

# Spis treści

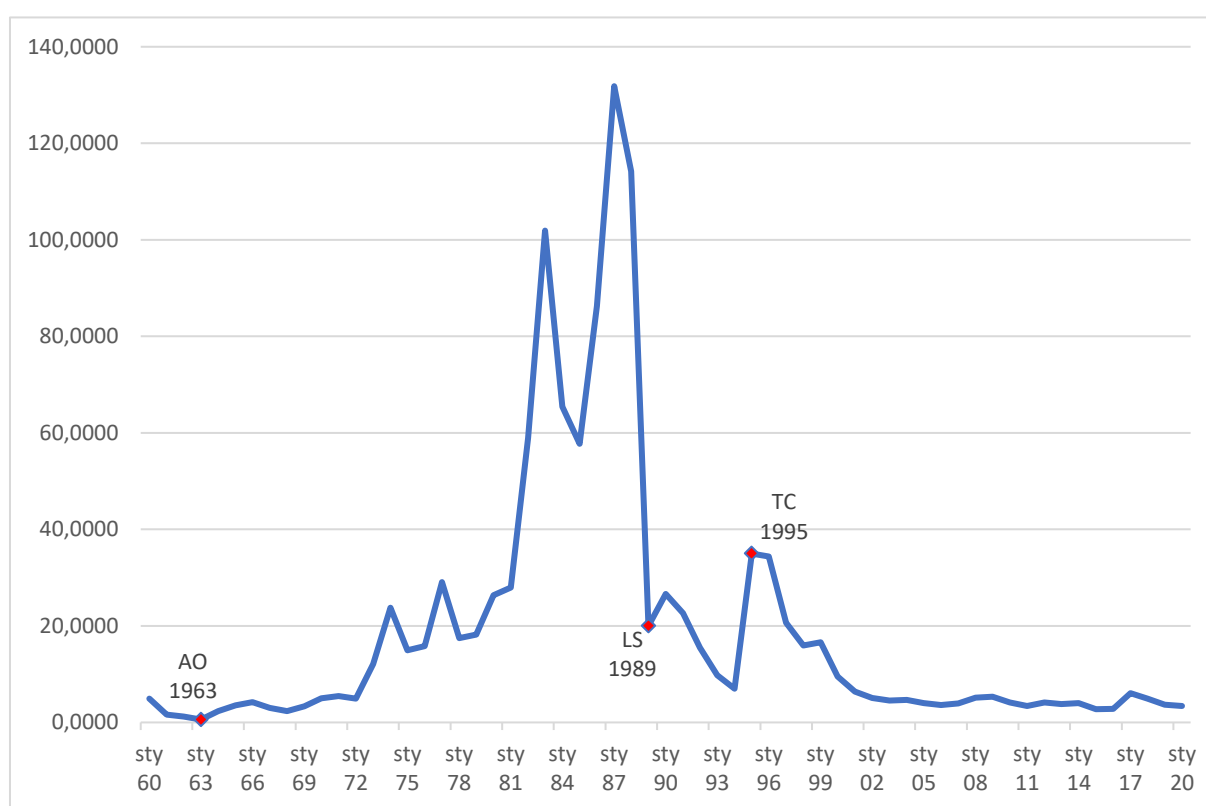
<b>I. SZEREG NIESEZONOWY</b>	2
1.1. Opis danych	2
1.2. Dekompozycja szeregu	3
1.3. Modele ekstrapolacyjne	3
1.3.1. Wygładzanie wykładnicze	3
1.3.2. Model Holta	4
1.4. Model ARIMA	6
1.4.1. Stacjonarność szeregu	6
1.4.2. Identyfikacja i weryfikacja	7
1.4.3. Prognoza i porównanie	9
<b>II. SZEREG SEZONOWY</b>	11
2.1. Opis danych	11
2.2. Dekompozycja szeregu	12
2.3. Model ekstrapolacyjny	12
2.4. Model SARIMA	14
2.4.1. Sezonowość i stacjonarność szeregu	14
2.4.2. Identyfikacja i weryfikacja	16
2.4.3. Prognoza i porównanie	17
<b>PODSUMOWANIE</b>	19

## I. SZEREG NIESEZONOWY

### 1.1. Opis danych

Szereg nie sezonowy, który zostanie podany analizie dotyczy wskaźnika cen towarów i usług konsumpcyjnych (CPI) dla Meksyku. Dane pochodzą z bazy danych "The World Bank Data"<sup>1</sup> i mają charakter roczny (dostęp online na dzień 02.04.2021). Liczba obserwacji wynosi 61, a pierwsza z nich odnosi się do 1960 roku natomiast ostatnia do 2020 roku.

Rysunek 1. CPI dla Meksyku w latach 1960-2020.



Źródło: Opracowanie własne na podstawie danych pochodzących z The World Bank Data

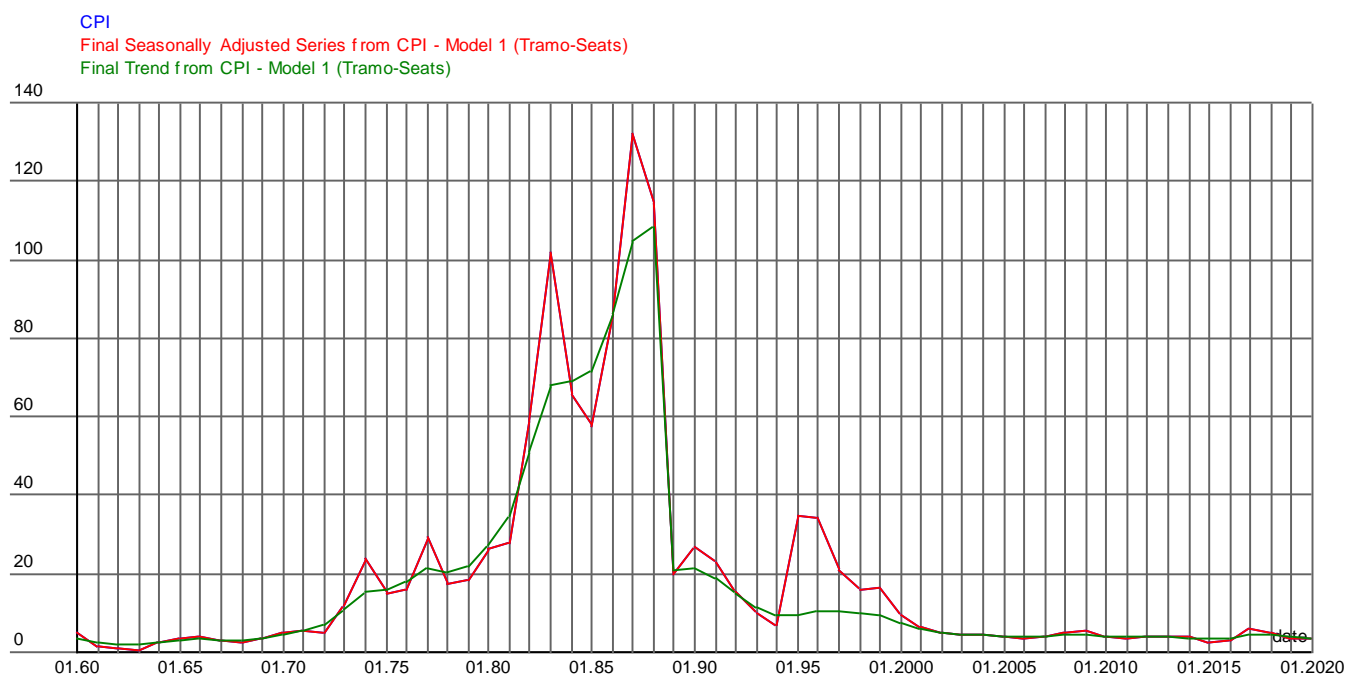
Rysunek 1 przedstawia wykres dotyczący kształtowania się CPI Meksyku. Czerwone punkty przedstawiają wartości odstające, które zostały wyróżnione przez program Demetra. Niestety w przypadku pierwszej z nich, nie udało się ustalić co mogło być przyczyną dla różnic w obserwacjach. Jeśli chodzi o drugą podejrzaną obserwację, w roku 1989 uchwalono plan Bakera, którego celem było rozwiązanie problemów gospodarczych Meksyku i Ameryki Łacińskiej. Natomiast obserwacja z 1995 roku jest odwzorowaniem kryzysu gospodarczego.

<sup>1</sup> <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=MX>

## 1.2. Dekompozycja szeregu

Wstępnej analizy szeregu dokonano w programie Demetra, natomiast samą dekompozycję szeregu przeprowadzono przy użyciu metody TRAMO/Seats.

Rysunek 2. Dekompozycja szeregu niesezonowego.



W powyższym szeregu, z pewnością nie ma sezonowości. Dekompozycja szeregu nie doprowadziła do zauważalnego wyodrębnienia jego składowych, a program potwierdził brak wahań sezonowych. Model przeszedł pomyślnie testy diagnostyczne. Program zasugerował, aby do interpretacji szeregu posłużyć się przekształceniem logarytmicznym, co też postanowiono wykorzystać w poniższym badaniu.

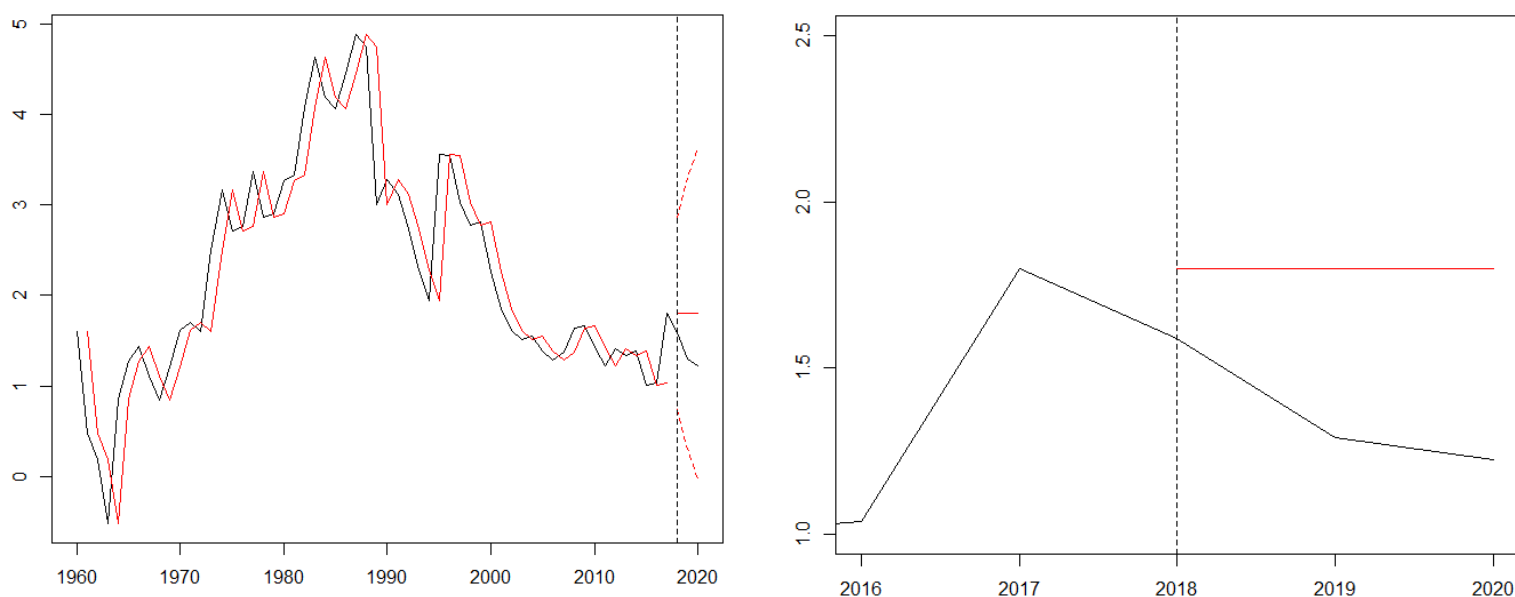
## 1.3. Modele ekstrapolacyjne

### 1.3.1. Wygładzanie wykładnicze

Proces analizy szeregu rozpoczęto od przeprowadzenia wygładzenia wykładniczego. Do analizy szeregu wykorzystano program statystyczny R. Dodatkowo, próba została podzielona na okresy in-sample (dane od 1960 roku do 2017 roku) oraz i out-of-sample (dane od 2018 roku do 2020 roku). W ramach wygładzania pozwolono programowi samemu wybrać najbardziej optymalną wartość, która wynosi 0.9999208.

Na rysunkach 3 i 4 postanowiono przedstawić graficznie wyniki uzyskanej estymacji dla alfy wyznaczonej przez program. Czerwonym kolorem oznaczono wartości prognozowane, natomiast czarną linią zaznaczono rzeczywiste dane. Na rysunku nr.3 uwzględniono również linie przedziałów ufności.

Rysunek 3, 4. Prognozy przy użyciu wykładniczego dla  $\alpha=0.9999208$

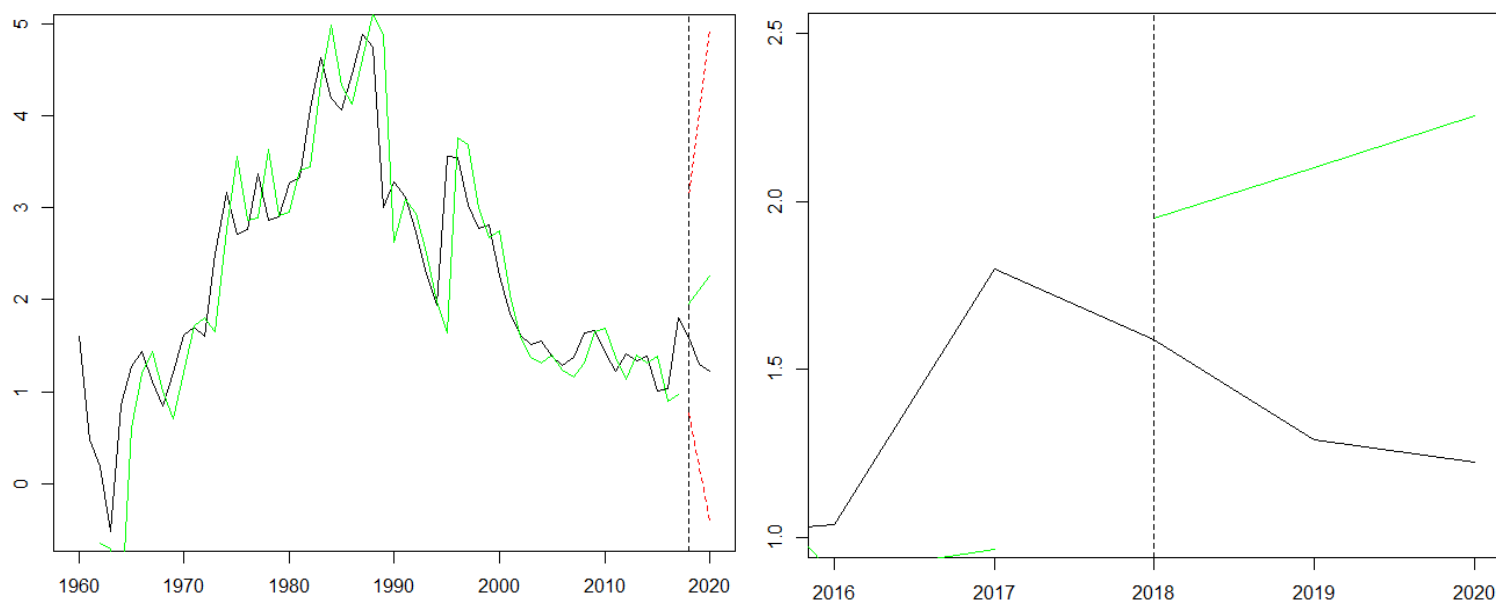


Z powyższych rysunków widać, iż wykładnicze nie jest najlepszą metodą estymacji, ponieważ uzyskane oszacowanie jest linią poziomą. Co więcej prognoza nie wydaje się być trafna, ponieważ w rzeczywistości, wartości CPI spadły. Dlatego też, zaleca się tego sposobu, tylko w momencie gdy chcemy przewidzieć dopasowanie na jeden okres do przodu.

### 1.3.2. Model Holta

W ramach niezadowolającej powyższej prognozy postanowiono wykorzystać inny sposób oszacowania, również pochodzący z klasy modeli ekstrapolacyjnych. Posłużono się zatem modelem Holta i wybrano iteracyjnie przez program wartości  $\alpha = 1$  i  $\beta = 0.2682065$ . Tak jak i w przypadku rysunków 3, 4 postanowiono przedstawić wyniki dla zoptymalizowanych danych w postaci graficznej. Tym razem kolorem zielonym zaznaczono wartości dopasowane poprzez model Holta, a kolorem czarnym oznaczono wartości rzeczywiste.

Rysunek 5,6. Prognozy przy użyciu modelu Holta dla  $\alpha=1$ ,  $\beta=0.2682065$



Co prawda uzyskana prognoza nie jest już w postaci poziomej, lecz zdecydowanie odbiega od danych rzeczywistych. Niemniej jednak, oszacowanie zachowało się zgodnie z tym jak wyglądały wartości w latach 2016-2017. Pomimo dwukrotnie nieudanej próby przewidzenia CPI od 2018 roku, postanowiono porównać obydwa modele pod względem ich błędów MAE, MSE, MAPE, AMAPE.

Tabela 1. Błędy dopasowania dla wykładniczego oraz modelu Holta w przedziale in-sample oraz out-of-sample

	Wykładnicze wykładnicze		Model Holta	
	$\alpha = 0.9999208$		$\alpha = 1$ $\beta = 0.2682065$	
	In sample	Out of sample	In sample	Out of sample
<b>MAE</b>	0.3833282	0.4306678	0.4047158	0.7343118
<b>MSE</b>	0.2833984	0.2106866	0.3679873	0.6167553
<b>MAPE</b>	0.2645105	0.3316115	0.2880637	0.5659243
<b>AMAPE</b>	0.1994191	0.1387862	0.1921173	0.2124798

Porównując wartości in-sample dla obu metod, ciężko stwierdzić, który z modeli jest lepszy ze względu na zbliżone wartości statystyk. Jednakże w przypadku out-of-sample

ewidentnie widać, iż wygładzenie wykładnicze cechuje się mniejszymi błędami, niż model Holta zarówno w porównaniu statystyk MAE/MSE, jak i MAPE/AMAPE.

## **1.4. Model ARIMA**

### **1.4.1. Stacjonarność szeregu**

Po wstępnym modelowaniu poprzez modele ekstrapolacyjne, postanowiono wyestymować model ARIMA dla szeregu niesezonowego. Ze względu na fakt, iż jest to sposób bardziej restrykcyjny, na początku należy sprawdzić, czy w szeregu istnieje problem niestacjonarności zmiennej. Do zweryfikowania, czy taka zależność ma miejsce, rozpoczęto procedurę badania autokorelacji zmiennych testem Breuscha-Godfreya. Ze względu na to, iż analizowane są dane roczne, sprawdzono tylko pierwsze rozszerzenie, które wydaje się być wystarczające do poprawnej oceny. P-value testu BG wynosi 0.7872, tak więc w moim szeregu nie występuje problem autokorelacji zmiennych.

W związku z uzyskaniem pozytywnego wyniku, przeprowadzono test Dickey-Fullera, którego hipoteza zerowa mówi o występowaniu zmiennej niestacjonarnej. W tym teście istotna jest analiza wartości *test-statistic*, która w moim przypadku wynosi -0.7503, co nie pozwala mi odrzucić hipotezy zerowej dla 3 punktów krytycznych.

Celem uzyskania stacjonarności, zróżnicowano zmienną jednokrotnie i ponownie przeprowadzono test na autokorelację zmiennych. P-value w teście BG wyniosło 0.9244, co umożliwiło ponowne przeprowadzenie testu D-F już na zmiennych zróżnicowanych. Statystyka w teście wyniosła -7.3185, co pozwoliło mi stwierdzić, iż różnicowanie pierwszego stopnia było wystarczające. Celem potwierdzenia uzyskanego wyniku, przeprowadzono test KPSS, którego hipoteza zerowa jest odwrotna tzn. głosi, iż zmienna jest stacjonarna. Wynik badania wyniósł 0.1376, tak więc nie miałem podstaw do odrzucenia hipotezy zerowej dla wszystkich trzech wartości krytycznych.

Rysunek 7. Test Dickeya-Fullera  
dla zróżnicowanej zmiennej

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1 -0.9646      0.1318  -7.318 1.13e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5202 on 55 degrees of freedom
Multiple R-squared:  0.4934,    Adjusted R-squared:  0.4842
F-statistic: 53.56 on 1 and 55 DF,  p-value: 1.134e-09

value of test-statistic is: -7.3185

critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.6 -1.95 -1.61
```

Rysunek 8. Test KPSS dla  
zróżnicowanej zmiennej

```
#####
# KPSS Unit Root Test #
#####

Test is of type: mu with 3 lags.

value of test-statistic is: 0.1376

critical value for a significance level of:
      10pct  5pct 2.5pct 1pct
critical values 0.347 0.463 0.574 0.739
```

Ponadto sprawdzono jeszcze występowanie białego szumu. W tym celu posłużono się dwoma testami Boxa-Pierce'a oraz Ljunga-Boxa. W przypadku drugiego testu, warto mieć na uwadze, iż jest on znacznie skuteczniejszy dla małej ilości danych. W moim szeregu jest to ważne (in-sample tylko 58 danych), ponieważ dla testu Ljunga-Boxa p-value wynosi 0.01304, więc tylko dla p-value=0.01 mam brak podstaw do odrzucenia hipotezy zerowej mówiącej o występowaniu białego szumu. Natomiast w teście Boxa-Pierce'a p-value wynosi 0.1921, co nie pozwala mi odrzucić hipotezy  $h_0$ .

Rysunek 9,10. Testy na biały szum

Box-Ljung test

```
data: Dy
x-squared = 41.96, df = 24, p-value = 0.01304
```

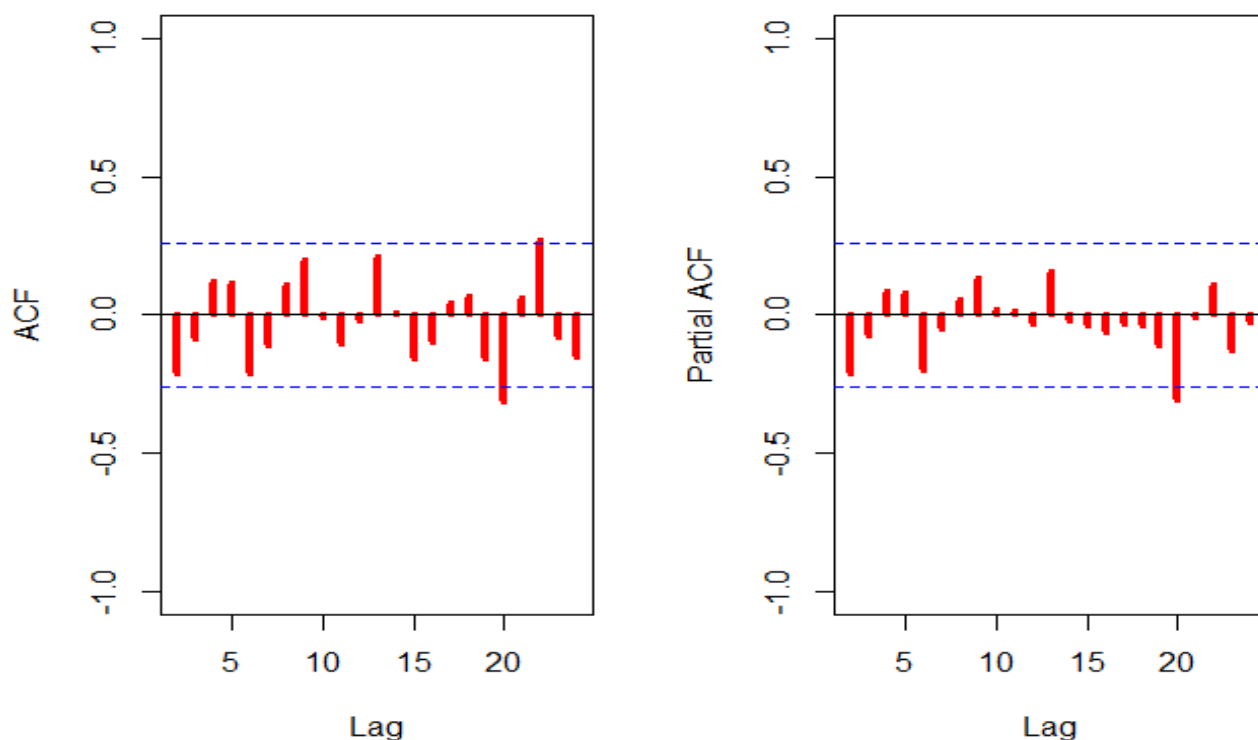
Box-Pierce test

```
data: Dy
x-squared = 29.782, df = 24, p-value = 0.1921
```

### 1.4.2. Identyfikacja i weryfikacja

W tym etapie estymacji modelu przeprowadzono próbę określenia rzędów AR i MA. Posłużono się graficzną interpretacją funkcji ACF i PACF, z której wynika, iż najbardziej odpowiednim modelem wydaje się być ARIMA(0,1,0). Co prawda widać, że w dalekich opóźnieniach pojedyncze wypustki wychodzą ponad przedział ufności, jednakże ze względu na roczny, niesezonowy charakter danych, można je uznać za nieistotne.

Rysunek 11, 12. Funkcje ACF i PACF dla zróżnicowanych zmiennych



Celem sprawdzenia, czy aby na pewno model ARIMA(0,1,0) jest najlepszy do interpretacji, poprzez metodę od ogółu do szczegółu oraz kryteria informacyjne będą postarano się potwierdzić wyniki uzyskane dzięki analizie ACF i PACF.

Tabela 2. Metoda od ogółu do szczegółu i kryteria informacyjne dla szeregów klasy ARIMA

Model	AIC	BIC
ARIMA(3,1,3)	93.32513	109.6695
ARIMA(3,1,2)	92.1737	106.4751
ARIMA(2,1,3)	92.17383	106.4752
ARIMA(2,1,2)	90.17509	102.4334
ARIMA(2,1,1)	96.74318	106.9584
ARIMA(1,1,2)	97.27439	107.4897
ARIMA(1,1,1)	95.14917	103.3214
ARIMA(0,1,1)	95.75218	101.8813
ARIMA(1,1,0)	95.81183	101.941
ARIMA(0,1,0)	93.88473	97.97083

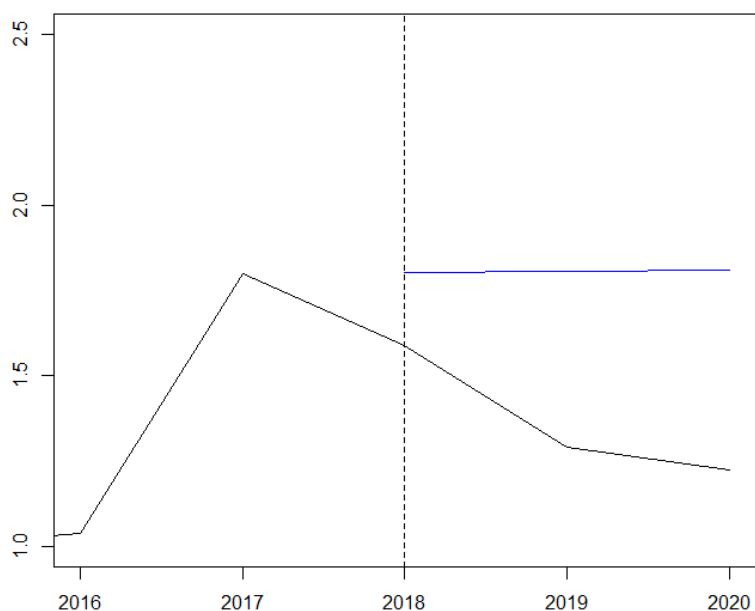


Zielonym kolorem zaznaczono najniższe wartości AIC i BIC. Z jednej strony, można żałować, iż nie pokrywają się one dla jednego modelu, ale z drugiej strony udało się uzyskać najniższą wartość BIC dla modelu ARIMA(0,1,0). W związku z tym, ponieważ wyniki uzyskane z wykresów ACF i PACF oraz z kryteriów informacyjnych częściowo się pokrywają, to w dalszej analizie będzie wykorzystywany wspomniany już wcześniej model posiadający proces autoregresyjny oraz średnią ruchomą na poziomie 0. Dla wybranego modelu przeprowadzono testy na biały szum dla reszt, które wykazały, iż nie mam podstaw do odrzucenia hipotezy zerowej. Dlatego też, rozpoczęto procedurę prognozowania szeregu.

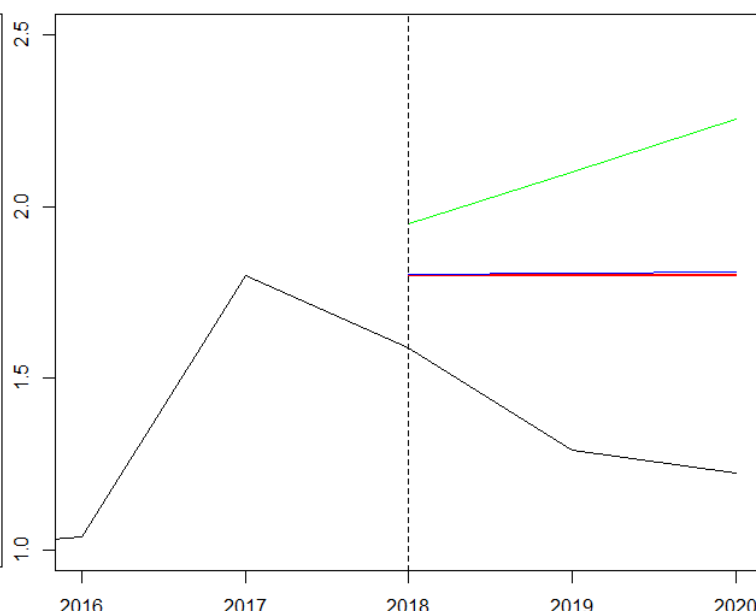
### 1.4.3. Prognoza i porównanie

Ostatnim krokiem w analizie jest oszacowanie dla okresu out-of-sample wartości dla ARIMY(0,1,0) oraz porównanie otrzymanego wyniku z modelami ekstrapolacyjnymi. Na poniższym rysunku dla lat 2016-2020 przedstawiono: czarnym kolorem rzeczywiste wartości, zielonym kolorem model Holta, czerwonym kolorem wykładnicze oraz niebieskim kolorem oszacowany model ARIMA.

Rysunek 13. Oszacowanie wartości out-of-sample dla modelu ARIMA(0,1,0)



Rysunek 14. Porównanie wszystkich oszacowanych modeli



Jak widać na powyższym wykresie, prognoza dla ARIMY jest bardzo zbliżona do prognozy dla wykładniczego. Na rysunku 14 wręcz widać, iż obie linie prawie na

siebie nachodzą. Można również stwierdzić, iż oba te modele wypadają najlepiej w porównaniu do sytuacji rzeczywistej. W przypadku modelu Holta, niestety prognoza wskazuje na zupełnie inny kierunek, dlatego też nie można jej uznać za rzetelną.

Tabela 3. Średnie oszacowania błędów prognoz dla wszystkich wyestymowanych modeli

	<b>Wygladzenie wykładnicze</b>	<b>Model Holta</b>	<b>ARIMA(0,1,0)</b>
<b>MAE</b>	0.4306678	0.7343118	0.4378793
<b>MSE</b>	0.2106866	0.6167553	0.2178295
<b>MAPE</b>	0.3316115	0.5659243	0.3371746
<b>AMAPE</b>	0.1387862	0.2124798	0.1407302

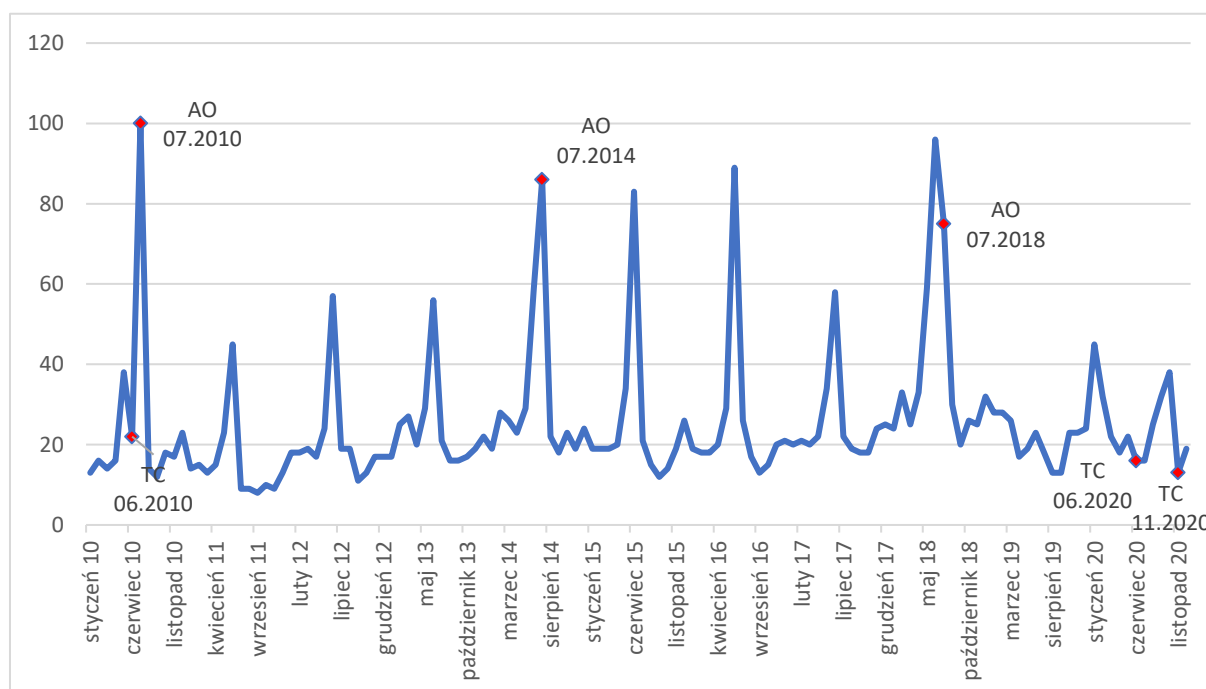
Analizując uśrednione błędy estymacji, najlepszym modelem pod względem błędów MAE/MSE jest wygladzenie wykładnicze, jednakże jego błędy są wręcz tożsame z modelem ARIMA(0,1,0). Natomiast w przypadku błędów MAPE/AMAPE najlepsza prognoza wydaje się być również dla wygladzenia wykładniczego, jednakże ponownie te wartości są bardzo zbliżone do modelu ARIMA(0,1,0). Niestety w tym zestawieniu model Holta w obu przypadkach wypada najgorzej ze wszystkich.

## II. SZEREG SEZONOWY

### 2.1. Opis danych

Szereg sezonowy, który zostanie podany analizie dotyczy liczby wyszukiwań na całym świecie amerykańskiego koszykarza Lebrona Jamesa w stosunku do łącznej liczby szukanych fraz w platformie Google w skali od 0 do 100. Dane pochodzą z bazy danych "Google Trends"<sup>2</sup> i mają charakter miesięczny (dostęp online na dzień 02.04.2021). Liczba obserwacji wynosi 132, a pierwsza z nich odnosi się do stycznia 2010 roku, natomiast ostatnia do grudnia 2020 roku.

Rysunek 15. Liczba wyszukiwań w Google Lebrona Jamesa w latach 2010-2020.



Na rysunku nr. 15 czerwone punkty przedstawiają wartości odstające, które zostały wyróżnione przez program Demetra. Na wykresie można zaobserwować 6 obserwacji nietypowych, przy czym 3 z nich (czerwiec 2010, czerwiec 2020 oraz listopad 2020) nie posiadają jasnego uzasadnienia historycznego. Natomiast w przypadku pozostałych odstających obserwacji, takie zależności już istnieją. W lipcu 2010 nastąpił transfer zawodnika do klubu Miami Heat, w lipcu 2014 powrócił on do swojego rodzimego klubu Cleveland

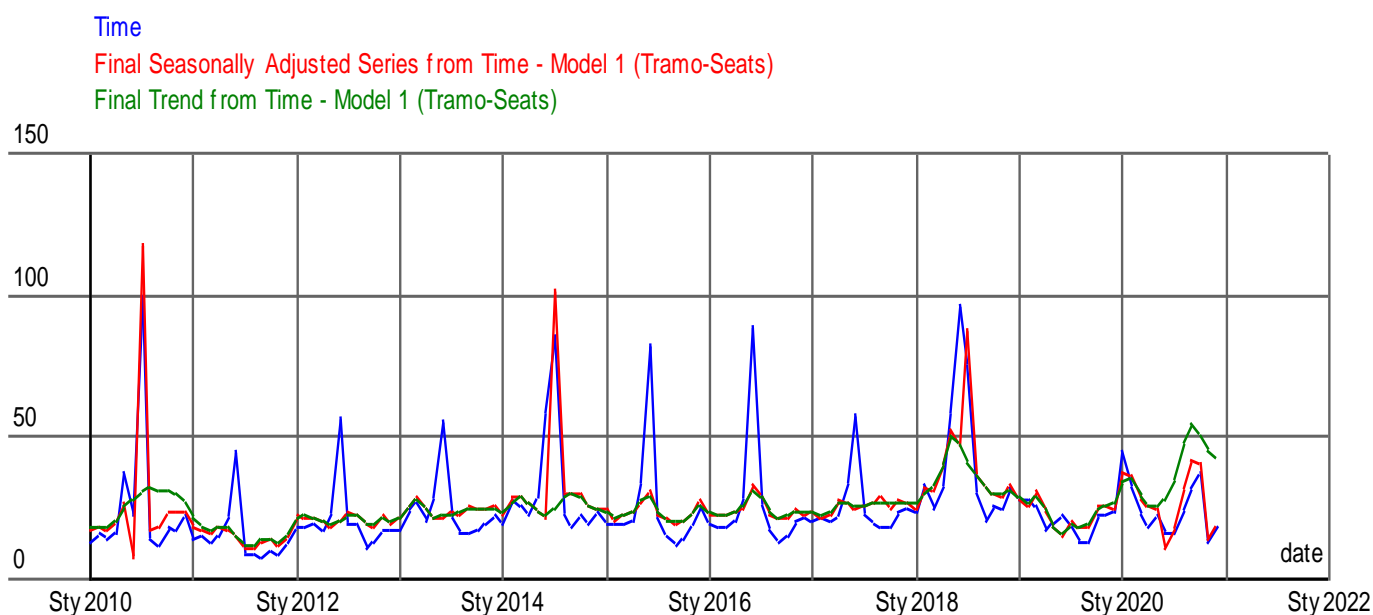
<sup>2</sup> <https://trends.google.com/trends/explore?date=all&q=lebron%20james>

Cavaliers. W przypadku lipca 2018 był to kolejny transfer zawodnika, jednakże do klubu Los Angeles Lakers.

## 2.2. Dekompozycja szeregu

Wstępnej analizy szeregu dokonano w programie Demetra, natomiast samą dekompozycję szeregu przeprowadzono przy użyciu metody TRAMO/Seats.

Rysunek 16. Dekompozycja szeregu sezonowego.



W powyższym szeregu, z pewnością można wyróżnić sezonowość. Najwięcej wyszukiwań corocznie jest w okresie wakacyjnym, ponieważ jest to moment zakończenia sezonu w NBA, co wiąże się z powstawaniem nowych informacji o potencjalnym transferze. Amplituda wahań jest względnie stała, więc można sądzić, iż zaobserwowano sezonowość addytywną. Program zasugerował, aby do interpretacji szeregu posłużyć się przekształceniem logarytmicznym, co też postanowiono wykorzystać w poniższej analizie. Taki zabieg pomoże zmniejszyć wartość wariancji.

## 2.3. Model ekstrapolacyjny

Do pierwszej próby wyestymowania szeregu użyto modeli ekstrapolacyjnych Holta-Wintersa dla szeregów sezonowych. Najbardziej optymalną formą wydaje się być postać

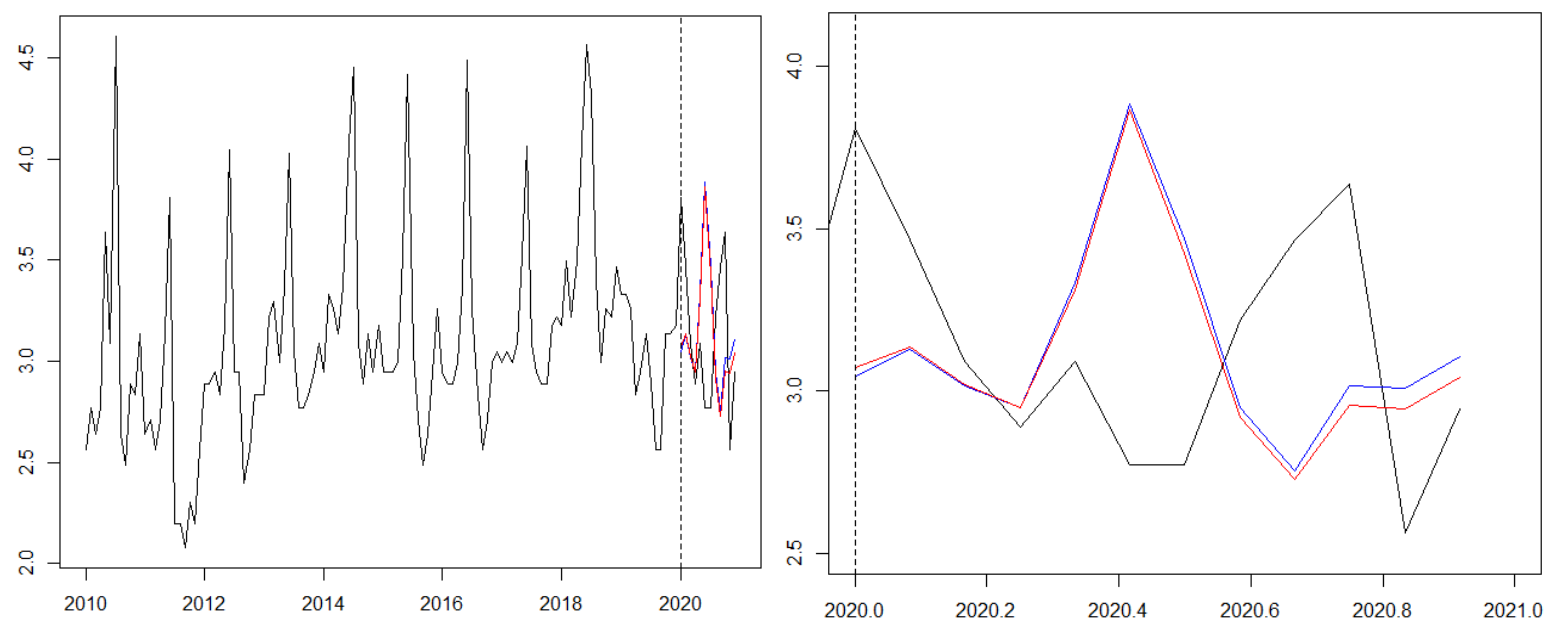
addytywna, lecz w celu upewnienia się i porównania, wykorzystano również postać multiplikatywną. Jednakże na samym początku badania, podzielono dane na próbę in-sample, która zawiera wartości od stycznia 2010 do grudnia 2019 oraz na próbę out-of-sample będącą w przedziale od stycznia 2020 do grudnia 2020 (odpowiadając pełnemu cyklowi sezonowemu).

W obydwu modelach pozwoliliśmy programowi znaleźć najoptymalniejsze wartości alfa, beta oraz gamma. W przypadku modelu addytywnego wynoszą one:  $\alpha = 0.4132592$ ,

$\beta = 0.01770069$ ,  $\gamma = 0.5891683$ , natomiast dla modelu multiplikatywnego:  $\alpha = 0.5515236$ ,

$\beta = 0.01068498$ ,  $\gamma = 0.6794855$ . Na poniższych rysunkach przedstawiono graficzne porównywanie obydwu wyestymowanych modeli wraz ze zlogarytmowanymi wartościami rzeczywistymi. Kolor czerwony odpowiada szeregowi multiplikatywnemu, kolor niebieski szeregowi addytywnemu, a kolor czarny realnym wartościom.

Rysunek 17,18. Porównanie wyestymowanych modeli sezonowych



Z powyższych wykresów można wywnioskować, iż prognozy są całkiem do siebie zbliżone, a okresem w którym obie prognozy się ze sobą pokrywają są pierwsze dwa kwartały 2020 roku. Celem weryfikacji czy model addytywny jest lepszy, sprawdzono jeszcze średnie błędy dla obu okresów.

Tabela 4. Błędy dopasowania dla modelu addytywnego i multiplikatywnego

	Sezonowość addytywna		Sezonowość multiplikatywna	
	In sample	Out of sample	In sample	Out of sample
<b>MAE</b>	0.2393679	0.4576338	0.2199155	0.4464328
<b>MSE</b>	0.1369323	0.3059704	0.1081440	0.2984963
<b>MAPE</b>	0.08064253	0.14671729	0.0723249	0.1422227
<b>AMAPE</b>	0.04025651	0.07157057	0.03577615	0.06995770

Okazuje się, iż minimalne mniejsze błędy zarówno dla okresów in-sample jak i out-of-sample są obserwowane dla sezonowości multiplikatywnej. W związku z tym, należy uznać, że ten model lepiej dopasował się do wartości realnych. Warto jednak nadmienić, iż różnice w błędach pomiędzy oboma modelami są niewielkie.

## 2.4. Model SARIMA

### 2.4.1. Sezonowość i stacjonarność szeregu

Aby w pełni, móc budować model SARIMA, nadrzędnym krokiem jest odsezonowanie zmiennych oraz doprowadzenie ich do stacjonarności. Tak jak i w przypadku modeli ekstrapolacyjnych, okresy in-sample i out-of-sample pozostają bez zmian. Początkowo w celu zweryfikowania sezonowości przeprowadzono test Dickeya, Haszy, Fullera, którego hipoteza zerowa, głosi iż występuje sezonowy pierwiastek jednostkowy.

W związku z tym, w pierwszej kolejności sprawdziłem występowanie autokorelacji za pomocą testu Breuscha-Godfrey. Niestety, nie dla wszystkich stopni (6 włącznie) posiadałem brak autokorelacji zmiennych. Dlatego też, zastosowano rozszerzony test DHF. Okazało się, iż dopiero dla drugiego stopnia tego testu, jestem w stanie w sposób satysfakcjonujący rozpatrzeć przypadek o braku autokorelacji zmiennych.

Rysunek 19. Rozszerzony test Dickeya, Haszy, Fullera

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.43647 -0.06580  0.06361  0.23925  1.42961

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
lag12lking   -0.01186    0.01347   -0.880  0.38088
lagd12lking    0.31248    0.09437    3.311  0.00128 **
lag2d12lking    0.27981    0.09436    2.965  0.00376 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4357 on 103 degrees of freedom
(14 observations deleted due to missingness)
Multiple R-squared:  0.2582,    Adjusted R-squared:  0.2365
F-statistic: 11.95 on 3 and 103 DF,  p-value: 0.0000008915

```

Istotna statystyka dla tego testu została zaznaczona kolorem żółtym i wynosi -0.880. Korzystając z tablic dla powyższej statystyki dla danych miesięcznych, gdzie ich liczba oscyluje wokół  $n=120$  dla  $p\text{-value}=0.05$ , wartość krytyczna wynosi -5.86. Zostałem więc zmuszony do odrzucenia hipotezy zerowej, co wiąże się z wprowadzeniem różnicowania sezonowego.

Następnie, przeprowadzono test Dickeya-Fullera celem sprawdzenia stacjonarności zmiennych. Niestety i tym razem, nie stwierdzono braku autokorelacji zmiennych, w związku z czym zastosowano rozszerzoną wersję testu, dla opóźnienia = 1. Statystyka  $\tau$  wyszła równa -4.0498, dlatego też odrzucono hipotezę zerową mówiącą o niestacjonarności zmiennych. Celem upewnienia, iż na pewno nie powinienem wprowadzić różnic regularnych po uwzględnieniu różnic sezonowych, przeprowadziłem test KPSS. Statystyka testowa wyniosła = 0.1157, dlatego też nie miałem podstaw do odrzucenia hipotezy zerowej mówiącej o stacjonarności zmiennych.

Rysunek 20. Rozszerzony test  
Dickeya Fullera

```

value of test-statistic is: -4.0498

critical values for test statistics:
    1pct   5pct  10pct
tau1 -2.58 -1.95 -1.62

```

Rysunek 21. Test KPSS

```

#####
# KPSS Unit Root Test #
#####

Test is of type: mu with 4 lags.

value of test-statistic is: 0.1157

critical value for a significance level of:
    10pct   5pct   2.5pct   1pct
critical values 0.347 0.463  0.574 0.739

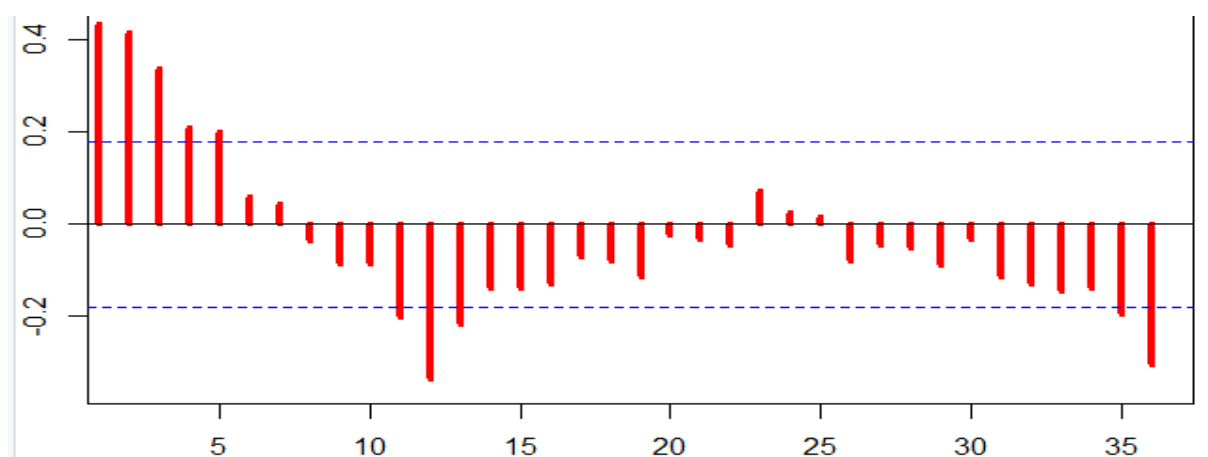
```

Po przeprowadzeniu powyższych testów, sprawdzono czy w uzyskanym szeregu nie występuje zjawisko białego szumu. Zarówno p-value'a dla testu Ljunga-Boxa, jak i Boxa-Pierce'a wyszły znacznie poniżej 0.000, więc odrzucono hipotezę zerową mówiącą o występowaniu białego szumu.

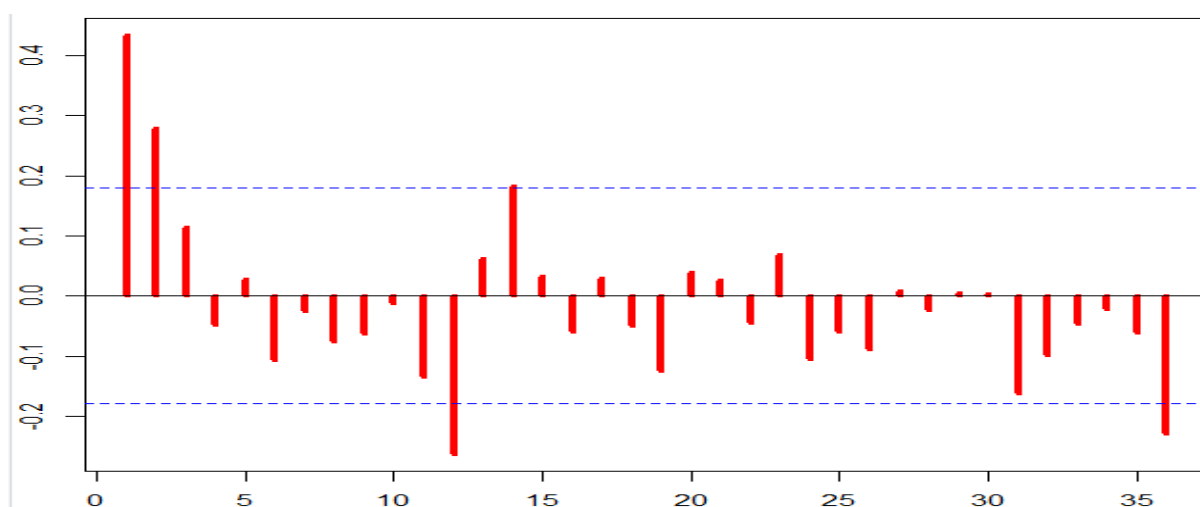
#### 2.4.2. Identyfikacja i weryfikacja

W następnym kroku należy zidentyfikować rzędy p, q, P, Q. Dlatego też, posłużono się graficzną interpretacją funkcji ACF i PACF, z której wynika, iż ciężko dokładnie określić co będzie najbardziej odpowiednim modelem.

Rysunek 22. Wykres funkcji ACF



Rysunek 23. Wykres funkcji PACF





Niemniej jednak, podjęto próbę oszacowania graficznego. Wydaje się, iż początkowo najbardziej optymalnym modelem do estymacji jest SARIMA(2,0,3)(2,1,2). Jednakże, sądzę, że nie byłoby błędem przypuszczać, że może to być nawet model SARIMA(1,0,1)(1,1,1). Dlatego też przeprowadzono metodę od ogółu do szczegółu, porównując kryteria informacyjne poszczególnych modeli.

Tabela 5. Metoda od ogółu do szczegółu i kryteria informacyjne dla szeregów klasy SARIMA

Model	AIC	BIC
SARIMA(2,0,3)(2,1,2)	92.00194	118.8233
SARIMA(2,0,3)(1,1,2)	87.63456	111.7737
SARIMA(2,0,3)(1,1,1)	91.83354	113.2906
SARIMA(2,0,3)(0,1,1)	91.55092	110.3258
SARIMA(2,0,3)(1,1,0)	91.55092	110.3258
SARIMA(2,0,2)(0,1,1)	91.46991	107.5627
SARIMA(1,0,2)(0,1,1)	89.62458	103.0352
SARIMA(2,0,1)(0,1,1)	89.60518	103.0158
SARIMA(1,0,1)(0,1,1)	90.16135	100.8899
SARIMA(0,0,1)(0,1,1)	111.94801	119.9944
SARIMA(1,0,0)(0,1,1)	102.47470	110.5211

Kolorami zielonymi zaznaczono najmniejsze wartości statystyk dla kryteriów informacyjnych AIC i BIC. Co prawda SARIMA(2,0,3)(1,1,2) wykazuje najmniejszy wskaźnik AIC, jednakże ten model okazuje się być nieistotny. Inaczej wygląda sytuacja w przypadku modelu SARIMA(1,0,1)(0,1,1), który posiada najmniejszą statystykę BIC oraz jest jak najbardziej istotny. Podjęto zatem decyzję, o użyciu tego modelu do dalszych prognoz. Dla wybranego modelu przeprowadzono testy na białe szumy dla reszt, które wykazały, iż nie mam podstaw do odrzucenia hipotezy zerowej. Dlatego też, rozpoczęto prognozę szeregu.

#### 2.4.3. Prognoza i porównanie

Ostatnim już krokiem w analizie jest oszacowanie dla okresu out-of-sample wartości dla SARIMY(0,1,0)(0,1,1) oraz porównanie otrzymanego wyniku z modelami

ekstrapolacyjnymi. Na poniższym rysunku dla okresu ostatnich 13 miesięcy przedstawiono: czarnym kolorem rzeczywiste wartości, niebieskim sezonowość multiplikatywną, czerwonym sezonowość addytywną oraz zielonym oszacowany model SARIMA.

Rysunek 24. Oszacowanie wartości out-of-sample dla SARIMA(1,0,1)(0,1,1)



Rysunek 25. Porównanie wszystkich oszacowanych modeli



Jak widać na powyższym wykresie, wszystkie prognozy są w miarę do siebie zbliżone, a szczególnie w okresie wakacyjnym. Na podstawie rysunku nr. 25 można stwierdzić, iż model multiplikatywny najlepiej prognozuje potencjalne rzeczywiste wartości. Dla potwierdzenia tej tezy, wyliczono średnie błędy dla prognoz.

Tabela 6. Średnie oszacowania błędów prognoz dla wszystkich wyestymowanych modeli

	Sezonowość addytywna	Sezonowość multiplikatywna	SARIMA(1,0,1)(0,1,1)
<b>MAE</b>	0.4576338	0.4464328	0.50645764
<b>MSE</b>	0.3059704	0.2984963	0.37277098
<b>MAPE</b>	0.14671729	0.1422227	0.16097518
<b>AMAPE</b>	0.07157057	0.06995770	0.07944275

Niestety w powyższym zestawieniu, model SARIMA(1,0,1)(0,1,1) pod względem średnich błędów dla prognoz wypada najgorzej. Dlatego też, należy stwierdzić, iż model

multiplikatywny dla tego zlogarytmowanego szeregu najlepiej zaprognozował wartości out-of-sample.

## **PODSUMOWANIE**

W pracy przedstawiono wyniki analizy szeregów czasowych: niesezonowego, który dotyczył kształtowania się CPI w Meksyku w latach 1960-2020 oraz sezonowego, gdzie dane miesięczne opierały się na liczbie wyszukiwań w platformie Google koszykarza Lebrona Jamesa w latach 2010-2020.

Rezultaty przeprowadzanych analiz wykazały, iż najmniejsze błędy w obu szeregach uzyskiwano dla modeli ekstrapolacyjnych, co przekłada się na lepsze dopasowanie szacowanych zmiennych do wartości rzeczywistych. Dodatkowo, w obu przypadkach błędy prognoz nie były względnie wysokie co można uznać za dobry wynik. Z przeprowadzonych analiz, z pewnością można wyciągnąć wniosek, iż tematyka badania szeregów czasowych oraz próby przewidywania przyszłości są niełatwą pracą. Każdy szereg wymaga indywidualnego podejścia oraz pewnej uniwersalności w obrębie przeprowadzania dla nich testów poprawności.