekistan"
"Venezuela"
"Vietnam"
"Yemen"
"Zambia"
"Zimbabwe"
```



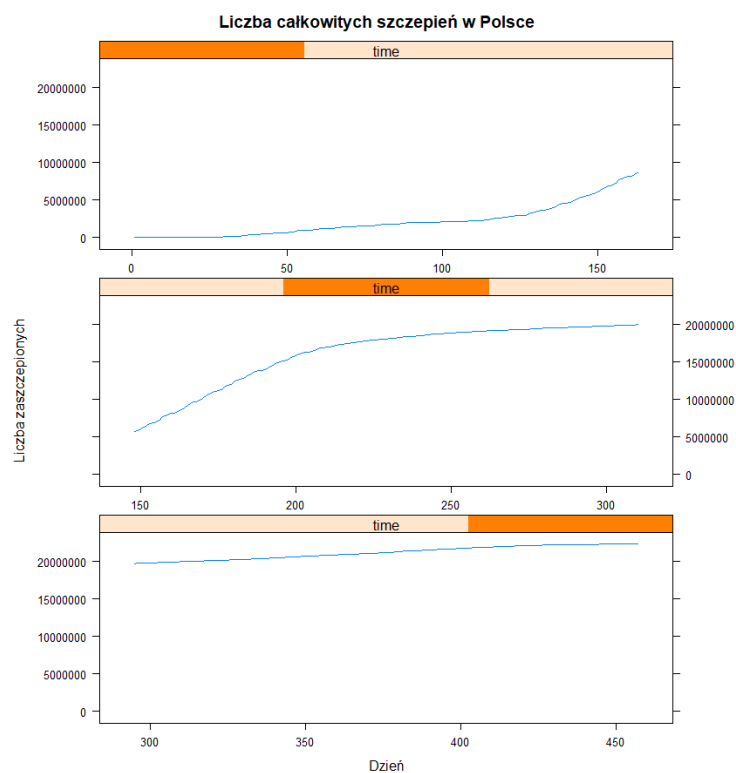
REPREZENTACJA SZEREGU CZASOWEGO NA RÓŻNEGO RODZAJU WYKRESACH



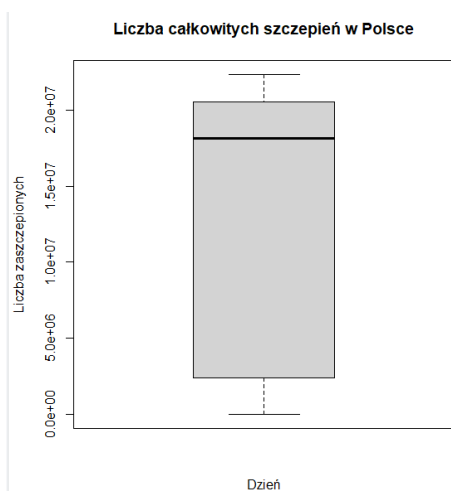
Wykres z użyciem xypplot()



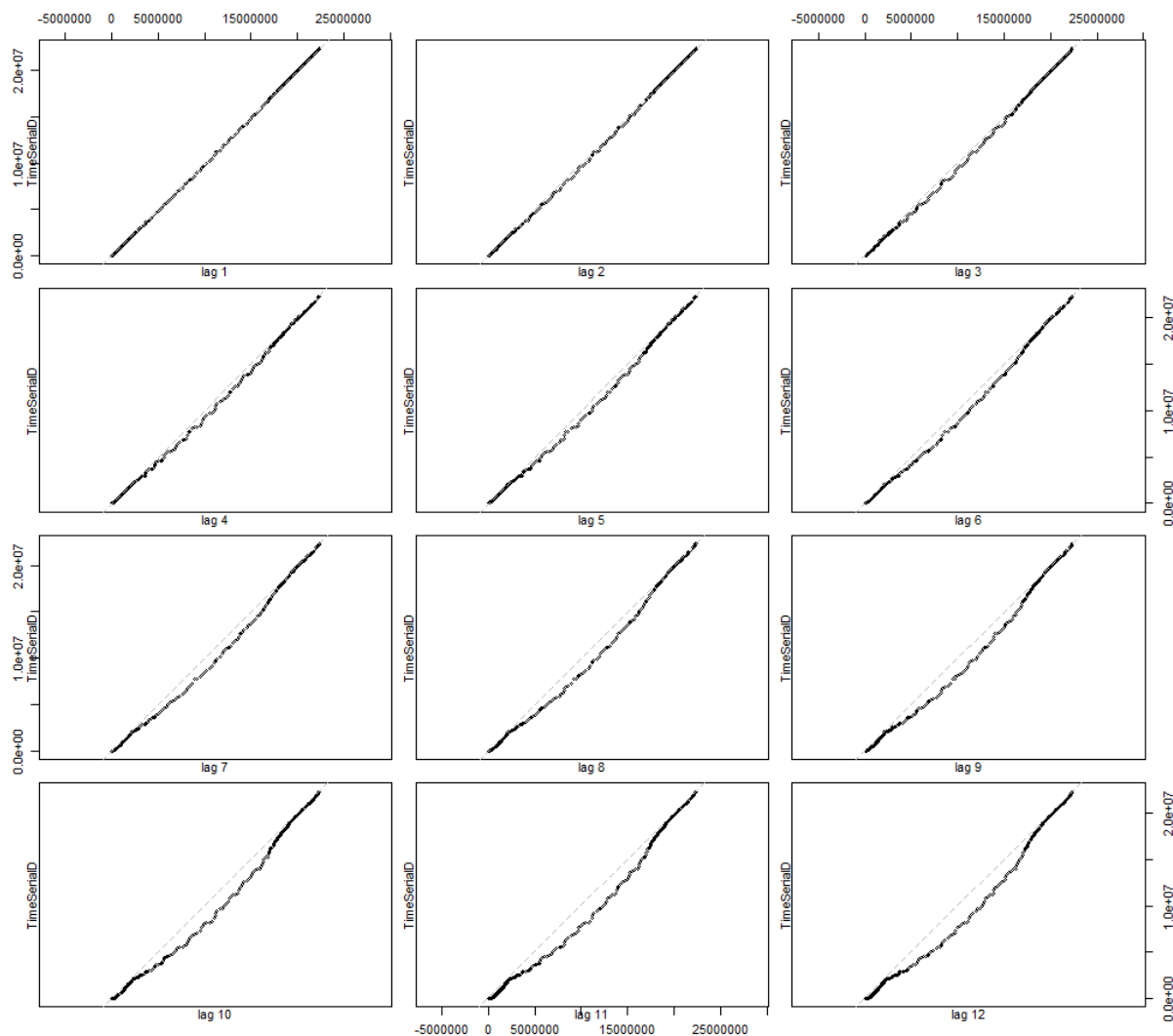
Wykres panelowy



Wykres pudełkowy

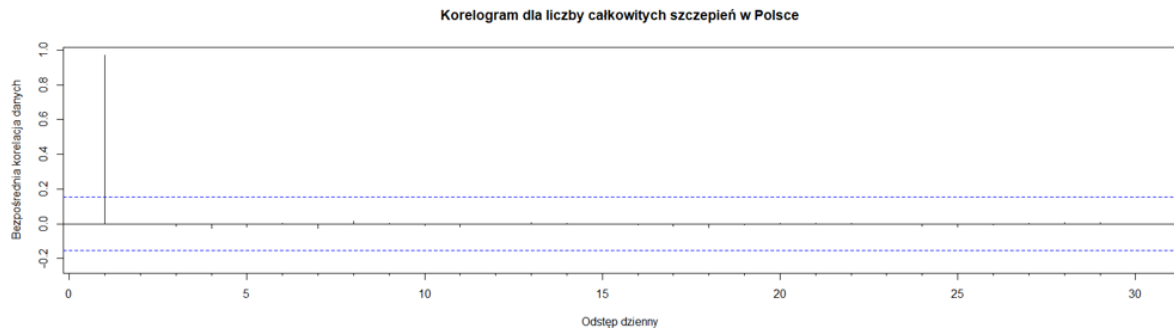


Wykres rozrzutu wartości opóźnionych



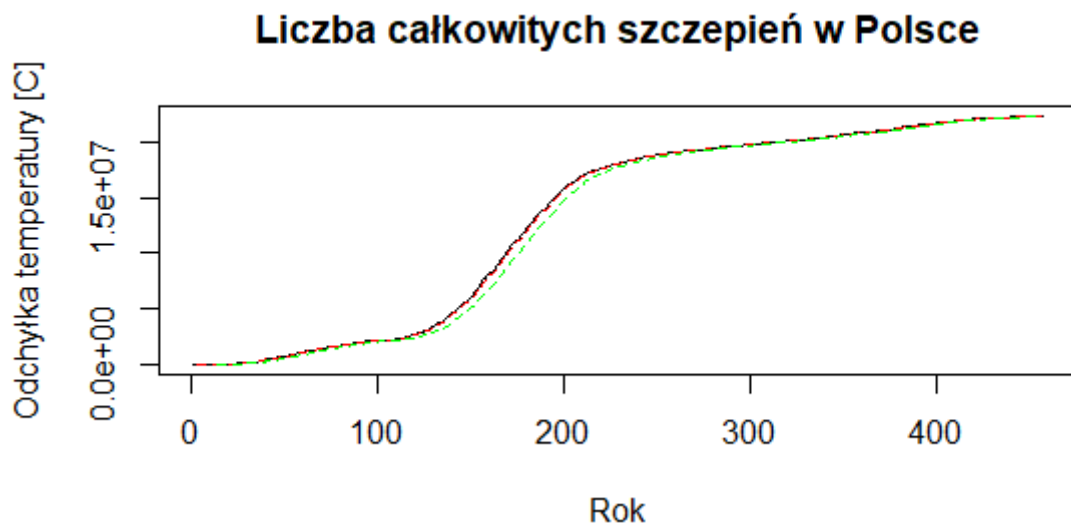
Dla danego szeregu widzimy największą korelację (zależność) danych dla opóźnienia z krokiem pierwszym, co pokazuje wyraźny trend wzrostowy. Z każdym kolejnym krokiem korelacja danych maleje, co oznacza gorzej dobrany współczynnik opóźnienia.

Wykresy funkcji autokorelacji – Acf, cząstkowej autokorelacji – Pacf



Widoczna jest wyraźny trend wzrostowy z opóźnieniem – jeden poprzez łagodne, regularne zmiany na wykresie funkcji Acf, oraz dzięki dużej „szpilce” przy odstępie sezonowym. Brak trendów z innymi opóźnieniami.

Wygładzenie danych poprzez użycie średniej ruchomej

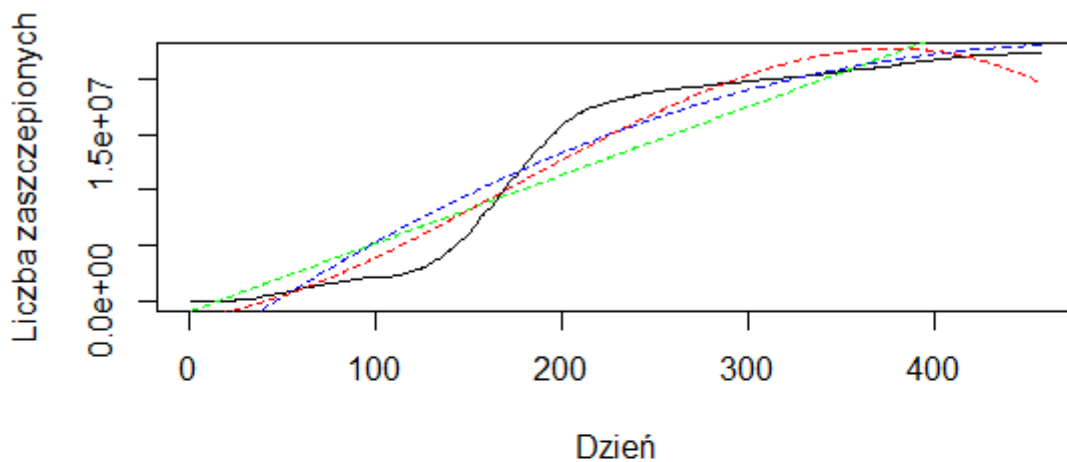


Wykres pokazuje wygładzenie szeregu dziennego dla odstępu: trzy/czternasto-dniowego, oznaczonych kolejno według wielkości spłaszczenia kolorami czerwonym i zielonym.

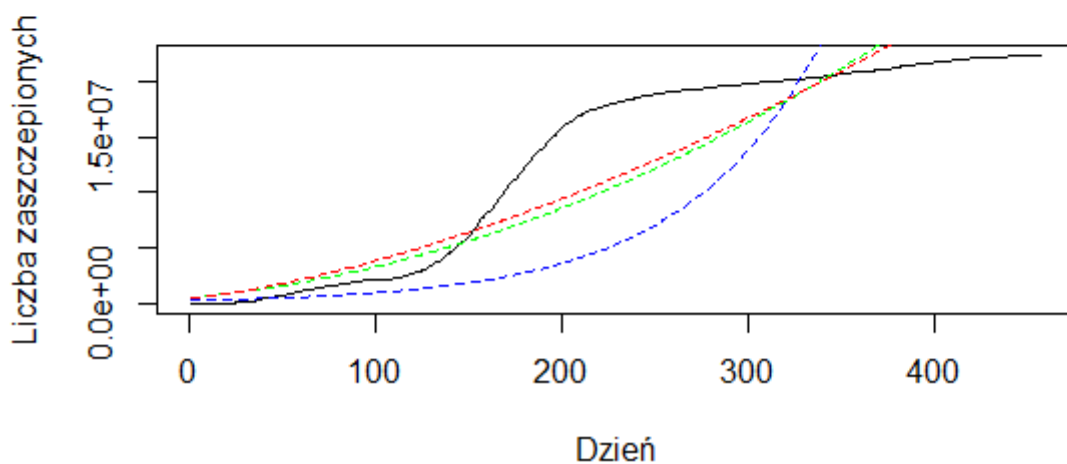


DEKOMPOZYCJE NA PODSTAWIE MODELU REGRESJI: TREND
LINIOWY/WIELOMIANOWY, TRANSFORMACJE BOXA-COXA

Liczba całkowitych szczepień w Polsce

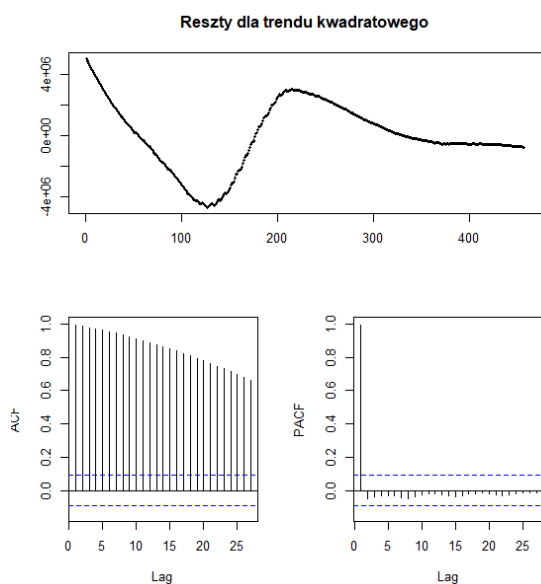
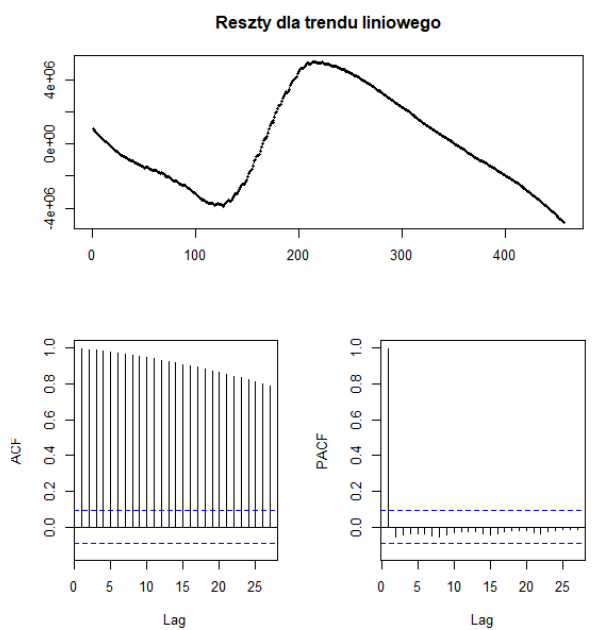


Liczba całkowitych szczepień w Polsce



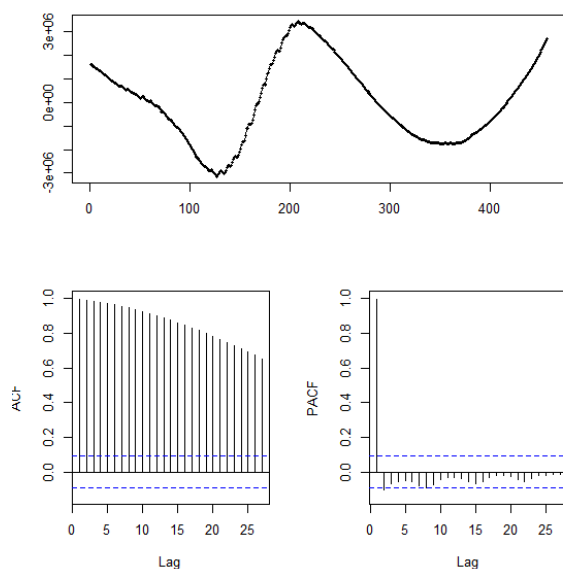
Wykres pierwszy pokazuje wykryte trendy kolejnych stopni: liniowy, kwadratowy, oraz stopnia trzeciego, oznaczonych kolejno kolorami: zielonym, czerwonym, niebieskim. Drugi z wykresów przedstawia transformacje Boxa-Coxa: pierwiastkową, logarytmiczną i z automatycznie dobranym współczynnikiem lambda, oznaczonych kolejno kolorami: zielonym, czerwonym, niebieskim. Na ich podstawie możemy stwierdzić, że koniecznym założeniem jest tutaj trend wielomianowy, najprawdopodobniej stopnia trzeciego.

Porównamy reszty z różnych metod dekompozycji na podstawie modelu regresji:



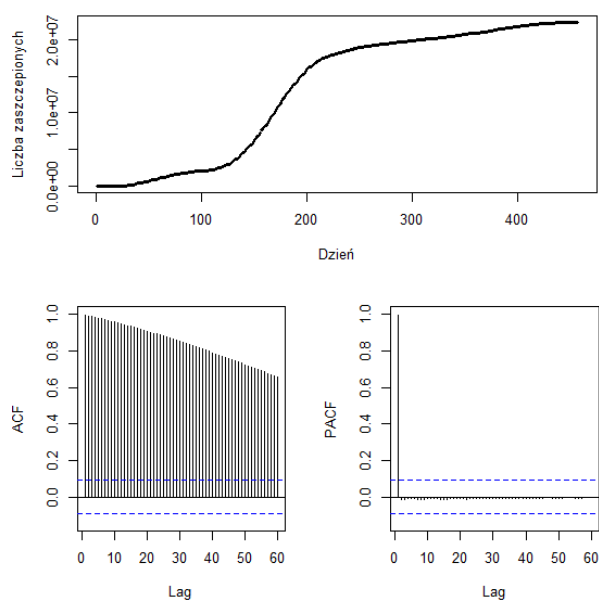


Reszty dla trendu stopnia trzeciego

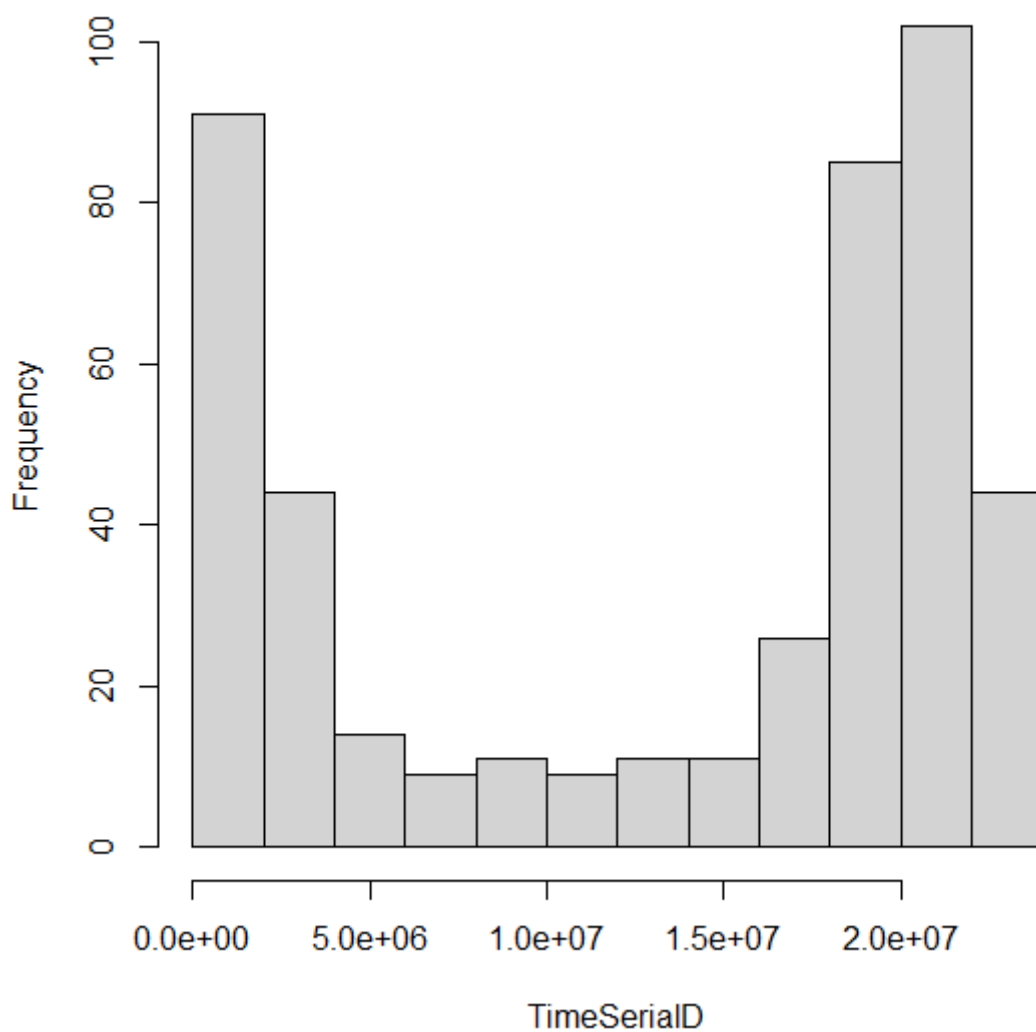


UCZYNIENIE SZEREGU STACJONARNYM – ELIMINACJA TRENDU,
SEZONOWOŚCI POPRZECZ RÓŻNICOWANIE I STABILIZACJA WARIANCJI
POPURZECZ TRANSFORMACJĘ BOXA-COXA

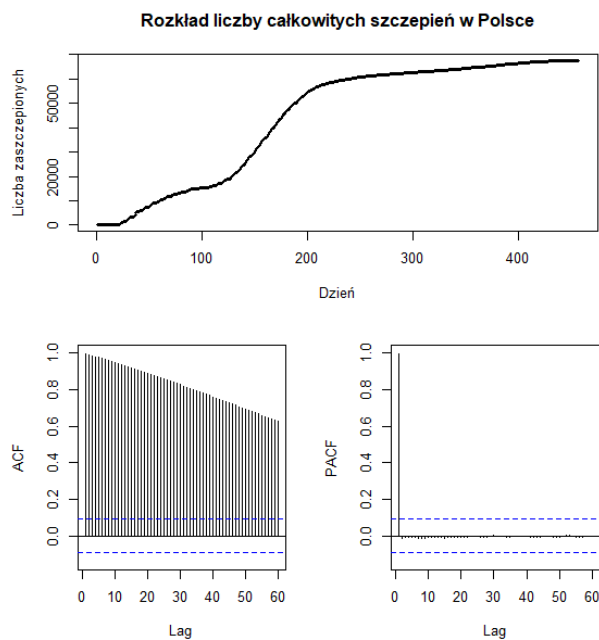
Rozkład liczby całkowitych szczepień w Polsce



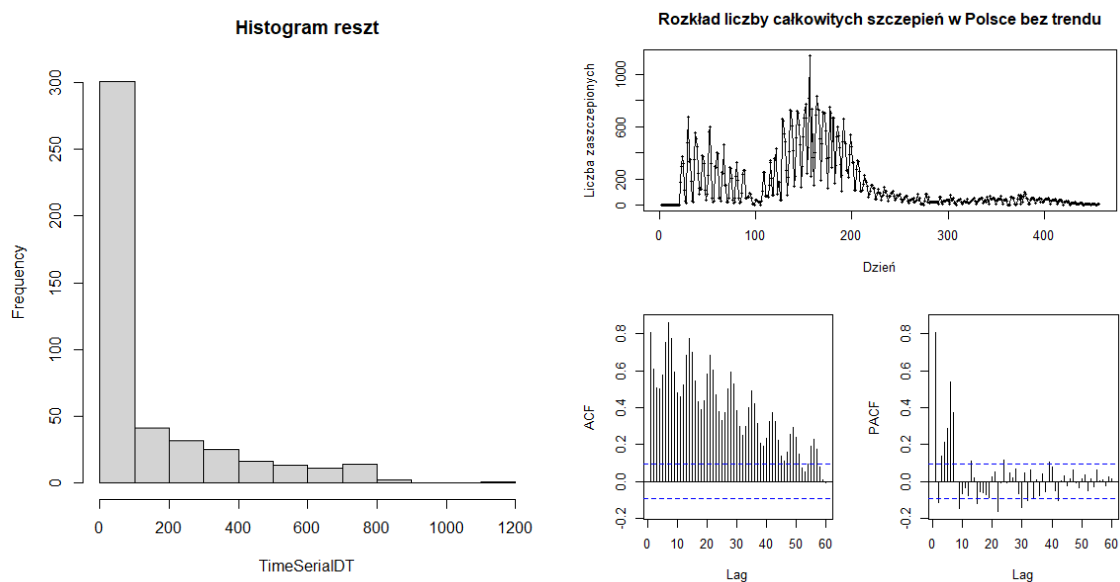
Histogram reszt



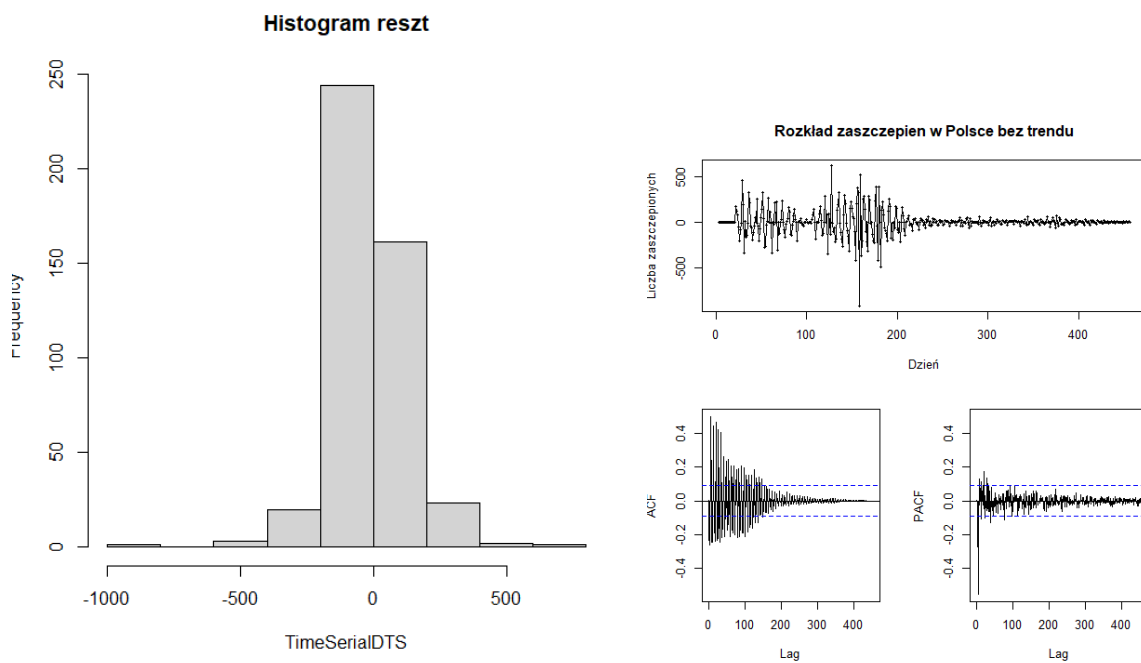
Po zastosowaniu transformacji Boxa-Coxa z automatycznie dobranym współczynnikiem lambda szereg i jego histogram, oraz wykresu autokorelacji wyglądają następująco:



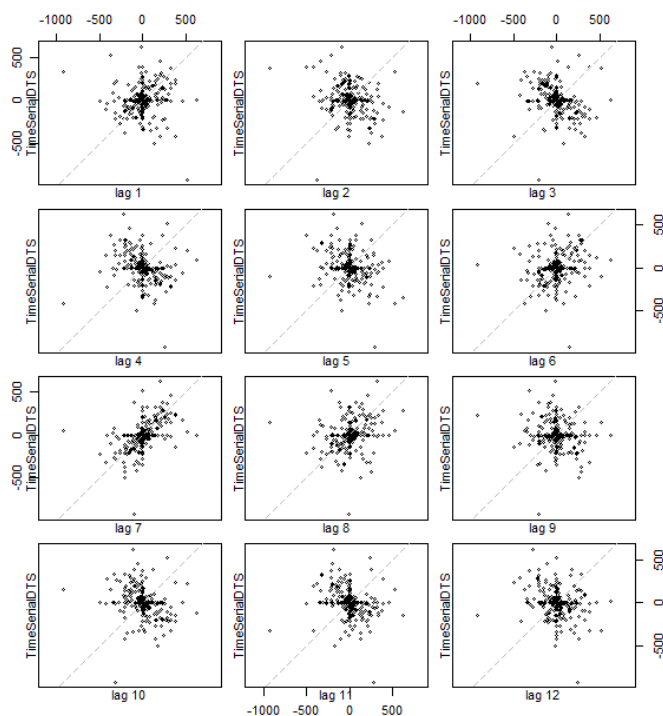
Po pierwszym zróżnicowaniu danych z opóźnieniem pierwszym wygląda to, jak na następnej stronie:



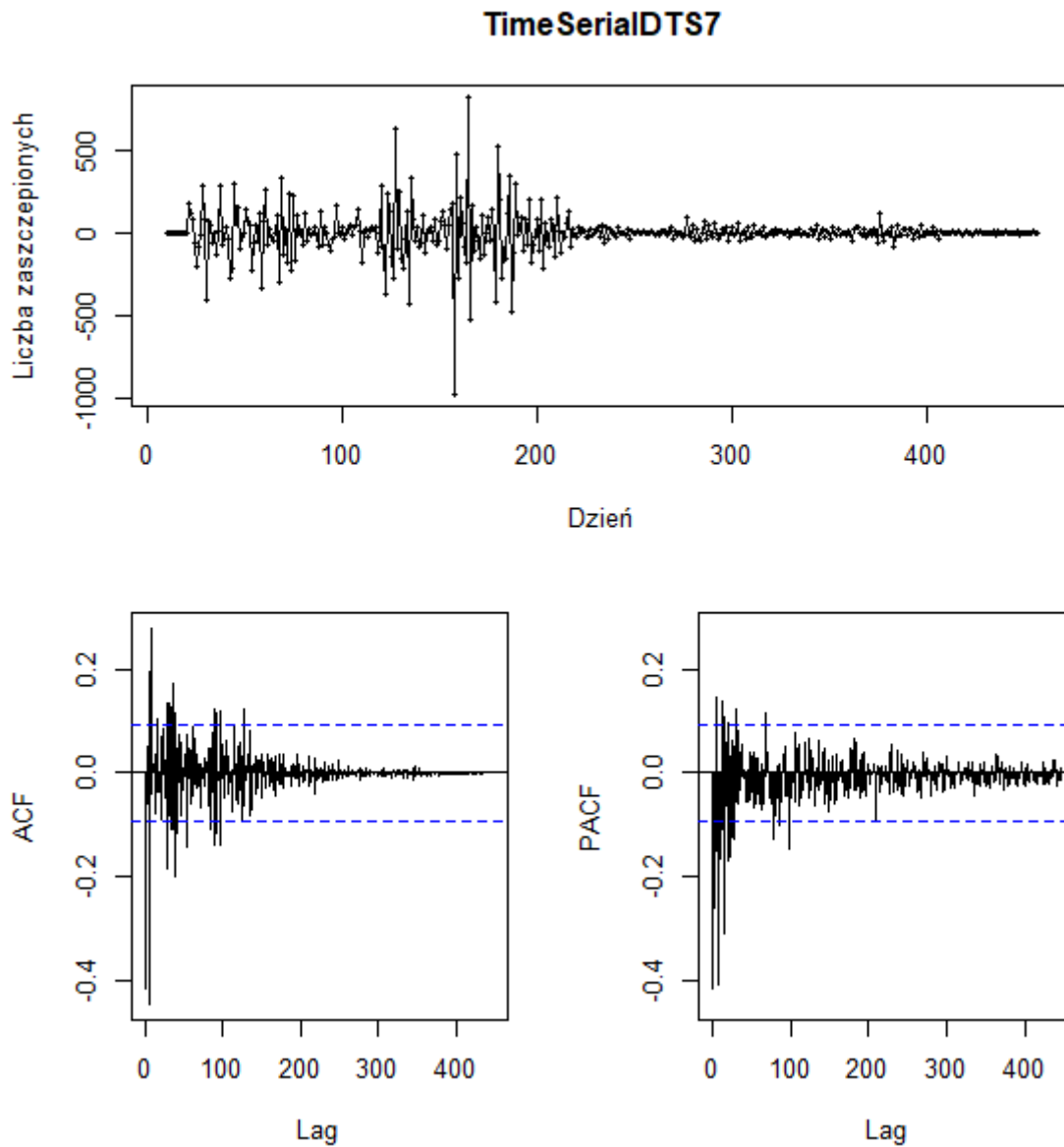
Po drugim zróżnicowaniu danych z opóźnieniem pierwszym wygląda to następująco:



Widoczna jednak dalej jest pewna sezonowość reszt. W celu doboru rzędu różnicowania sprawdzimy wykresy rozrzutu.



Po zróżnicowaniu z opóźnieniem – siedem wygląda to następująco:

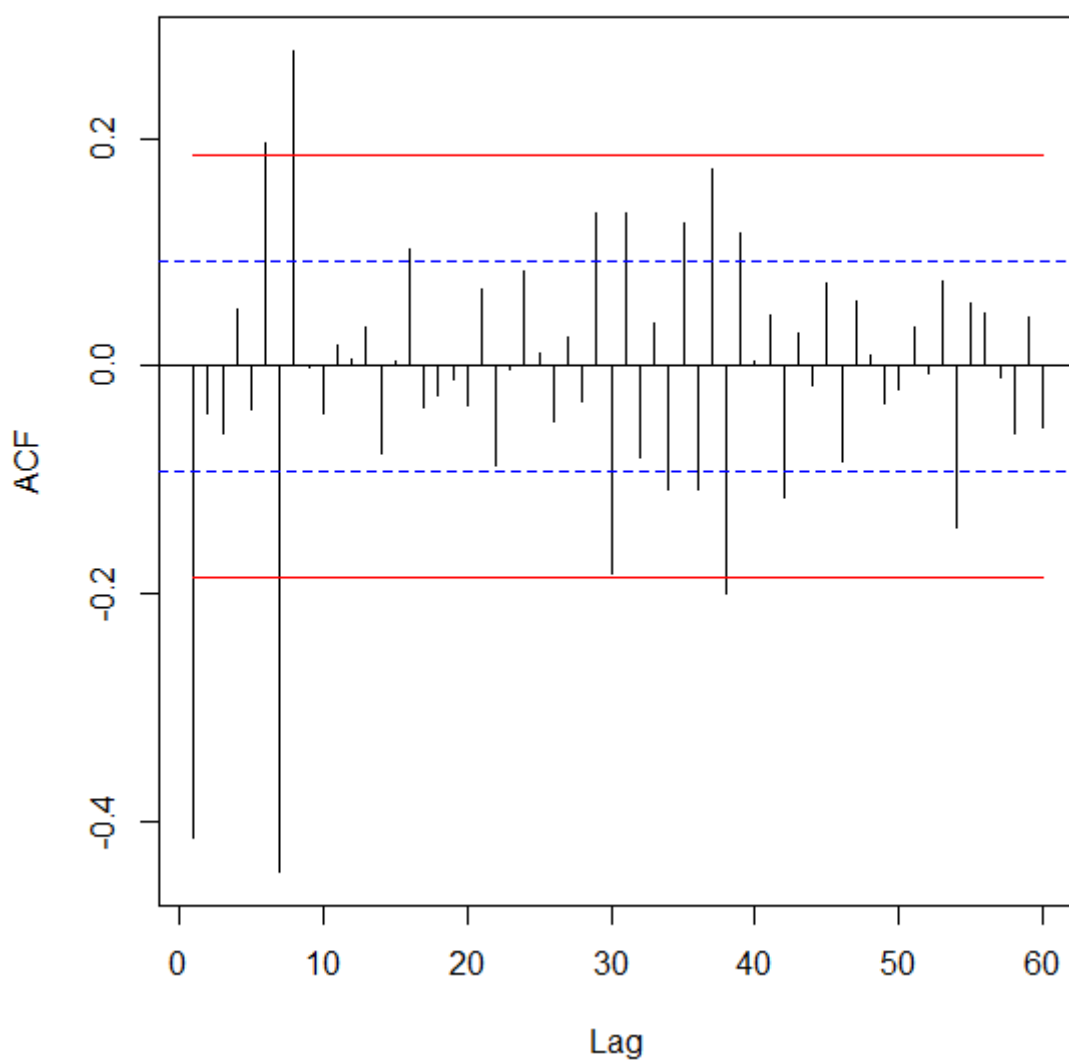


Po przyglądnięciu się wykresom funkcji autokorelacji bezpośredniej i cząstkowej możemy założyć pewne modele z konkretnymi współczynnikami w celu identyfikacji szeregu. Wartości wykraczające poza przedział ufności na wykresie ACF to: 1, 7, 8, 21, 23, 30, 31; dla wykresu Pacf to: 1, 3, 7, minimalnie 14, 21, 22. Poprawnym byłoby założenie więc modeli np.: MA(1), MA(31), AR(1), AR(7).



SPRAWDZENIE STACJONARNOŚCI OTRZYMANEGO SZEREGU

Series TimeSerialDTS7



Brak widocznych wyraźnych odchyłek od przedziału ufności, możemy więc przyjąć dany szereg za stacjonarny.

WYZNACZENIE WSPÓŁCZYNNIKÓW MODELU AUTOREGRESJI I PORÓWNANIE DOPASOWANIA RÓŻNYMI METODAMI ESTYMACJI

```
Call:
ar(x = TimeSeriesDNM, aic = FALSE, order.max = 1, method = c("yule-walker"))

Coefficients:
      1
-0.3017

Order selected 1 sigma^2 estimated as 771.6
Call:
ar(x = TimeSeriesDNM, aic = FALSE, order.max = 1, method = c("burg"))

Coefficients:
      1
-0.3019

Order selected 1 sigma^2 estimated as 761.5
Call:
ar(x = TimeSeriesDNM, aic = FALSE, order.max = 1, method = c("ols"))

Coefficients:
      1
-0.3021

Intercept: 0.02444 (2.238)

Order selected 1 sigma^2 estimated as 766.5
Call:
ar(x = TimeSeriesDNM, aic = FALSE, order.max = 1, method = c("mle"))

Coefficients:
      1
-0.3002

Order selected 1 sigma^2 estimated as 761.5
Call:
ar(x = TimeSeriesDNM, aic = FALSE, order.max = 1, method = c("yw"))

Coefficients:
      1
-0.3017

Order selected 1 sigma^2 estimated as 771.6
```

Automatyczny dobór rzędu dzięki kryterium „aic”:

```
Call:
ar(x = TimeSeriesDNM, aic = TRUE, order.max = 100, method = c("yule-walker"))

Coefficients:
      1      2      3      4      5      6      7
-0.4339 -0.3527 -0.3874 -0.2801 -0.2079 -0.1070 -0.2937

Order selected 7 sigma^2 estimated as 646.1
```

WYZNACZENIE WSPÓŁCZYNNIKÓW MODELU RUCHOMEJ ŚREDNIEJ Z UŻYCIEM FUNKCJI ARIMA()

ARIMA(0,0,1) with non-zero mean

```
Coefficients:
      ma1      mean
-0.6073  0.0219
s.e.    0.1164  0.8493
```

```
sigma^2 estimated as 715: log likelihood=-723.81
AIC=1453.62 AICC=1453.78 BIC=1462.73
```

Training set error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|-------------|----------|----------|----------|----------|-----------|-----------|
| Training set | 0.004000951 | 26.56592 | 17.68786 | 24.05806 | 311.4821 | 0.5644599 | 0.1493788 |



ARIMA(0,0,7) with non-zero mean

Coefficients:

| | ma1 | ma2 | ma3 | ma4 | ma5 | ma6 | ma7 | mean |
|------|---------|---------|---------|---------|--------|--------|---------|--------|
| | -0.2768 | -0.3701 | -0.3430 | -0.0049 | 0.1319 | 0.4055 | -0.5425 | 0.1010 |
| s.e. | 0.0744 | 0.0808 | 0.0793 | 0.0722 | 0.0832 | 0.0857 | 0.0809 | 0.1115 |

sigma^2 estimated as 569.6: log likelihood=-706.35
AIC=1430.69 AICc=1431.94 BIC=1458.02

Training set error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|------------|----------|----------|---------|----------|-----------|------------|
| Training set | -0.2527391 | 23.23713 | 15.51878 | 137.842 | 370.8928 | 0.4952398 | -0.1249344 |

PORÓWNANIE MODELU RUCHOMEJ ŚREDNIEJ Z MODELEM AUTOREGRESJI, ANALIZA DOBROCI DOPASOWANIA

ARIMA(1,0,0) with non-zero mean

Coefficients:

| | ar1 | mean |
|------|---------|--------|
| | -0.3001 | 0.0190 |
| s.e. | 0.0766 | 1.7129 |

sigma^2 estimated as 771.5: log likelihood=-729.48
AIC=1464.97 AICc=1465.13 BIC=1474.08

Training set error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|-------------|----------|----------|----------|----------|-----------|-------------|
| Training set | 0.000248016 | 27.59571 | 18.05437 | 34.48005 | 240.9512 | 0.5761562 | -0.05071994 |

ARIMA(7,0,0) with non-zero mean

Coefficients:

| | ar1 | ar2 | ar3 | ar4 | ar5 | ar6 | ar7 | mean |
|------|---------|---------|---------|---------|---------|---------|---------|--------|
| | -0.4237 | -0.3549 | -0.3765 | -0.2759 | -0.2221 | -0.0877 | -0.3400 | 0.0341 |
| s.e. | 0.0760 | 0.0829 | 0.0854 | 0.0886 | 0.0857 | 0.0870 | 0.0811 | 0.6502 |

sigma^2 estimated as 631.1: log likelihood=-711.56
AIC=1441.12 AICc=1442.37 BIC=1468.46

Training set error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|-------------|----------|----------|----------|----------|-----------|--------------|
| Training set | 0.003049892 | 24.46013 | 15.93834 | 79.35744 | 203.1509 | 0.5086289 | -0.003157649 |



WYZNACZENIE OPTYMALNYCH MODELI Z WYKORZYSTANIEM FUNKCJI AUTO.ARIMA(), ORAZ ICH PORÓWNANIE NA PODSTAWIE KRYTERIÓW DOPASOWANIA, WYBÓR NAJLEPSZEGO

ARIMA(1,0,1) with zero mean

Coefficients:

| | ar1 | ma1 |
|------|--------|---------|
| | 0.4416 | -0.9686 |
| s.e. | 0.0813 | 0.0304 |

sigma^2 estimated as 629.4: log likelihood=-714.7
AIC=1435.41 AICC=1435.57 BIC=1444.52

Training set error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|---------|----------|----------|----------|----------|-----------|-------------|
| Training set | 1.21456 | 24.92532 | 16.50394 | 149.1697 | 213.2714 | 0.5266785 | -0.02032933 |

ARIMA(1,0,1) with zero mean

Coefficients:

| | ar1 | ma1 |
|------|--------|---------|
| | 0.4416 | -0.9686 |
| s.e. | 0.0813 | 0.0304 |

sigma^2 estimated as 629.4: log likelihood=-714.7
AIC=1435.41 AICC=1435.57 BIC=1444.52

Training set error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|---------|----------|----------|----------|----------|-----------|-------------|
| Training set | 1.21456 | 24.92532 | 16.50394 | 149.1697 | 213.2714 | 0.5266785 | -0.02032933 |

ARIMA(1,0,1) with zero mean

Coefficients:

| | ar1 | ma1 |
|------|--------|---------|
| | 0.4416 | -0.9686 |
| s.e. | 0.0813 | 0.0304 |

sigma^2 estimated as 629.4: log likelihood=-714.7
AIC=1435.41 AICC=1435.57 BIC=1444.52

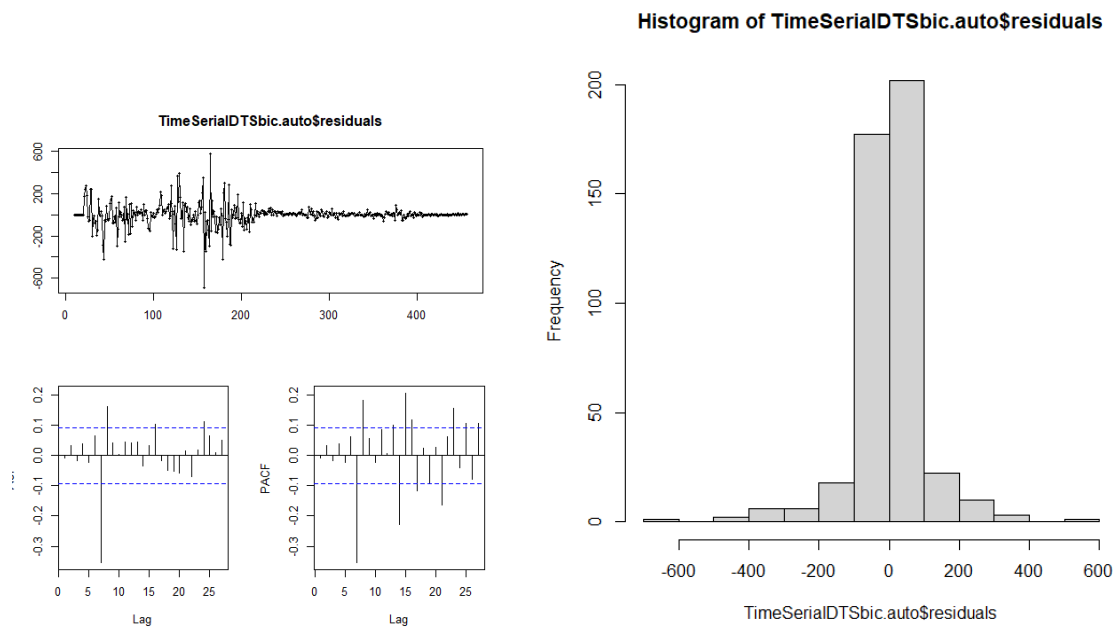
Training set error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|---------|----------|----------|----------|----------|-----------|-------------|
| Training set | 1.21456 | 24.92532 | 16.50394 | 149.1697 | 213.2714 | 0.5266785 | -0.02032933 |

Najlepszym modelem jest więc - ARIMA(1,0,1) wyznaczona identycznie (z identycznymi współczynnikami) przez każde z kryteriów dobroci dopasowania: aic, aicc, bic.



TEST STACJONARNOŚCI AUTOMATYCZNIE DOBRANEGO MODELU



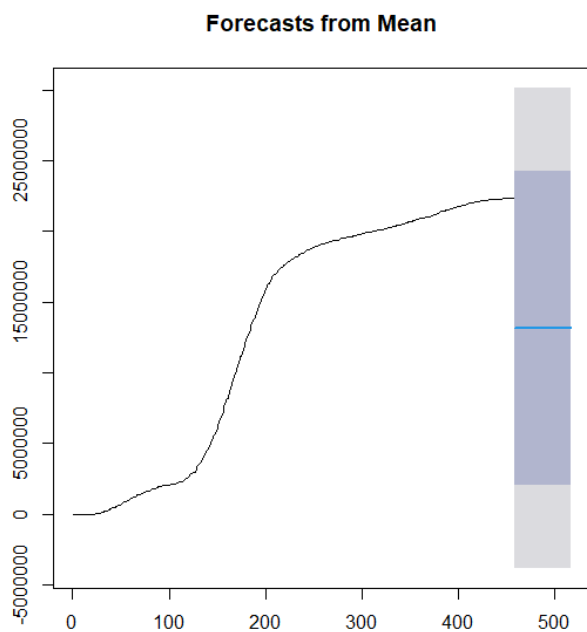
Jest to szereg stacjonarny, ponieważ żadna z wartości nie wystaje znacząco poza przedział ufności. Histogram jest bardzo zbliżony do rozkładu normalnego. Test Shapiro-Wilka potwierdza, że nim jest, ponieważ wartość zmiennej p nie przekracza: 0,05.

shapiro-wilk normality test

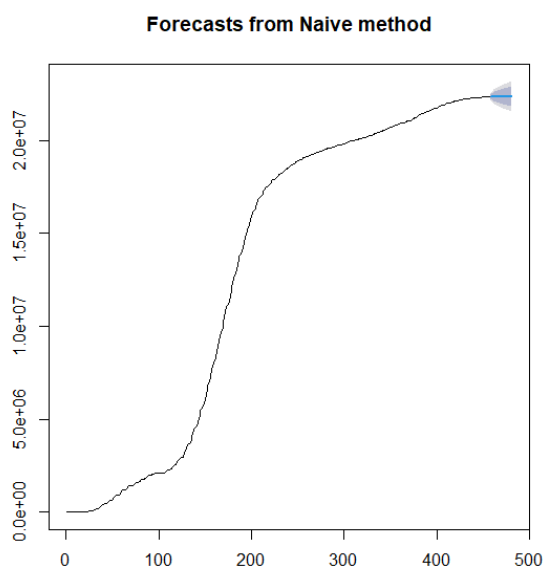
```
data: TimeSeriesMSaicc.auto$residuals  
W = 0.98575, p-value = 0.005308
```

PROGNOZOWANIE Z WYKORZYSTANIEM METOD NAIWNYCH

Prognoza oparta na średniej

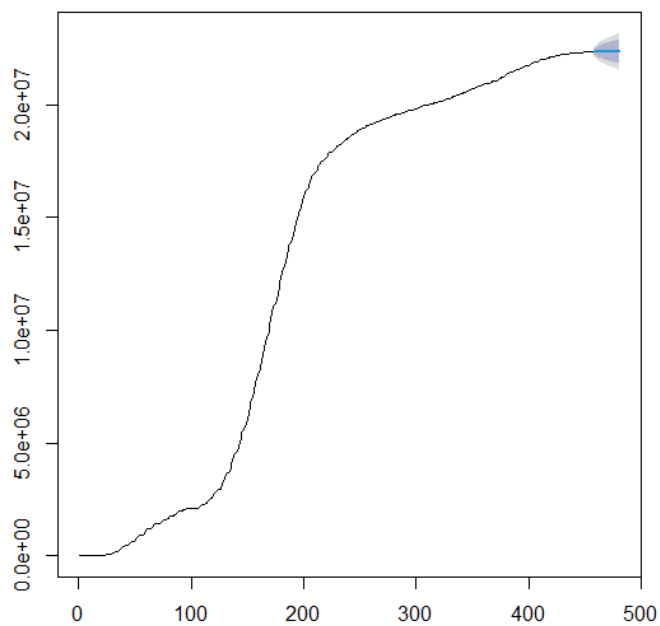


Prognozowanie naiwne z użyciem funkcji naive()



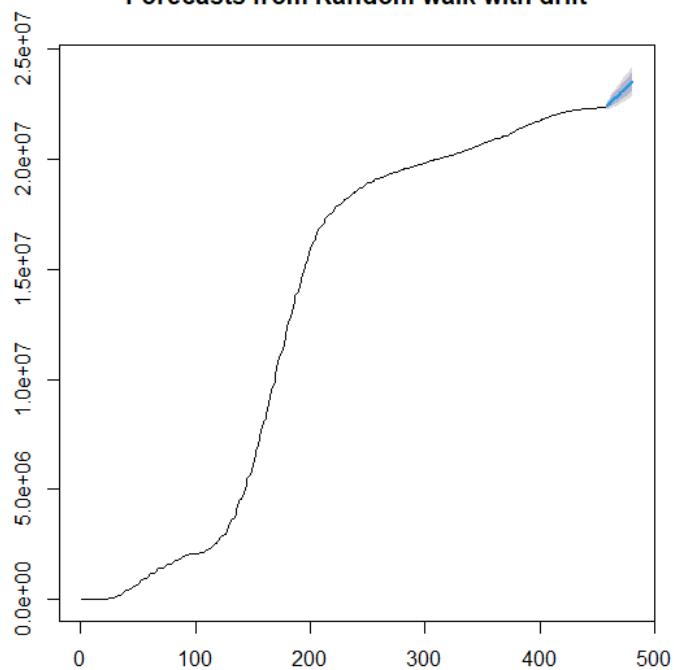
Prognozowanie z użyciem funkcji `snaive()`

Forecasts from Seasonal naive method



Prognozowanie naiwne z dryfem

Forecasts from Random walk with drift



Najlepszą metodą w celu prognozy liczby całkowitych szczepień jest zastosowanie tutaj prognozowania naiwnego z dryfem z powodu trendu wielomianowego w szeregu. Nadal nie będzie ona jednak wystarczająco poprawna.