**Zaawansowana ekonometria szeregów czasowych - projekt zaliczeniowy.**

**Wykonanie: Natalia Owczarek (nr albumu:380549) i Patrycja Czernik (nr albumu: 380593)**

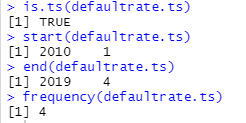
**Motywacja i cel projektu:**

Jedną z form działalności banku jest udzielanie kredytów, co wiąże się z dużym ryzykiem. Ryzyko to wynika m.in. z możliwości nieoczekiwanego niewywiązywania się klienta z warunków umowy. Definicja niewykonania zobowiązania przez dłużnika (default) obejmuje m.in. kryterium liczby dni niespłacalności oraz przesłanki wskazujące na prawdopodobieństwo niewykonania lub powrotu do stanu niewykonania zobowiązania. Kredytobiorca np. z powodu błędnej oceny jego zdolności kredytowej przez bank lub po prostu z powodu pogorszenia swojej sytuacji finansowej może spóźniać się z zapłatą rat kredytowych albo całkowicie zaprzestać spłacać zobowiązanie. Efektem tego są realne straty finansowe lub nawet zachwianie płynności finansowej banku. W celu uniknięcia dużych strat w związku z nieregulowaniem zobowiązań banki określają współczynnik PD (Probability of Default), który określa prawdopodobieństwo, że w przyszłości klient przestanie spłacać kredyt. Wyższe prawdopodobieństwo niewypłacalności wiąże się z wyższym ryzykiem dla instytucji. Porównanie współczynnika PD z akceptowanym przez bank poziomem ryzyka decyduje o odrzuceniu lub przyjęciu wniosku o kredyt czy pożyczkę. Pozwala to na przyjęcie do dalszej analizy wniosków, dla których potencjalne zyski są wyższe niż koszty związane z ryzykiem wystąpienia zaległości w spłacie kredytu. Określanie prawdopodobieństwa defaultu dostarcza informacji na temat pożyczkobiorców. Niespłacanie pożyczki pogorszy ocenę kredytową pożyczkobiorcy, utrudniając otrzymanie zatwierdzeń pożyczek w przyszłości. Nawet gdyby przyszła pożyczka została zatwierdzona, niski wynik kredytowy prawdopodobnie doprowadziłby do wyższego oprocentowania pożyczek. Współczynniki niewypłacalności mogą być reprezentatywne dla warunków ekonomicznych. Defaultrate jest wykorzystywany przez ekonomistów, jako jedna ze składowych do oceny ogólnego stanu gospodarki. Jest on wysoki w okresach spowolnienia gospodarczego i niski w okresach wzrostu gospodarczego.

Dokonywanie prognoz niespłacalności na kolejne okresy, pozwala na oszacowanie ryzyka na które narażony jest bank. Informacje te można wykorzystać, aby pomóc instytucji finansowej w podejmowaniu lepszych decyzji kredytowych i zarządzaniu ryzykiem. Banki powinny opracować i wdrożyć zasady zarządzania ryzykiem kredytowym, obejmujące, skuteczne procedury służące identyfikacji i szacowaniu ryzyka kredytowego, jego przegląd, ocenę i raportowanie oraz odpowiednie mechanizmy kontroli, w tym wewnętrznej, obejmujące należyte procedury administracyjne i księgowe. Ponadto, bank powinien posiadać dokumentację dotyczącą poszczególnych dłużników. W celu redukcji ryzyka kredytowego banki wprowadzają adekwatną do sytuacji gospodarczej strategię kredytową oraz ceny kredytów dla poszczególnych grup klientów. Zazwyczaj im wyższe ryzyko niespłacalności zobowiązania, tym wyższa marża kredytu.

**ANALIZA DANYCH SZEREGU:**

**Wprowadzamy szereg default rate'u (defaultrate.ts):**



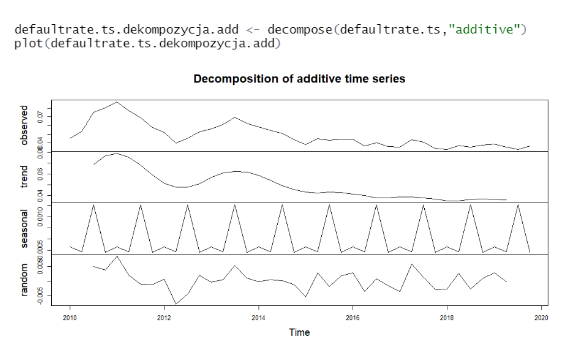
Szereg niespłacalności kredytów jest szeregiem czasowym. Zawiera on dane kwartalne, zaczynając od pierwszego kwartału 2010 roku, do czwartego kwartału 2019 roku.

**Przedstawienie danych na wykresie:**



Wykres pokazuje znaczny spadek współczynnika niespłacalności kredytów, od roku 2015 wartości tej zmiennej wykazują znacznie mniejsze wahania, oscylują około 4%. Podczas gdy w latach 2010-2014 wartości te wahały się od 4% do nawet 8%. Wysoki defaultrate w tych latach był efektem panującego kilka lat wcześniej ogólnoświatowego kryzysu ekonomicznego (2007-2009). Trudna sytuacja gospodarcza spowodowała problemy finansowe i była przyczyną trudności związanych ze spłatami zobowiązań.

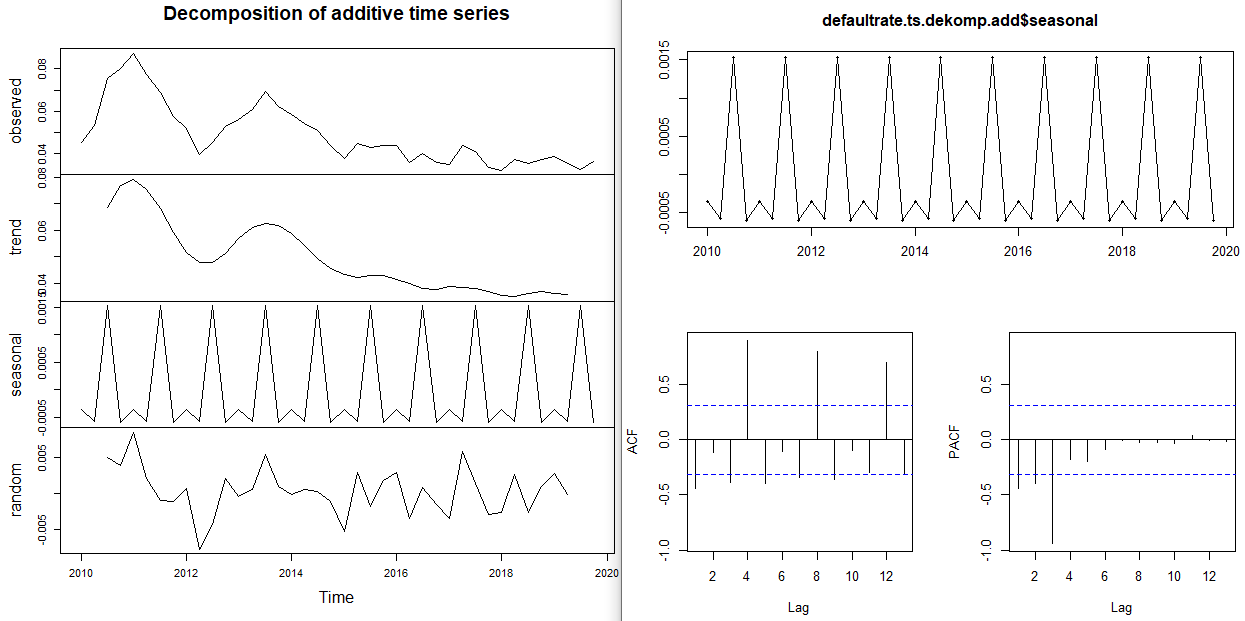
**Dekompozycja szeregu czasowego:**

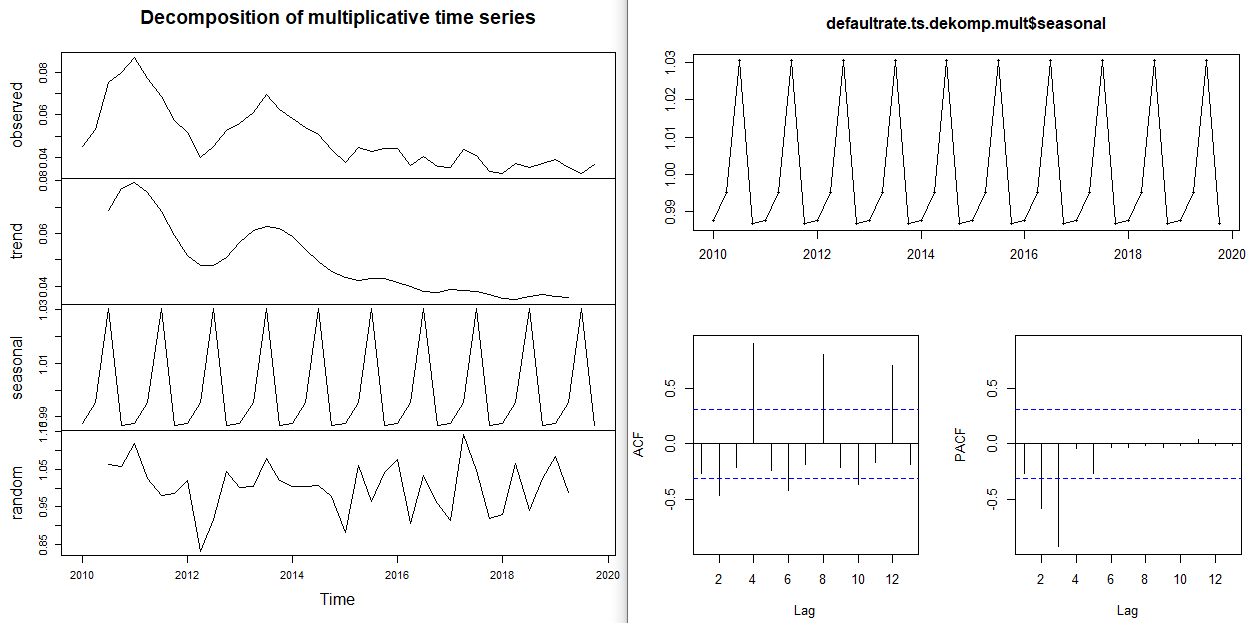
****

Na wykresie dekompozycji szeregu czasowego widoczny jest trend wielomianowy, na skutek pokłosia kryzysu finansowego z lat 2007-2009 trend charakteryzuje się tendencją wzrostową w latach 2010-2011, później w latach 2011-2012 występuje tendencja spadkowa. Kolejny wzrost rozpoczyna się w 2012 roku, a jedną z jego przyczyn jest wzrost bezrobocia w Polsce w tamtym czasie. Od 2013 roku obserwujemy tendencję spadkową. W całym analizowanym okresie pojawiają się skoki ilości niespłacanych pożyczek co może świadczyć o tym, że w tym szeregu występuje sezonowość. Wahania te są jednak na tyle niskie, że nie ma możliwości oceny występowania sezonowości przed wykonaniem szczegółowych badań ilościowych. Ponadto z wykresu dekompozycji wynika, że szereg zawiera wahania losowe.

**Dekompozycja addytywna i multiplikatywna ze względu na sezonowość:Obraz zawierający tekst

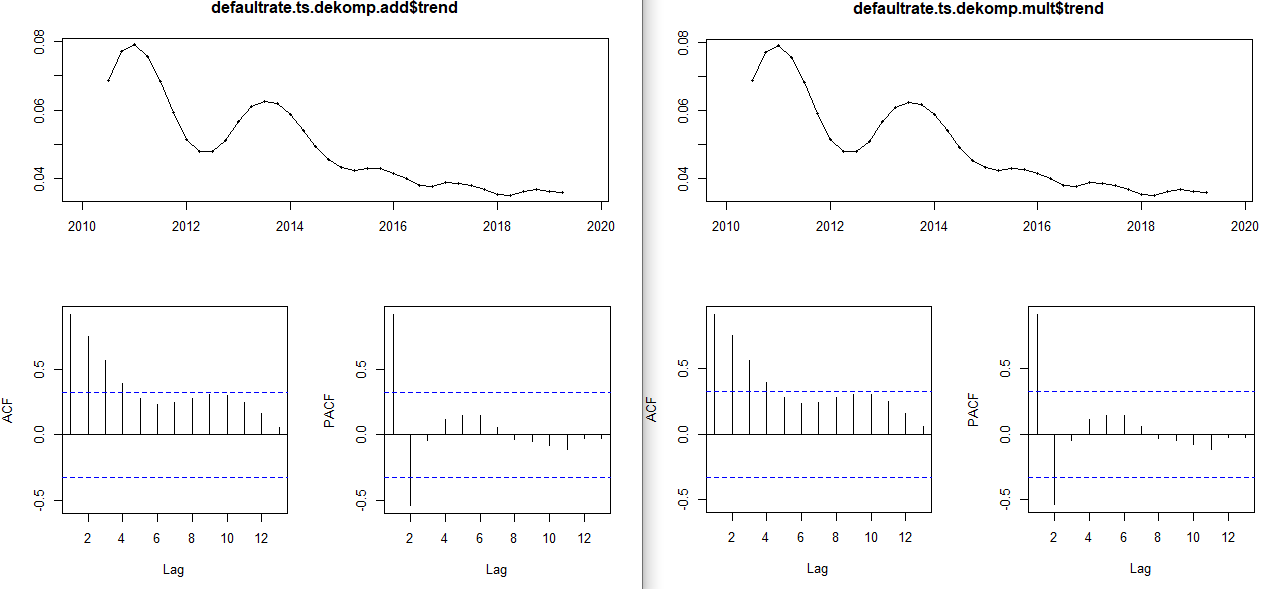
Opis wygenerowany automatycznie**

****

****

Wartość ACF oraz PACF sugerują korelację wartości w szeregu. Te dwa rodzaje korelacji pozwalają na ustalenie, że sezonowość jest istotnym czynnikiem wpływającym na dane.

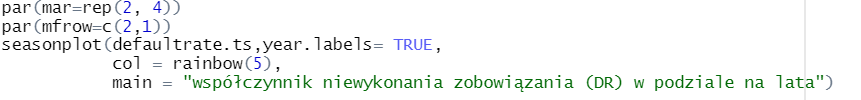
**Dekompozycja addytywna i multiplikatywna ze względu na trend:**

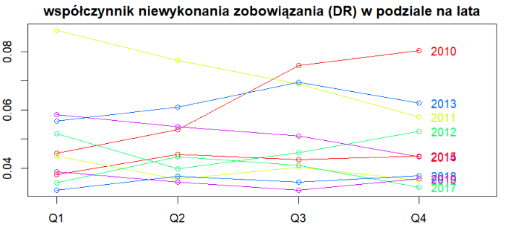


W badanym szeregu występuje trend. Na wykresie funkcji autokorelacji częściowej (PACF) występują dwa istotne opóźnienia. Oznacza to, że wartość defaultrate jest skorelowana z dwoma poprzednimi wartościami.

W dekompozycji addytywnej oznacza to, że trend jest skorelowany z dwoma poprzednimi wartościami trendu, co sugeruje, że trend nie jest liniowy.

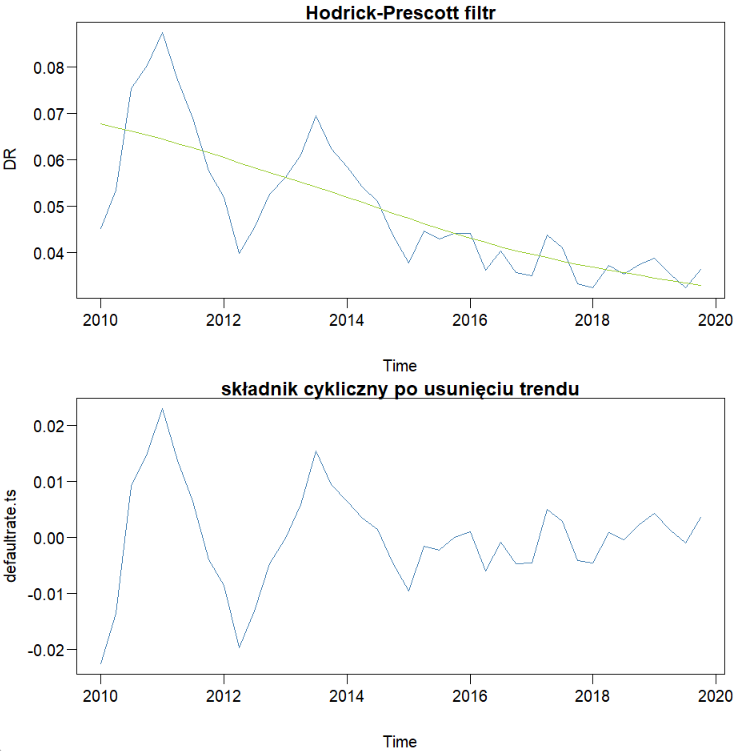
**Sprawdzenie występowania sezonowości:**



****

**Obraz zawierający tekst

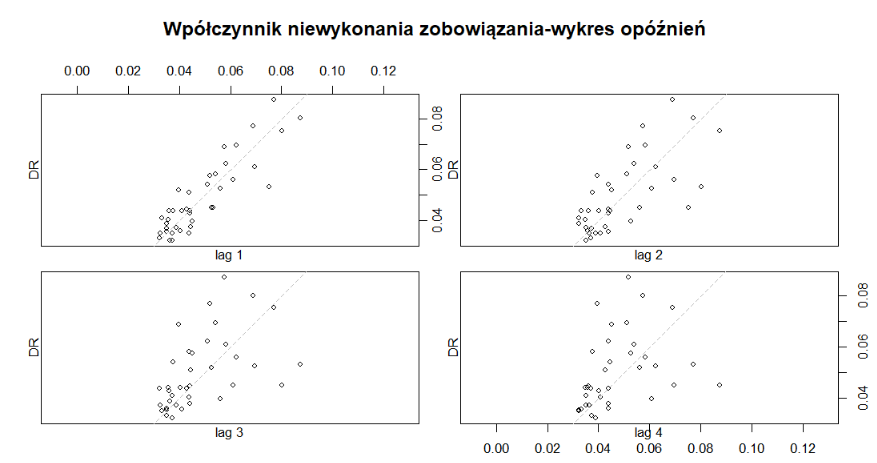
Opis wygenerowany automatycznie**

**

Zarówno wykres obserwowanych wartości w podziale na poszczególne kwartały jak i wykres po usunięciu trendu nie wskazują na występowanie sezonowości w badanym szeregu. Dane nie wykazują porównywalnych wzrostów lub spadków wartości w zależności od analizowanego kwartału w kolejnych latach.

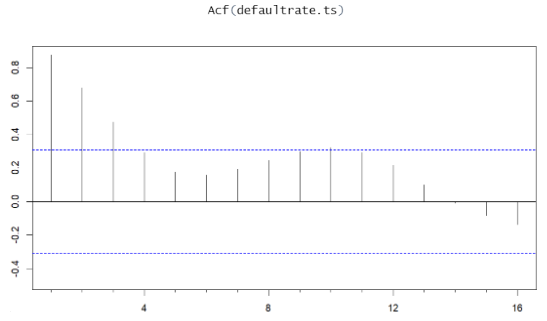
**BADANIE STACJONARNOŚCI SZEREGU:**

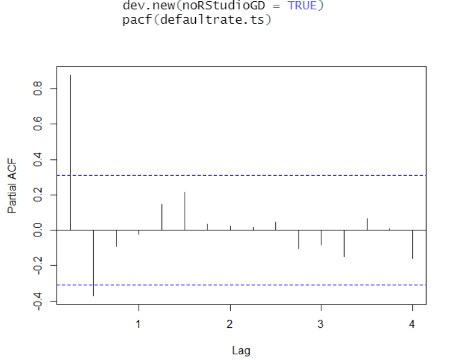
**Wykres opóźnień:**



Wraz ze wzrostem opóźnienia widoczna jest coraz mniejsza zależność opóźnień, ponieważ wraz ze wzrostem opóźnienia dane są w coraz mniejszym stopniu skupione wokół przekątnej. Może to wstępnie wskazywać na brak występowania stacjonarności, natomiast aby sprawdzić to dokładniej należy zbadać dane na podstawie wykresów ACF i PACF.

**Funkcja autokorelacji (ACF) i autokorelacji cząstkowej (PACF) badanego szeregu.**

****

****

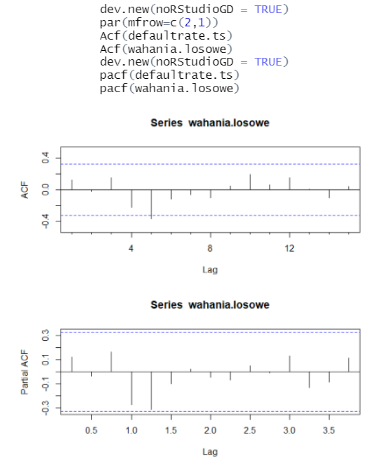
Funkcja ACF ma 3 istotne piki, dla procesu stacjonarnego ACF wraz ze wzrostem rzędu opóźnień powinien szybko zanikać. W tej sytuacji mamy do czynienia z silną korelacją między każdą wartością szeregu, a wartością występującą trzy okresy później. Co wskazuje na niestacjonarność szeregu.

Funkcja PACF wykazuje dwa istotne „piki”. Pierwszy z nich wskazuje na dodatnią korelację między kolejnymi opóźnieniami po eliminacji wpływu pośrednich opóźnień, drugi wskazuje na ujemną korelację. To znaczy, że szereg nie jest stacjonarny.

Zarówno dla wykresu ACF jak i PACF zaobserwowano istotne statystycznie opóźnienia. Na tej podstawie można stwierdzić, że DR zależy istotnie od swoich przeszłych wartości. Na podstawie funkcji ACF jak i PACF wnioskować można, że badany szereg jest niestacjonarny.

**BADANIE STACJONARNOŚCI SKŁADNIKA LOSOWEGO:**

**Funkcja autokorelacji (ACF) i autokorelacji cząstkowej (PACF):**

****

Zarówno dla wykresu ACF, jak i PACF nie ma żadnych istotnych „pików”, zatem wahania losowe są procesem stacjonarnym. Składnik losowy jest białym szumem. Wahania losowe nie wpływają istotnie na kolejne opóźnienia, co należy ocenić pozytywnie.

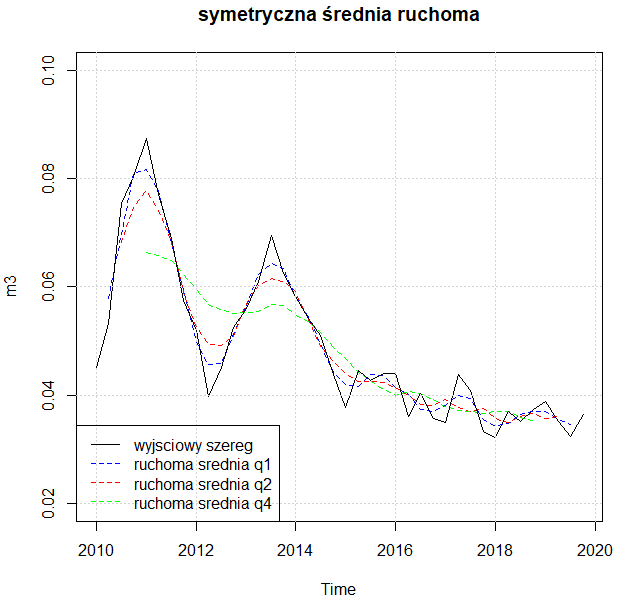
**BADANIE TRENDU:**

**Filtrowanie:**

Filtrowanie pozwala na wyszczególnienie trendu długookresowego z szeregu czasowego. Jedną z metod filtrowania jest średnia ruchoma prosta – która wraz ze wzrostem q eliminuje coraz więcej składowych szeregu o wysokich częstościach np. wahań losowych, tym samym bardziej wygładzając szereg i pozostawiając sam trend. Innymi metodami są filtry Hodricka-Prescotta (hp), Baxtter-Kinga (bk) oraz Christiano-Fitzgeralda (cf).

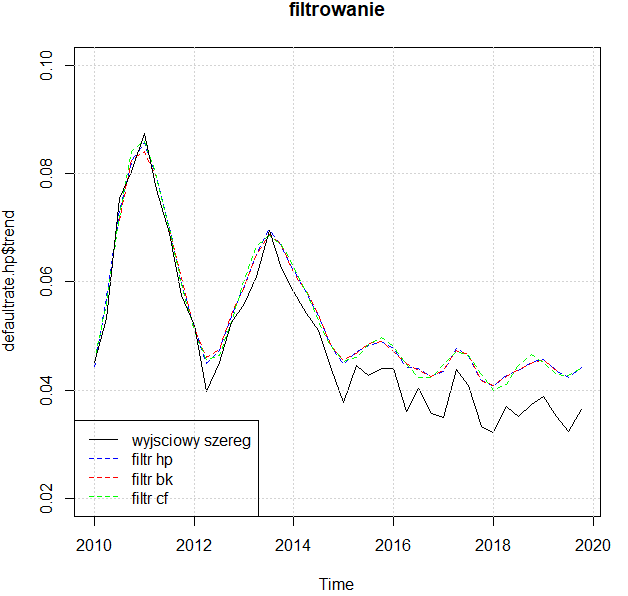
**Obraz zawierający tekst

Opis wygenerowany automatycznie**

****

**Obraz zawierający tekst

Opis wygenerowany automatycznie**

****

Zastosowanie tych dwóch metod umożliwiło usunięcie z modelu wahań krótkookresowych, czyli składników zakłócających widok trendu np. wahania losowe i sezonowe. Te działania sprawiły, że trend długookresowy jest bardziej wyraźny. Dane wykazują duże wahania, a linia trendu faluje, jednak biorąc pod uwagę cały okres można zauważyć tendencję malejącą. Szereg po zastosowaniu filtrów po roku 2014 przyjmuje wyższe wartości, niż szereg bazowy. Oznacza to, że filtry wyeliminowały z szeregu drobne wahania i szum. Zatem po 2014 roku na kształt szeregu bazowego miało wpływ więcej wahań losowych. Po 2015 roku szereg po zastosowaniu filtra nie dopasowuje się dobrze do wyjściowego szeregu, wahania krótkookresowe są istotne dla tego modelu, zatem do prognozowania wykorzystamy szereg wyjściowy.

**DOPASOWANIE TRENDU:**

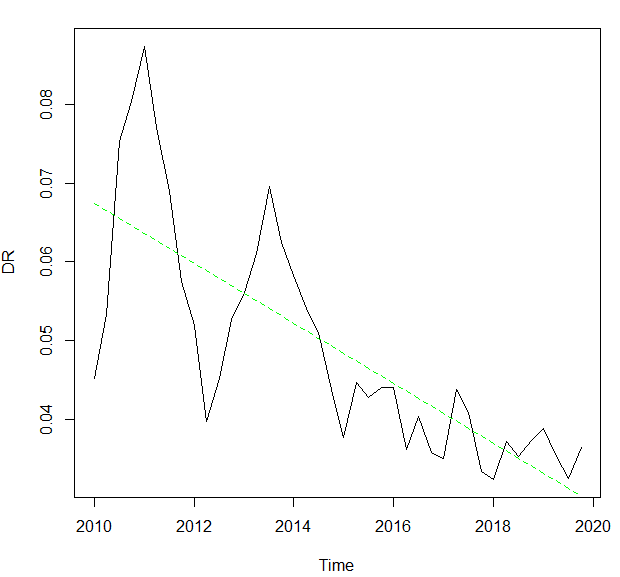
**Dekompozycja na podstawie modelu regresji:**

1. Trend

Obraz zawierający tekst

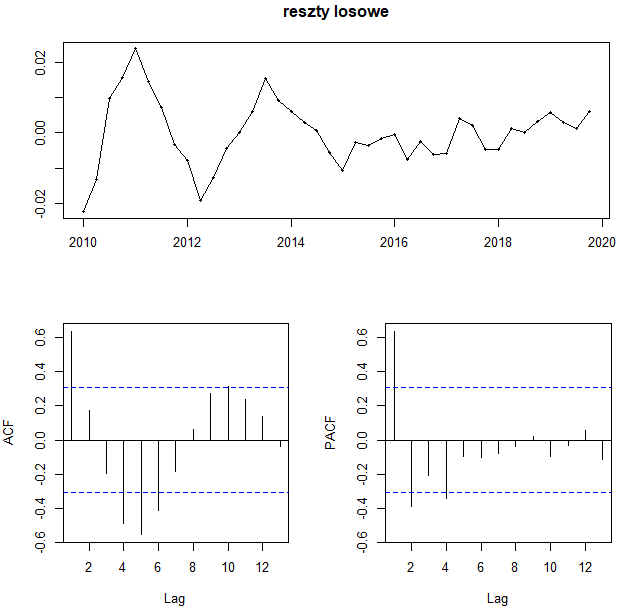
Opis wygenerowany automatycznie





Z uwagi na p-value <= 5% uznajemy, że zmienna mówiąca o trendzie jest istotna dla modelu regresji.





Reszty losowe nie są stacjonarne, zależą od swojej wartości w poprzednim okresie, co jest negatywne, ponieważ trend niedostatecznie objaśnia zjawisko niespłacalności kredytów.

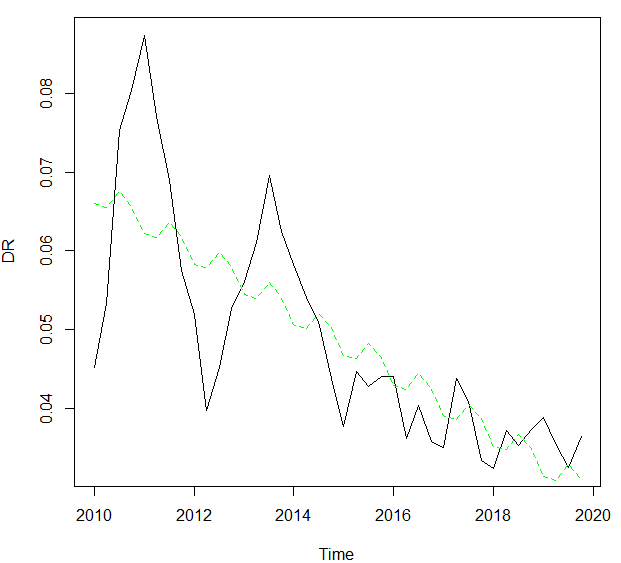
1. Trend i sezonowość

Obraz zawierający stół

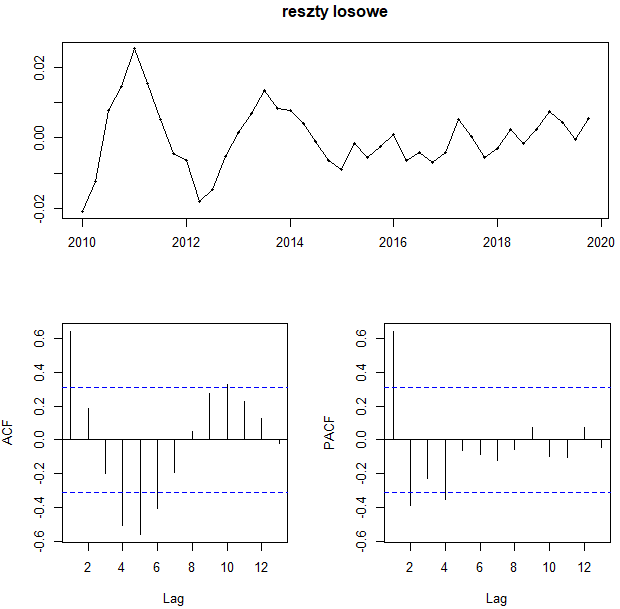
Opis wygenerowany automatycznie

Z uwagi na p-value <= 5% uznajemy, że zmienna mówiąca o trendzie jest istotna dla modelu regresji. P-value dla zmiennej mówiącej o sezonowości p-value > 5%, zatem nie jest ona istotna, co jest kolejnym potwierdzeniem braku sezonowości w modelu.



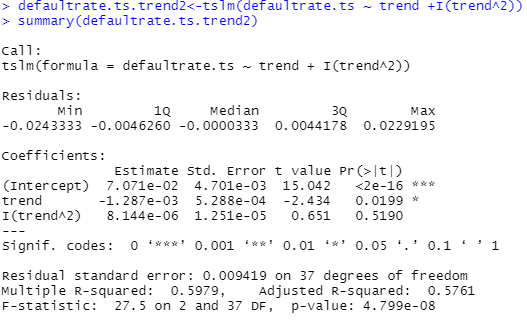




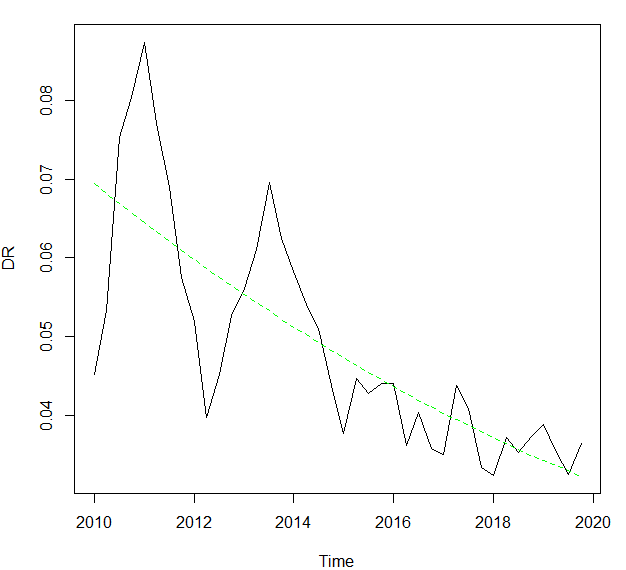


Porównanie 1 i 2 – model z trendem jest lepszy niż z trendem i sezonowością – wartość skorygowanego R^2 jest wyższa. Reszty w obu przypadkach nie są stacjonarne

1. Trend 2 stopnia





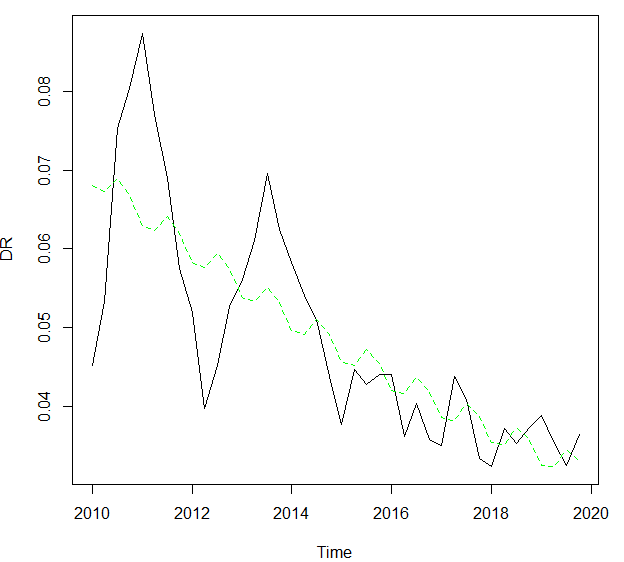


Trend wielomianowy drugiego stopnia nie dopasowuje się do bazowego szeregu danych ale z wartości współczynnika determinacji wynika, że jest on lepiej dopasowany do szeregu, niż trend z sezonowością.

1. Trend 2 stopnia i sezonowość

Obraz zawierający tekst

Opis wygenerowany automatycznie



Porównanie 3 i 4, ponownie model z trendem ma wyższą wartość skorygowanego R^2. Z powodu, że model z sezonowością pogarsza skorygowany R^2, dalej analizujemy sam trend, aby dopasować go lepiej do modelu.

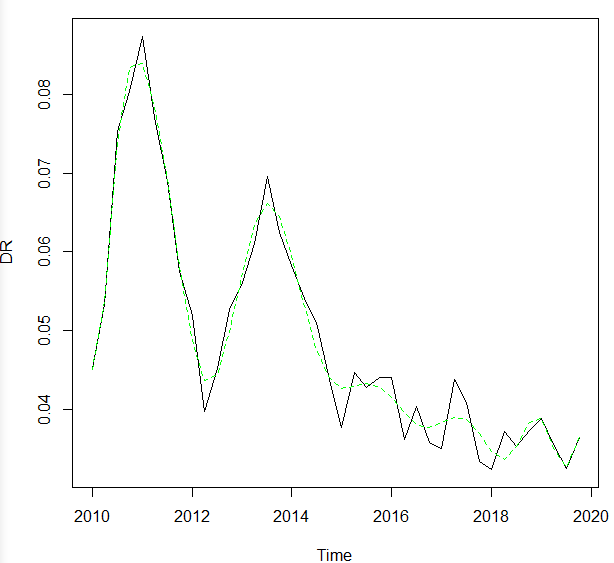
Trend niewystarczająco dopasował się do szeregu, dlatego próbujemy zwiększyć stopień wielomianu.

1. Trend 18 stopnia



Obraz zawierający stół

Opis wygenerowany automatycznie

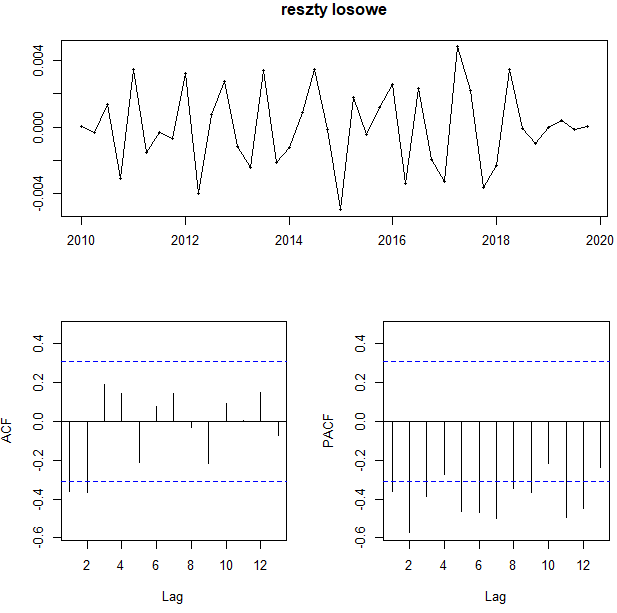


1. defaultrate.ts~ trend (R^2 =0,5825)
2. defaultrate.ts ~ trend ~sezonowość (R^2=0,5576)
3. defaultrate.ts ~ trend2, (R^2=0,5761)
4. defaultrate.ts ~ trend2 ~ sezonowość (R^2= 0,5499)
5. defaultrate.ts ~ trend18 (R^2=9552)

Najbardziej dopasowany model to ten zawierający trend 18 stopnia, ponieważ ma najwyższą wartość skorygowanego R^2.

Decydujemy się na zbadanie reszt losowych.





Reszty są niestacjonarne, więc trend osiemnastego stopnia nie wyjaśnia dobrze zjawiska niespłacalności kredytów.

**Sprawdzenie kryteriów informacyjnych:**

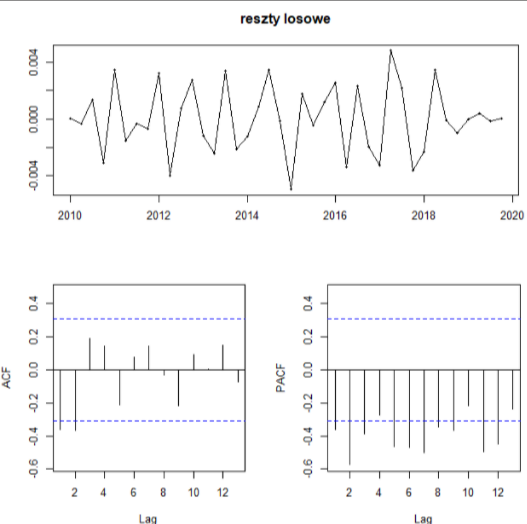
Wykorzystanie kryteriów informacyjnych do określenia dopasowania trendu. Z tej analizy wynika, że modele z sezonowością są gorsze niż bez sezonowości. Najlepszy model to ten z trendem stopnia 18, ponieważ ma najniższą wartość kryteriów informacyjnych i najwyższą wartość współczynnika determinacji. Powyższa analiza potwierdziła, że mamy do czynienia z wielomianowym trendem.

Obraz zawierający tekst

Opis wygenerowany automatycznie

**Badanie stacjonarności reszt trendu 18 stopnia**





Na powyższym wykresie widać niestacjonarność reszt w modelu trendu 18 stopnia, co dyskwalifikuje go w dalszej analizie.

**Budowanie modeli ARIMA:**



Obraz zawierający tekst

Opis wygenerowany automatycznie

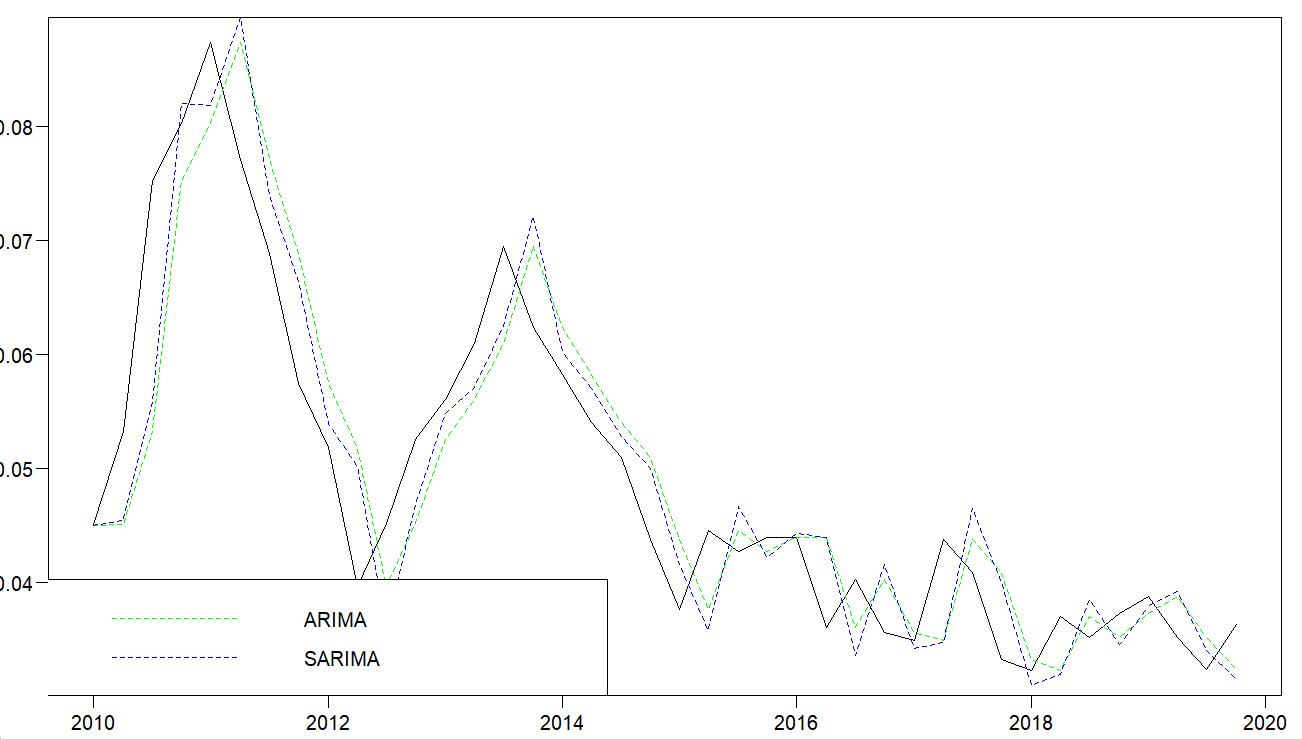
Funkcja z automatycznie dobieraną parametryzacją na podstawie kryteriów informacyjnych proponuje powyższy dobór opóźnień. Uwzględniła ona w modelu sezonowość, dlatego w późniejszych analizach zdecydowałyśmy się dodatkowo zbadać model z uwzględnieniem sezonowości, aby dobrać jak najlepszy model.

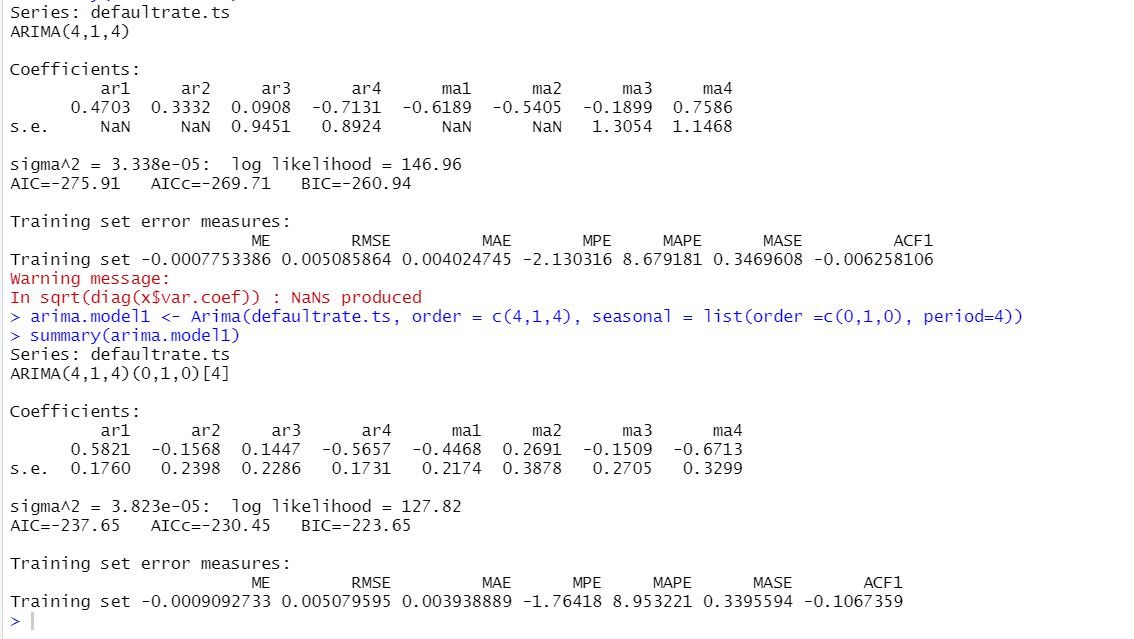
**Analiza ręcznych modeli ARIMA i SARIMA:**  
Na podstawie wykresów ACF/PACF wybieramy wartości do ręcznego modelu ARIMA i SARIMA. Wykres ACF odpowiada za liczbę istotnych opóźnień q, w modelu MA. Wykres PACF odpowiada za liczbę istotnych opóźnień p, w modelu AR. Z tego powodu w modelu ARIMA wybierzemy parametryzację (4,1,4).

Pomimo wykluczenia sezonowości w poprzednim etapie, decydujemy się na jej uwzględnienie w modelu, ze względu na chęć wykonania jak najdokładniejszej analizy.

Obraz zawierający tekst

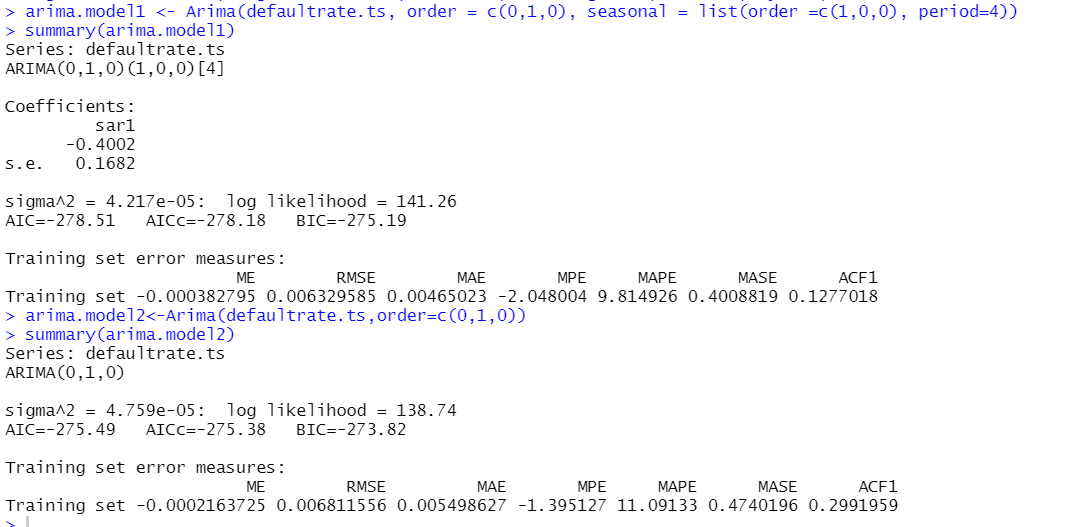
Opis wygenerowany automatycznie

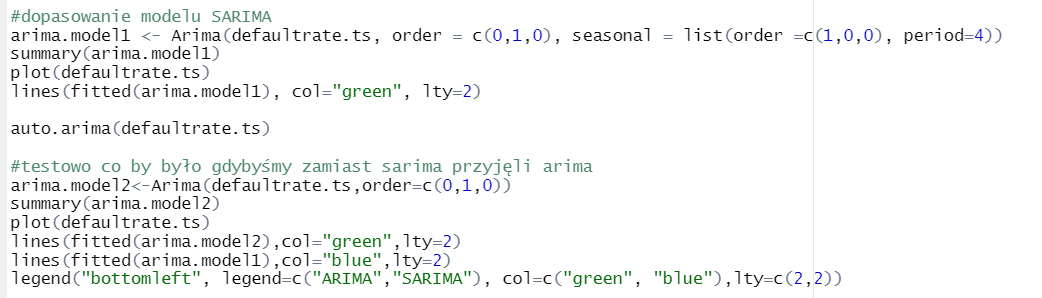


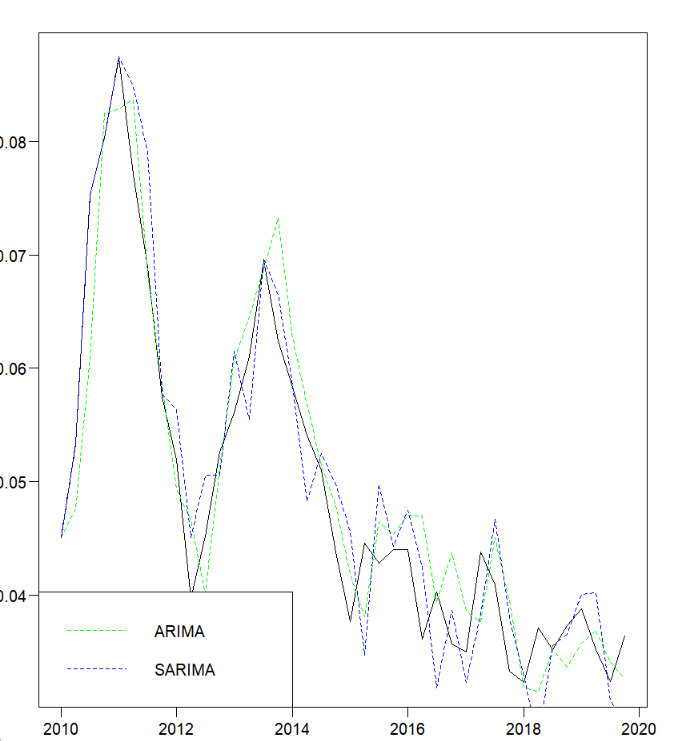


Jeżeli większość błędów przyjmuje niższe wartości w danym modelu, to oceniamy go jako lepszy model od modelu, w którym błędy w większości przyjmują wyższe wartości. Model z sezonowością okazuje się być lepszy, ze względu na niższe błędy.

**Modele na podstawie auto ARIMA:**

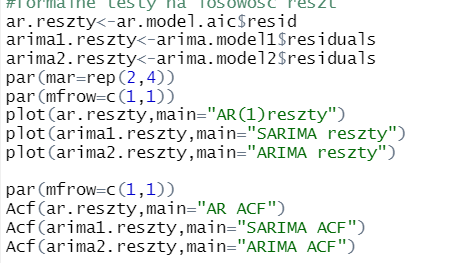


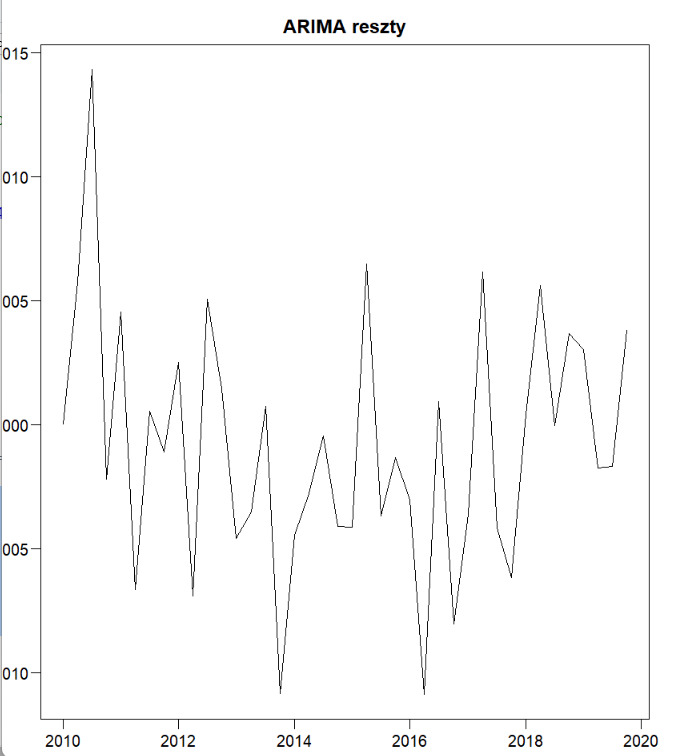


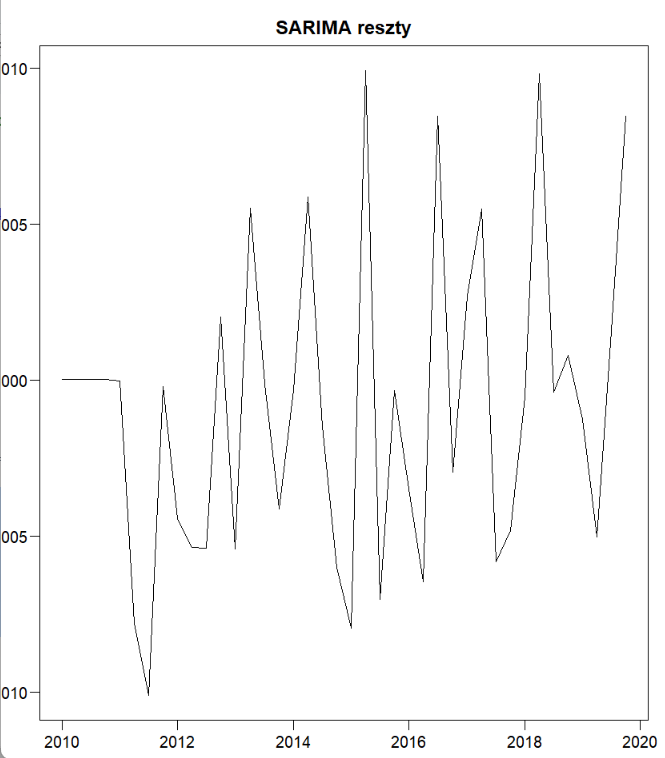


Również w przypadku automatycznego doboru opóźnień model z sezonowością wypada lepiej niż bez niej. Jednak porównując model automatyczny z tym dobranym ręcznie wnioskujemy, że modele utworzone ręcznie są lepiej dopasowane, ze względu na niższą wartość błędów. Model (4,1,4) plus sezonowość wypada lepiej niż model z automatycznie dobieraną parametryzacją na podstawie kryteriów informacyjnych.

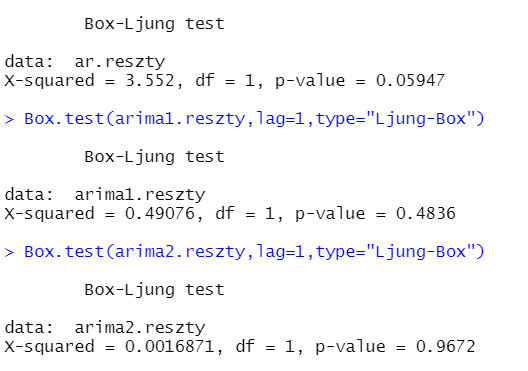
Badanie reszt modeli:







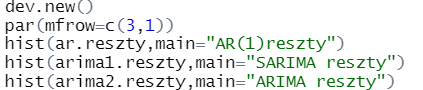


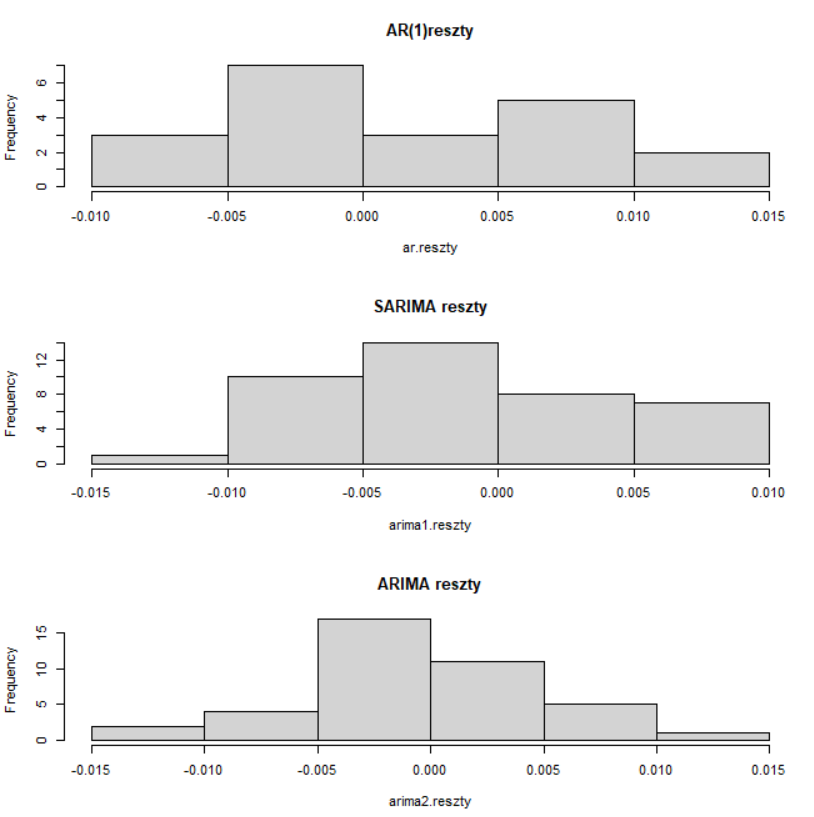


Test Ljunga-Boxa:

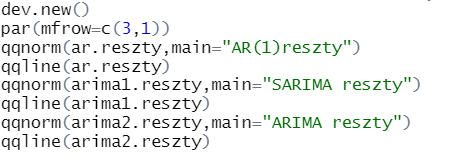
H0: brak autokorelacji.

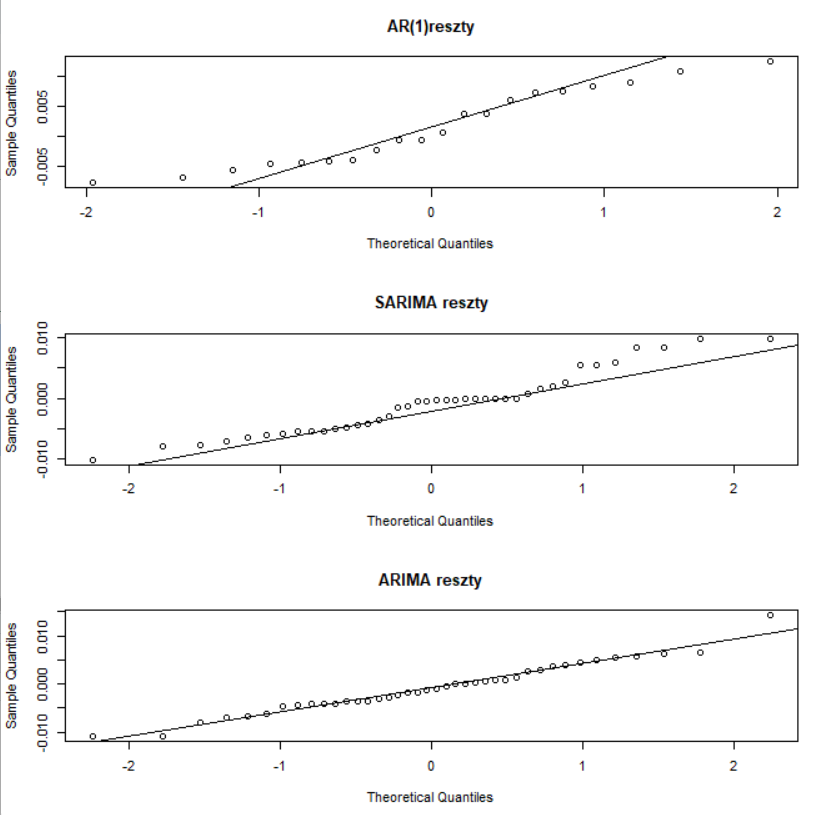
We wszystkich modelach p-value jest większe niż 5%, stąd brak podstaw do odrzucenia hipotezy zerowej, co oznacza brak autokorelacji, a tym samym stacjonarność reszt.





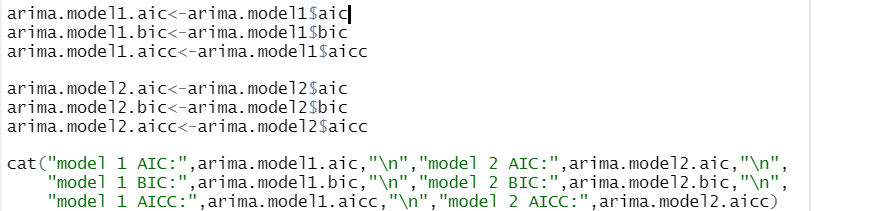
Histogramy reszt - wykazują, że rozkład reszt najbardziej zbliżony do normalnego mają modele SARIMA i ARIMA.

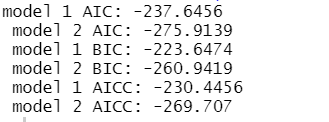




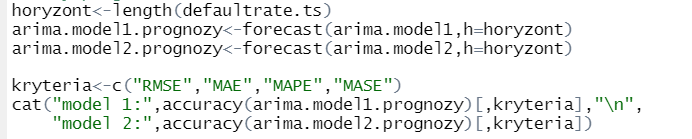
Funkcja qqline dodaje teoretyczną linię zgodnie z rozkładem normalnym, na podstawie czego można stwierdzić, że rozkład najbardziej zbliżony do normalnego ma model ARIMA oraz SARIMA.

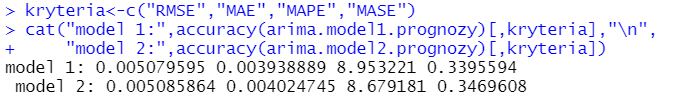
**Porównanie modelu ARIMA i SARIMA:**



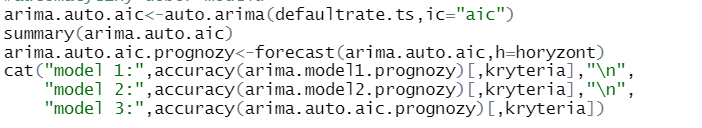


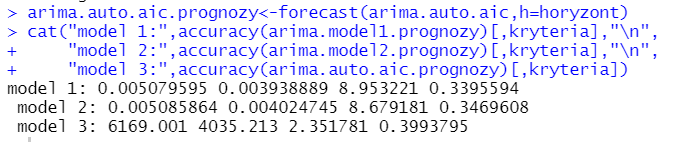
Niższe wartości kryteriów informacyjnych są w modelu 2 - ARIMA, więc jest to lepszy model niż model pierwszy.





Model 1 (SARIMA) ma niższe wartości błędów - jest lepszy niż model drugi. Jest to sprzeczne z wnioskami uzyskanymi po analizie kryteriów informacyjnych.





Na podstawie powyższych wyników nie można stwierdzić, czy model z sezonowością, czy bez niej będzie lepszy do prognozowania. Decydujemy się dodać do modeli zmienne makroekonomiczne i wtedy porównać błędy prognoz.

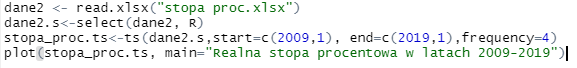
**Dodajemy zmienne makro:**

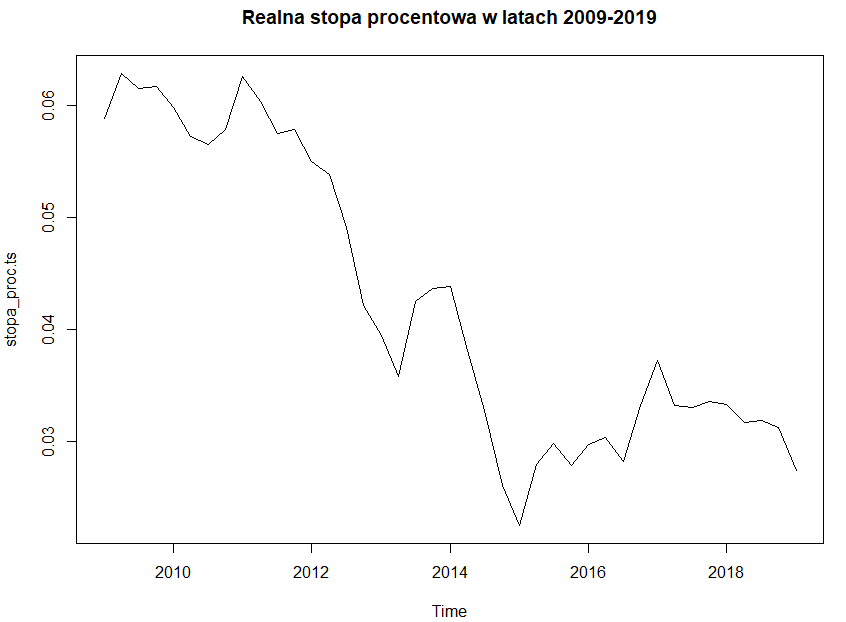
Dane makroekonomiczne uwzględniają okres wcześniejszy o 4 kwartały - rok w porównaniu z szeregiem bazowym *defaultrate.ts*. Dodatkowe zmienne, które zostaną wprowadzone to:

* stopa procentowa
* stopa wzrostu PKB
* stopa inflacji
* stopa bezrobocia

Analizujemy każdą zmienną pojedynczo, żeby zbadać jej stacjonarność. Jeżeli zmienna okaże się stacjonarna, wtedy dodajemy ją do modelu. Zmienne do modelu dodajemy stopniowo, aby porównać jakość modeli z poszczególnymi zmiennymi. Stacjonarność oceniana jest na podstawie funkcji ACF I PACF oraz testów na stacjonarność ADF I KPSS. W przypadku gdy testy pokazują sprzeczne wyniki, to używamy testu PP i sugerujemy się wynikiem, który przeważa.

**Analizujemy pierwszą zmienną:**

****





Obraz zawierający tekst

Opis wygenerowany automatycznie



Obraz zawierający tekst

Opis wygenerowany automatycznie

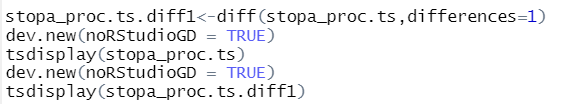
ADF: H0-szereg niestacjonarny.

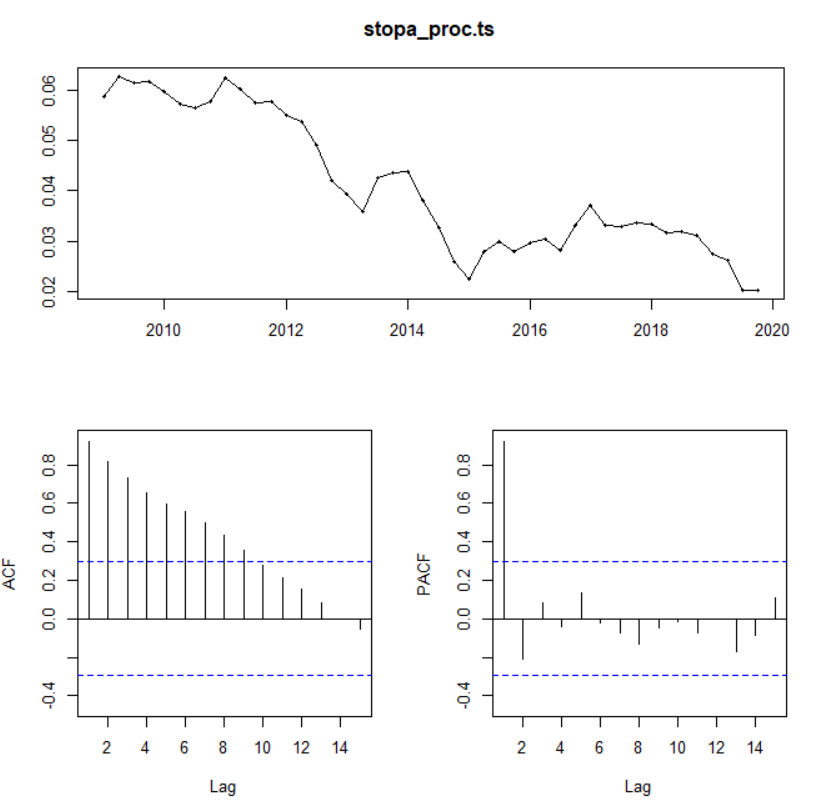
p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

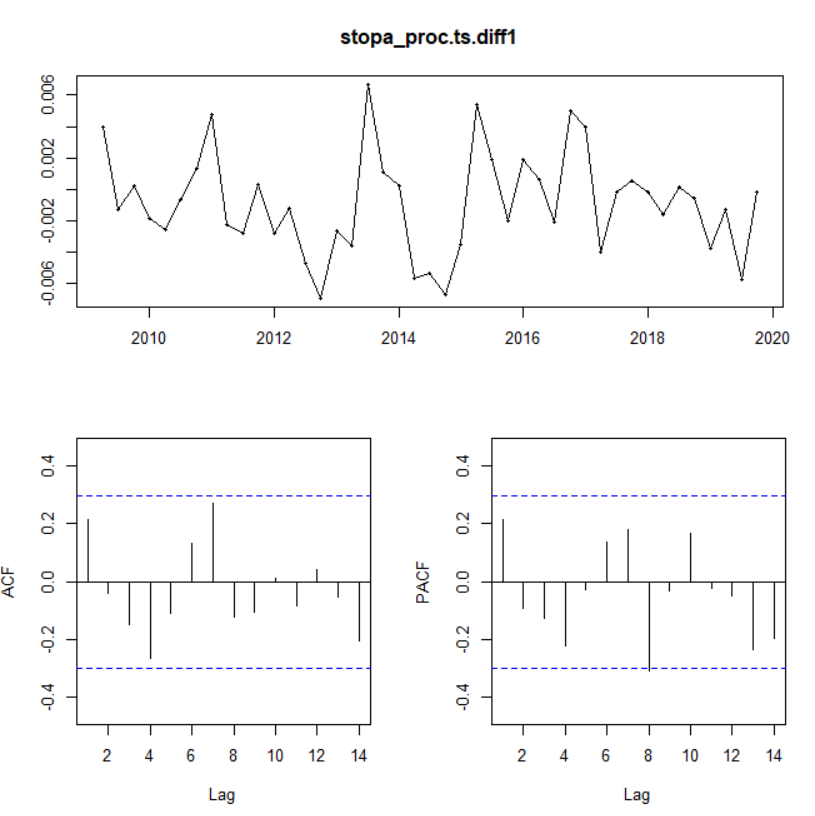
KPSS: H0-szereg stacjonarny.

p-value jest mniejsze niż 5%, co daje podstawy do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

Niestacjonarność szeregu oznacza, że wymaga on różnicowania, w celu doprowadzenia go do stacjonarności.

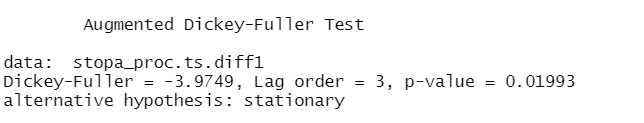


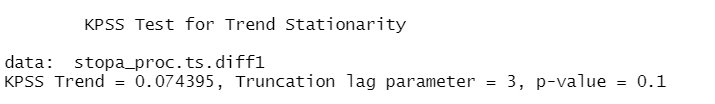




Obraz zawierający tekst

Opis wygenerowany automatycznie





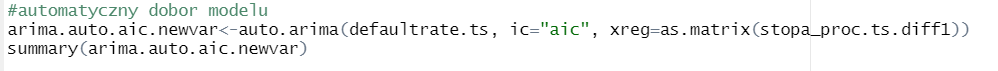
ADF: H0-szereg niestacjonarny.

p-value jest mniejsze niż 5%, co pozwala na odrzucenie hipotezy zerowej, więc szereg jest stacjonarny.

KPSS: H0-szereg stacjonarny.

p-value jest większe niż 5%, co daje brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest stacjonarny.

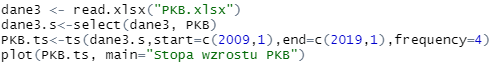
Jednokrotne różnicowanie sprawiło, że szereg jest stacjonarny i można go dalej rozważać w modelu. Model z tą zmienną nazywamy *arima.auto.aic.newvar.*

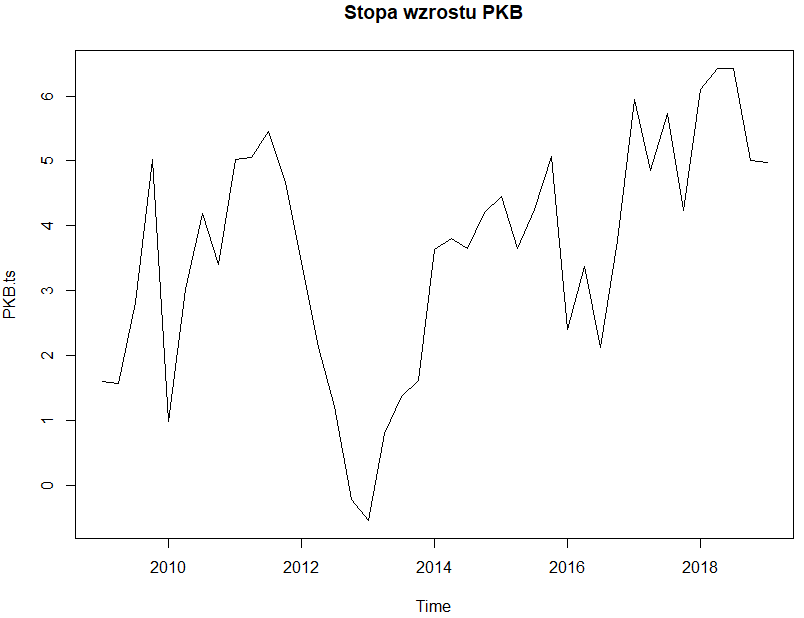


Obraz zawierający tekst

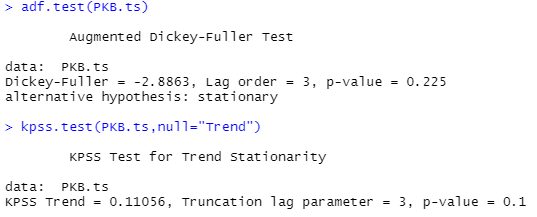
Opis wygenerowany automatycznie

**Analizujemy drugą zmienną jaką jest PKB:**





Sprawdzamy czy zmienna jest stacjonarna:



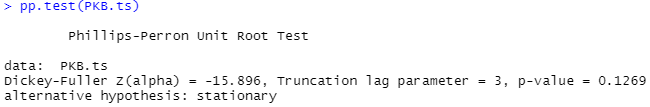
ADF: H0-szereg niestacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

KPSS: H0-szereg stacjonarny.

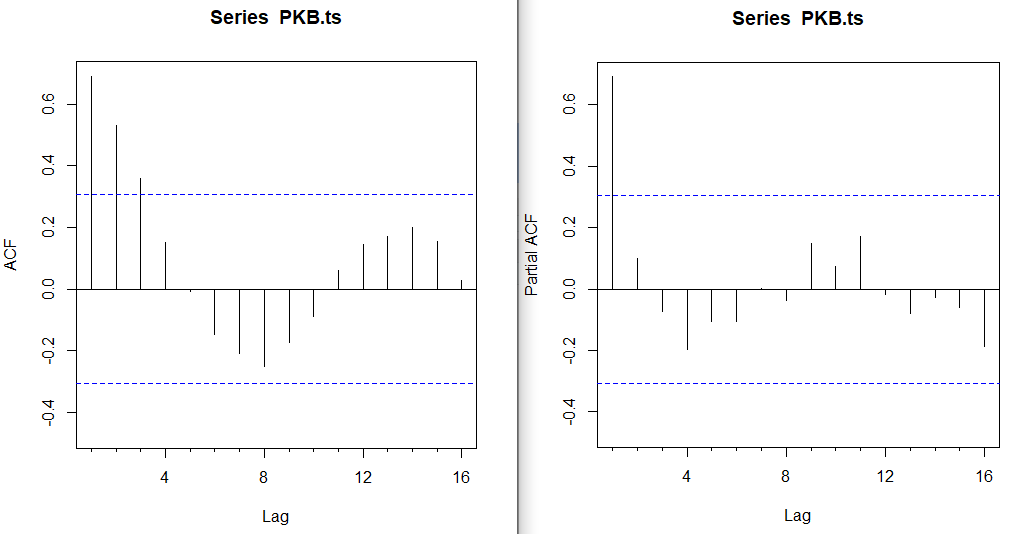
p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest stacjonarny.

Dwa z przeprowadzonych testów wskazują na sprzeczne obserwacje. Przeprowadzamy trzeci test (PP) oraz analizę ACF i PACF.

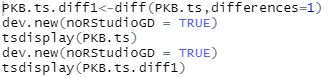


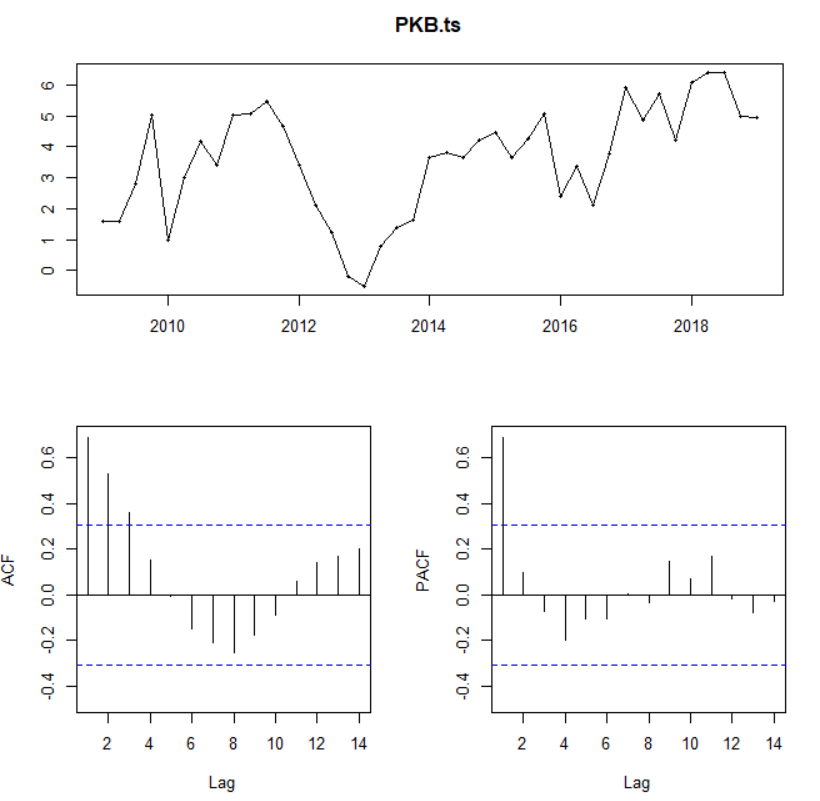
Test PP: H0-szereg niestacjonarny

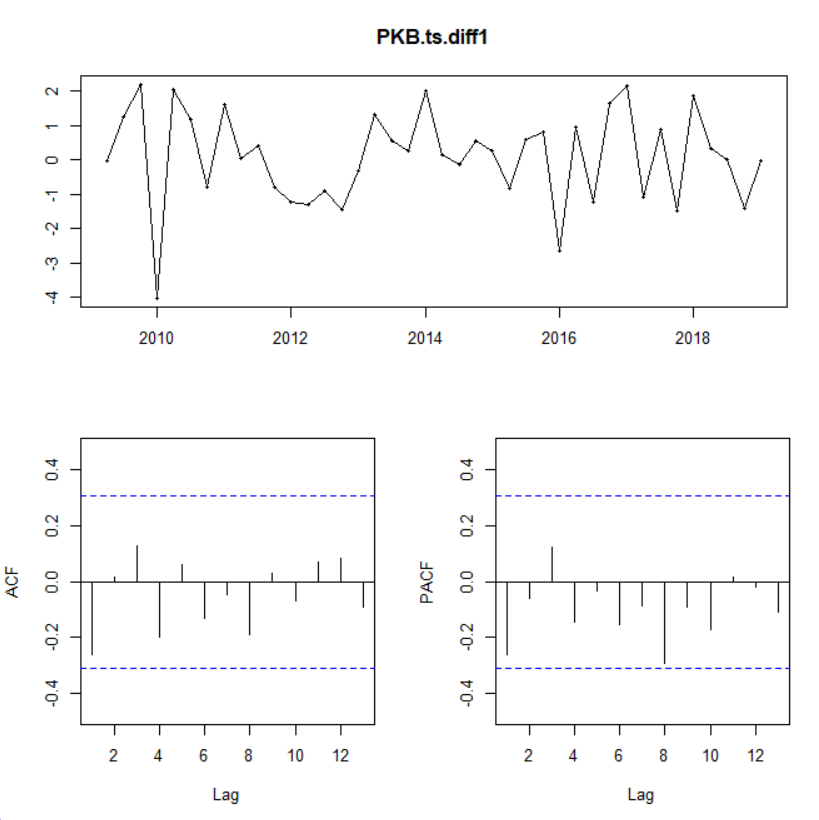
P-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, zatem szereg jest niestacjonarny.



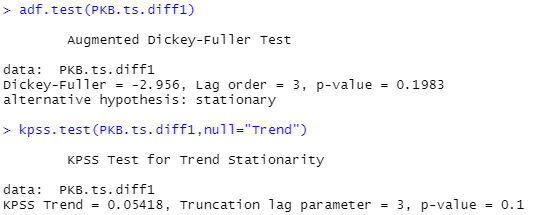
Analiza wykresu ACF i PACF również potwierdza niestacjonarność szeregu. Aby móc dalej pracować na tym szeregu należy go zróżnicować.







Jak wynika z wykresów ACF i PACF jednokrotne różnicowanie powoduje, że szereg jest stacjonarny. Sprawdzamy to testami.

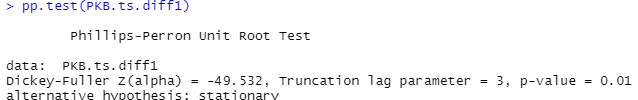


ADF: H0-szereg niestacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

KPSS: H0-szereg stacjonarny.

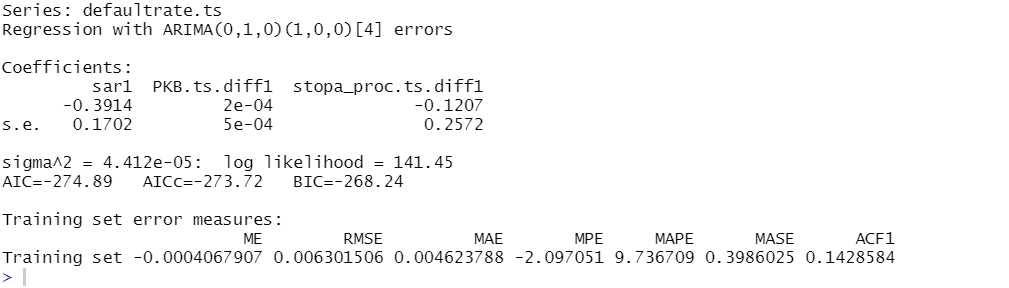
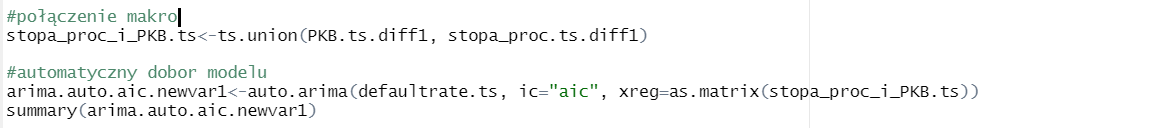
p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest stacjonarny



Test PP: H0-szereg niestacjonarny

P-value jest mniejsze niż 5%, co oznacza odrzucenie hipotezy zerowej, zatem szereg jest stacjonarny.

Wykres ACF/PACF oraz dwa z trzech testów wskazują na stacjonarność szeregu, zatem wnioskujemy, że szereg jest stacjonarny i dodajemy go do modelu. Model zawierający tą zmienną jak i zmienną analizowaną wcześniej nazywamy *arima.auto.aic.newvar1.*



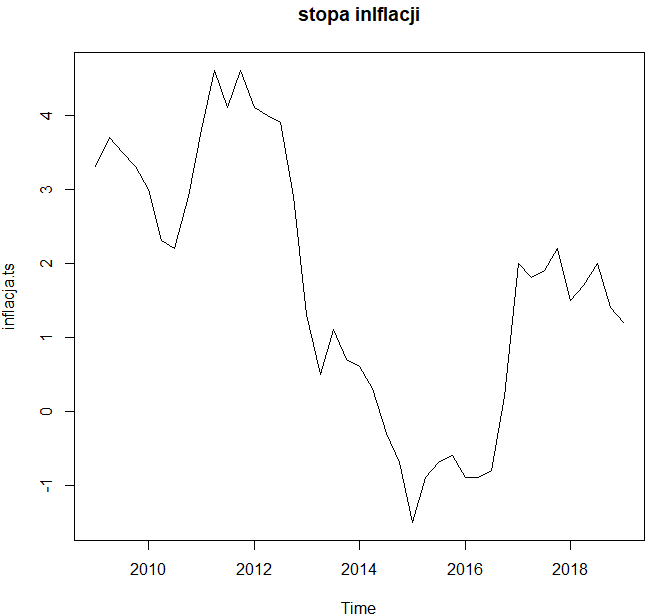
Porównanie wszystkich błędów modelu *arima.auto.aic.newvar* i *arima.auto.aic.newvar1:* Większość sprawdzonych błędów tzn. 5 z 7 wskazuje, że model ze stopą procentową jest lepszy niż model ze stopą procentową i PKB.

**Rozważamy trzecią zmienną – inflację:**

**Obraz zawierający tekst

Opis wygenerowany automatycznie**





Obraz zawierający tekst

Opis wygenerowany automatycznie

ADF: H0-szereg niestacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

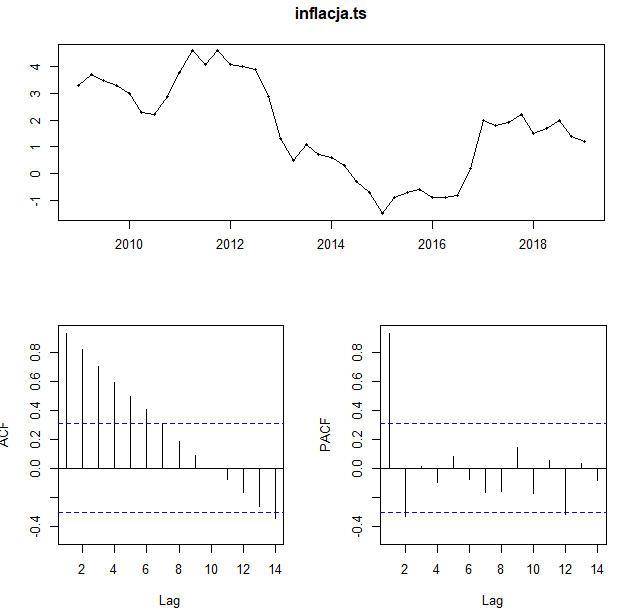
KPSS: H0-szereg stacjonarny.

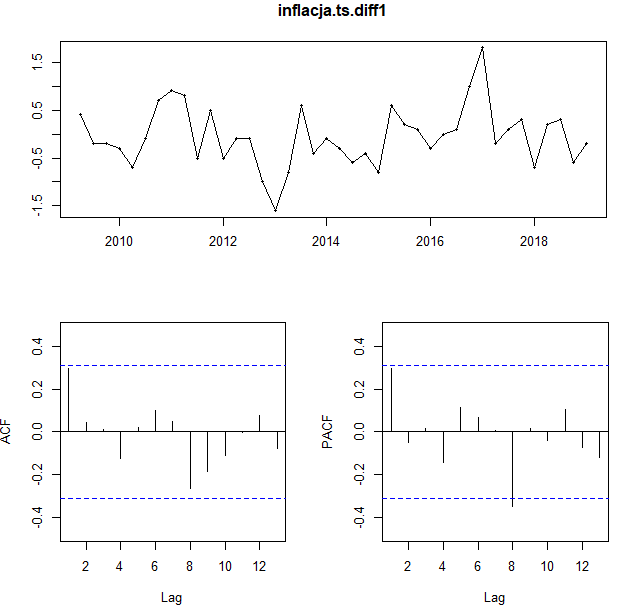
p-value jest mniejsze niż 5%, co oznacza że można odrzucić hipotezę zerową, więc szereg jest niestacjonarny

Należy zróżnicować szereg:

Obraz zawierający tekst

Opis wygenerowany automatycznie





Wykresy ACF/PACF delikatnie przekraczają wyznaczony przedział, zatem sprawdzamy ponownie testy.

Obraz zawierający tekst

Opis wygenerowany automatycznie

ADF: H0-szereg niestacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

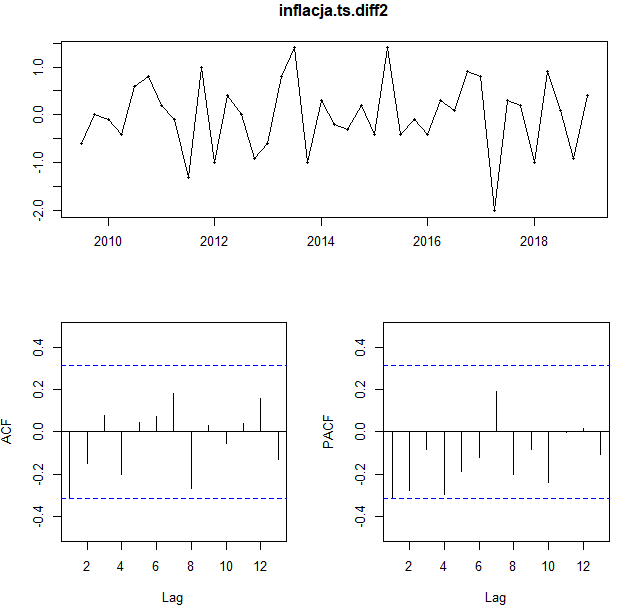
KPSS: H0-szereg stacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest stacjonarny

Testy też nie wskazują jednoznacznie, różnicujemy drugi raz.

Obraz zawierający tekst

Opis wygenerowany automatycznie



Obraz zawierający tekst

Opis wygenerowany automatycznie

Obraz zawierający tekst

Opis wygenerowany automatycznie

ADF: H0-szereg niestacjonarny.

p-value jest niższe niż 5%, co oznacza że można odrzucić hipotezę zerową, więc szereg jest stacjonarny.

KPSS: H0-szereg stacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest stacjonarny

Po drugim różnicowaniu szereg jest stacjonarny. Można to stwierdzić zarówno na podstawie testów jak i wykresów ACF/PACF.

Dodajemy szereg do modelu, model z trzema zmiennymi nazywamy *arima.auto.aic.newvar2*

Obraz zawierający tekst

Opis wygenerowany automatycznie

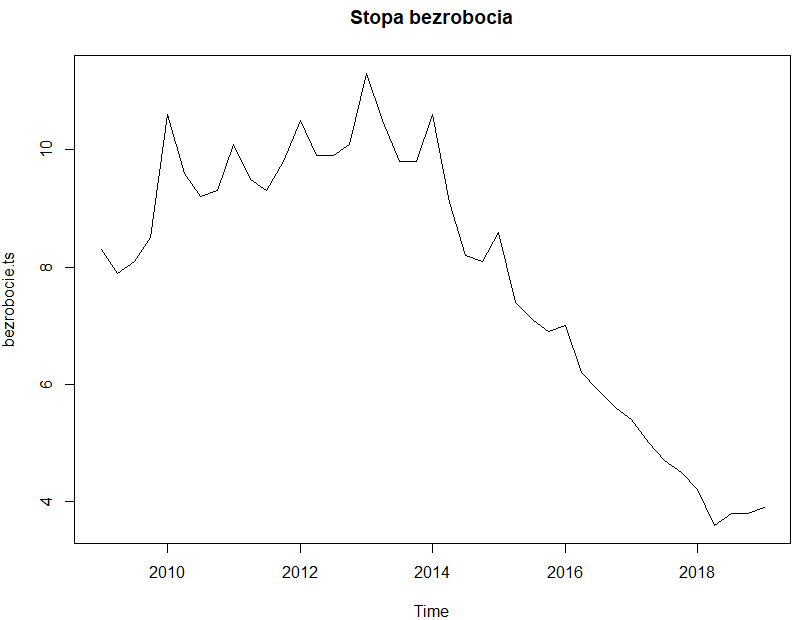
Wartości błędów tego modelu są najniższe na podstawie wszystkich błędów w porównaniu do modelu poprzedniego, ale także w porównaniu do modelu pierwszego. Ten model jest najlepszy jak do tej pory.

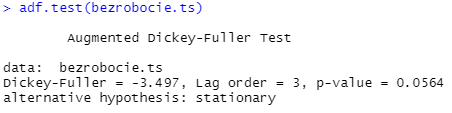
**Analiza czwartej zmiennej – bezrobocia:**

Obraz zawierający tekst

Opis wygenerowany automatycznie

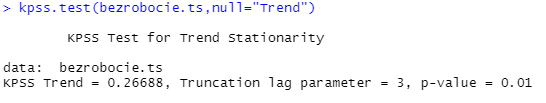






ADF: H0-szereg niestacjonarny.

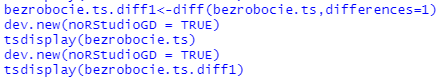
p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

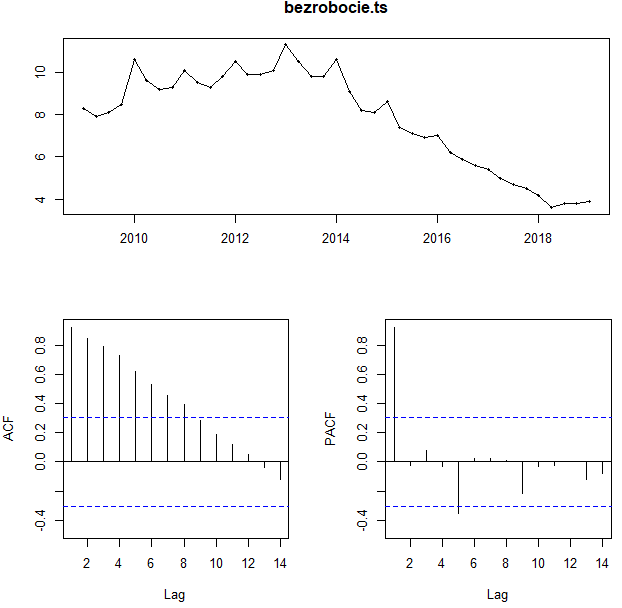


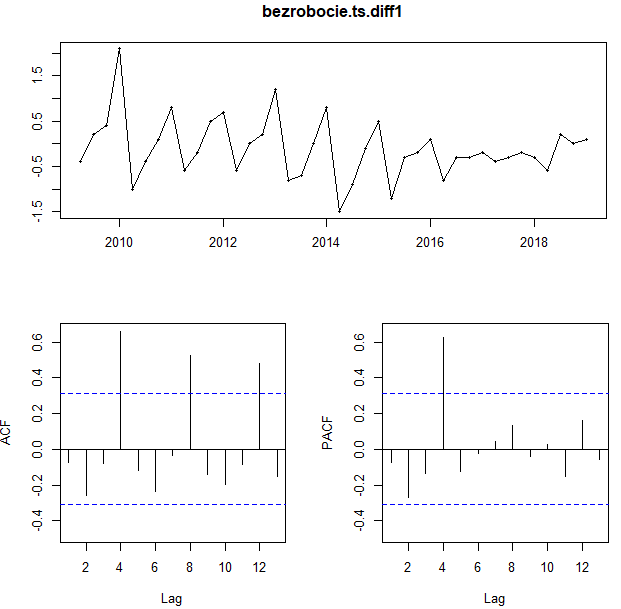
KPSS: H0-szereg stacjonarny.

p-value jest mniejsze niż 5%, co daje podstawy do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

Niestacjonarność szeregu oznacza, że wymaga on różnicowania, w celu doprowadzenia go do stacjonarności.



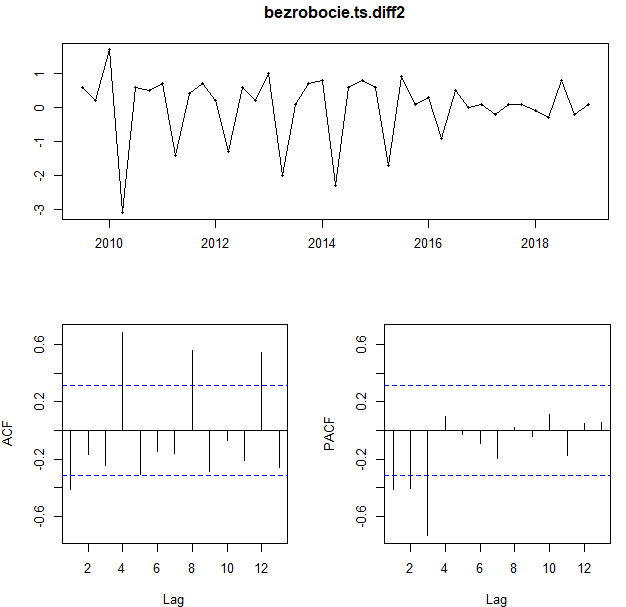




Pierwsze różnicowanie nie przynosi pożądanego efektu, można to wywnioskować na podstawie wykresów ACF/PACF.

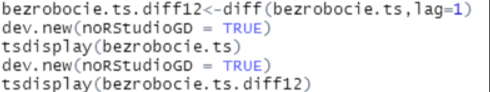
Obraz zawierający tekst

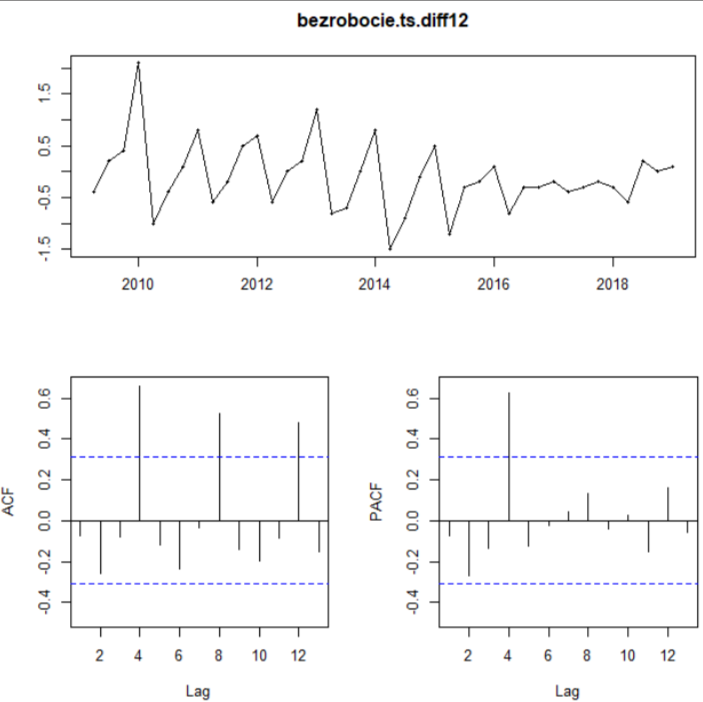
Opis wygenerowany automatycznie



Po drugim różnicowaniu szereg nadal jest niestacjonarny. Rezygnujemy z dalszego różnicowania sposobem różnicowania szeregu bazowego kolejnymi stopniami. Próbujemy innej metody różnicowania.

Wykonano różnicowanie na podstawie różnicowania szeregu, który wcześniej został zróżnicowany.

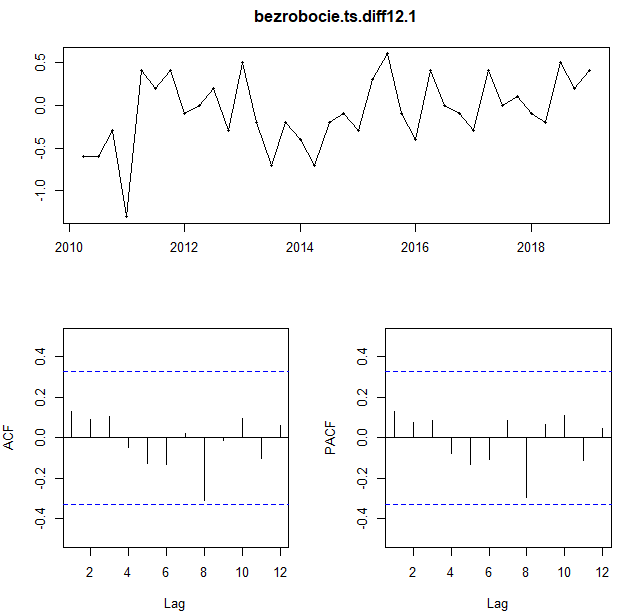




Z wykresu PACF wywnioskować można pewną sezonową zależność, dlatego można spróbować opóźnienia lag=4.

Obraz zawierający tekst

Opis wygenerowany automatycznie



Z wykresów ACF/PACF wynika, że szereg jest stacjonarny po pierwszym różnicowaniu.

Obraz zawierający tekst

Opis wygenerowany automatycznie

ADF: H0-szereg niestacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest niestacjonarny.

KPSS: H0-szereg stacjonarny.

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc szereg jest stacjonarny.

Przeprowadzamy dodatkowy test (PP):

Obraz zawierający tekst

Opis wygenerowany automatycznie

Test PP: H0-szereg niestacjonarny

P-value jest mniejsze niż 5%, co oznacza że odrzucamy hipotezę zerową, zatem szereg jest stacjonarny.

Wykres ACF/PACF oraz dwa z trzech przeprowadzonych testów potwierdzają stacjonarność szeregu. Na tej podstawie dodajemy bezrobocie i budujemy model – *arima.auto.aic.newvar3.*

Obraz zawierający tekst

Opis wygenerowany automatycznie

Model z bezrobociem ma tylko 3 (z 7) wartości błędów niższe niż model poprzedni. Model poprzedni - *arima.auto.aic.newvar2,* uwzględniający wpływ stopy procentowej, PKB oraz inflacji okazał się modelem najlepszym, błędy dla tego modelu okazały się najniższe.

**Prognozowanie:**

Arima.auto.aic.newvar3 to model ręczny z uwzględnieniem sezonowości i zmiennymi makroekonomicznymi , Arima.auto.aic.newvar4 to model ręczny bez sezonowości ze zmiennymi makroekonomicznymi, Arima.auto.aic.newvar2 to model automatyczny ze zmiennymi makroekonomicznymi. W skład zmiennych makroekonomicznych wchodzi inflacja, stopa procentowa i PKB.

**Obraz zawierający tekst

Opis wygenerowany automatycznie**

kryteria<-c("RMSE","MAE","MAPE","MASE")

Obraz zawierający tekst

Opis wygenerowany automatycznie

Model czwarty (model ręczny z sezonowością i zmiennymi makroekonomicznymi) ma najniższe wartości błędów RMSE, MAE i MASE, zatem jest on najlepszy.

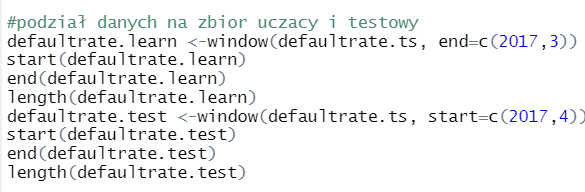
**Przeprowadzamy prognozowanie następującymi metodami:**

**Podział szeregu:**

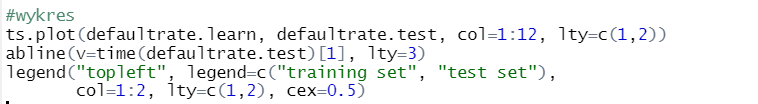
Długość bazowego szeregu:

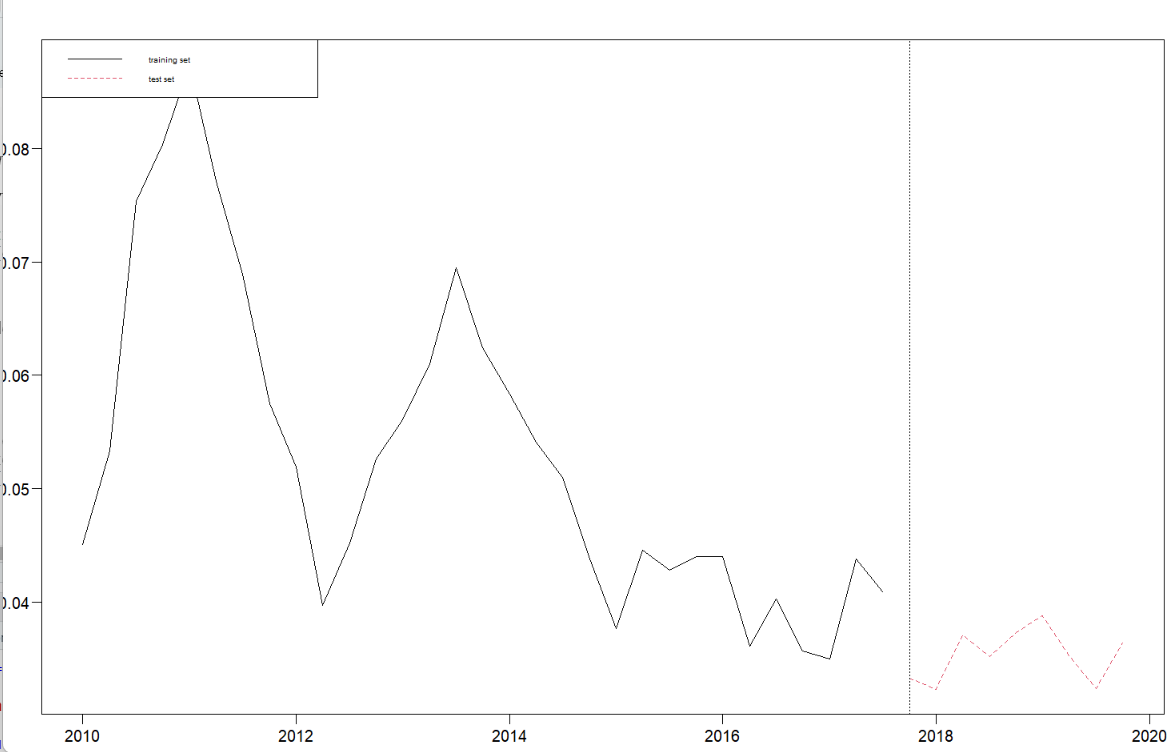


Podział szeregu na uczący i testowy:



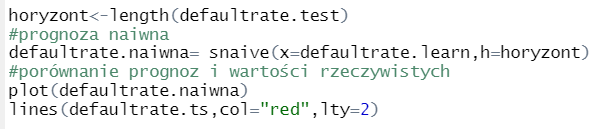
Zbiór testowy to dwa ostatnie lata bazowej obserwacji szeregu.

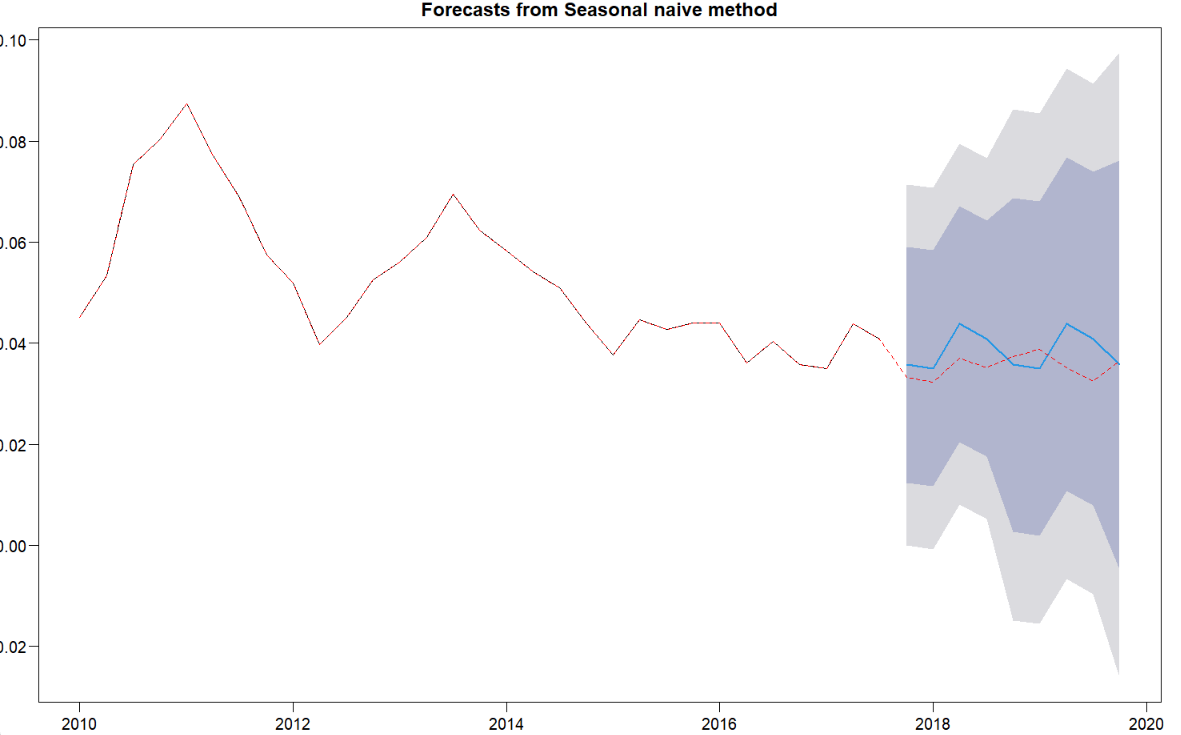




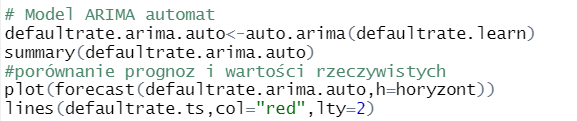
**Prognozy na szeregu testowym:**

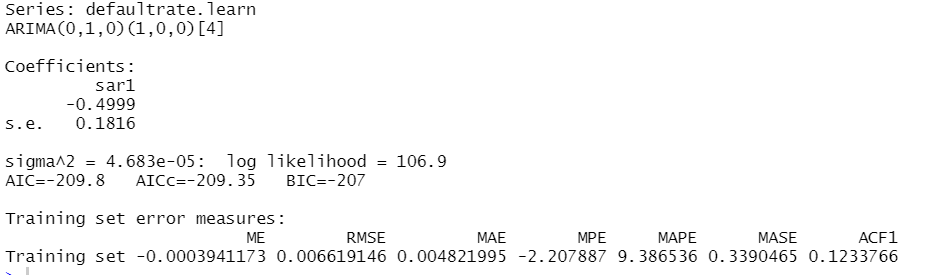
1. Metoda naiwna

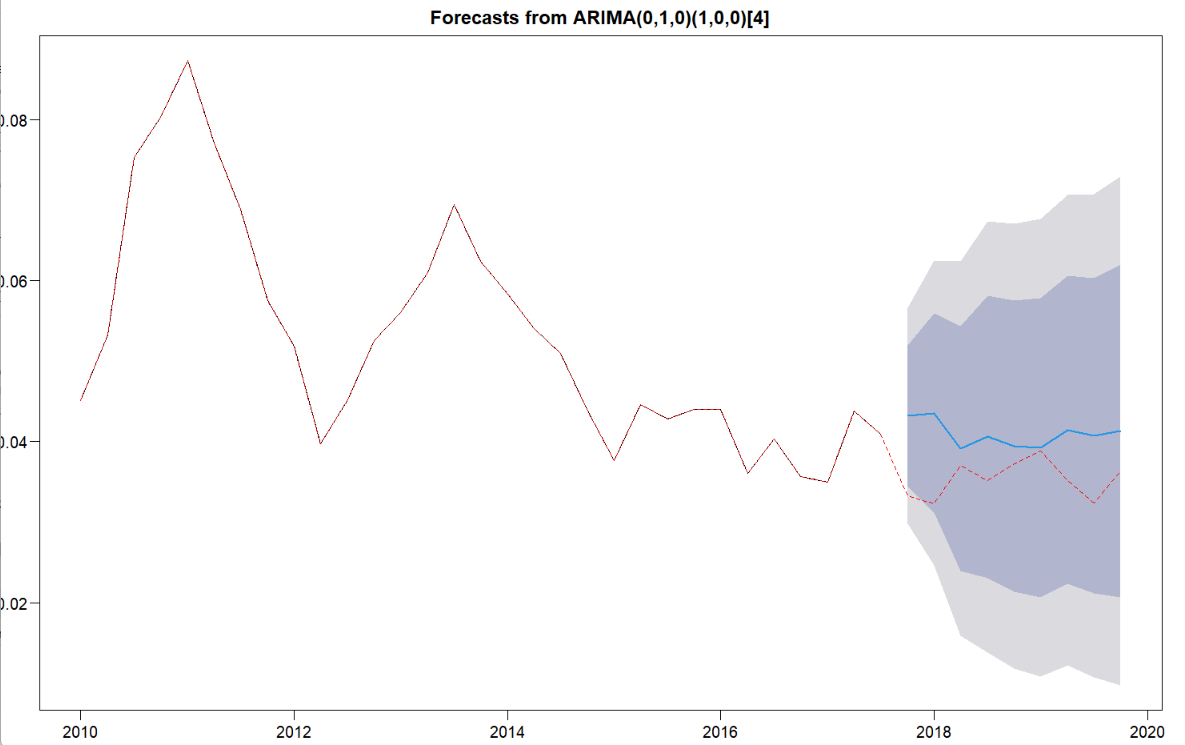


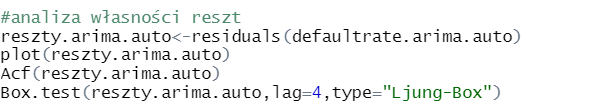


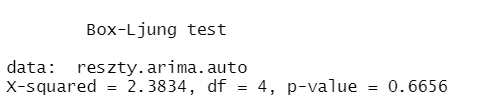
1. Auto arima na szeregu testowym







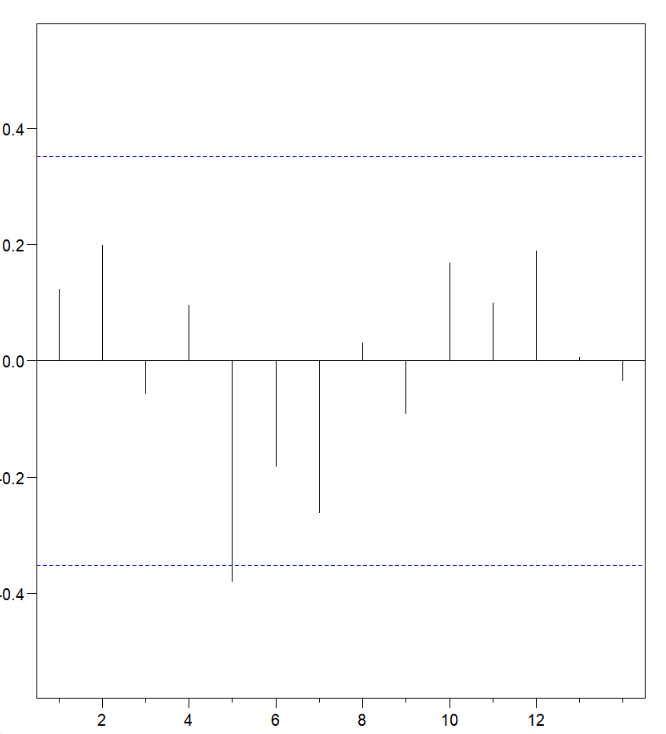




Test Ljung-Box

H0-brak autokorelacji (stacjonarność)

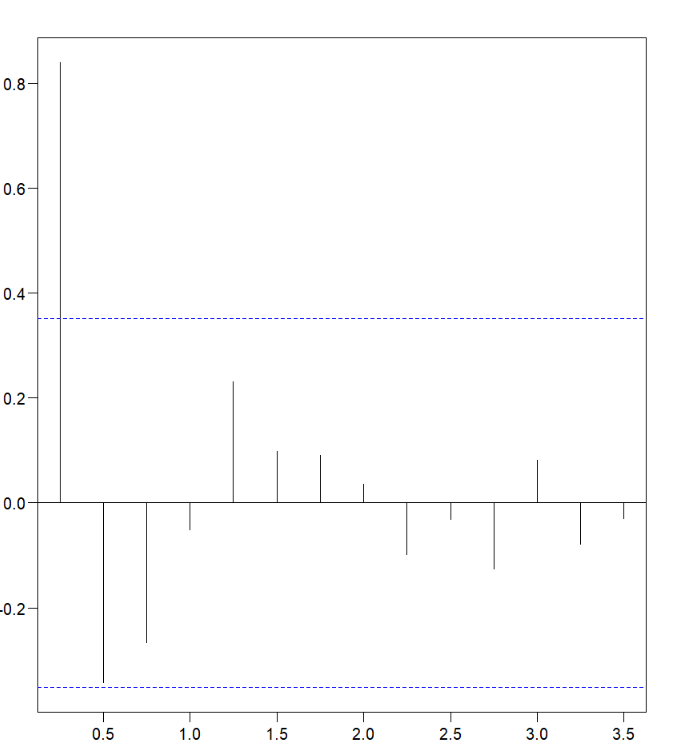
p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc nie występuje autokorelacja, a reszty są stacjonarne.



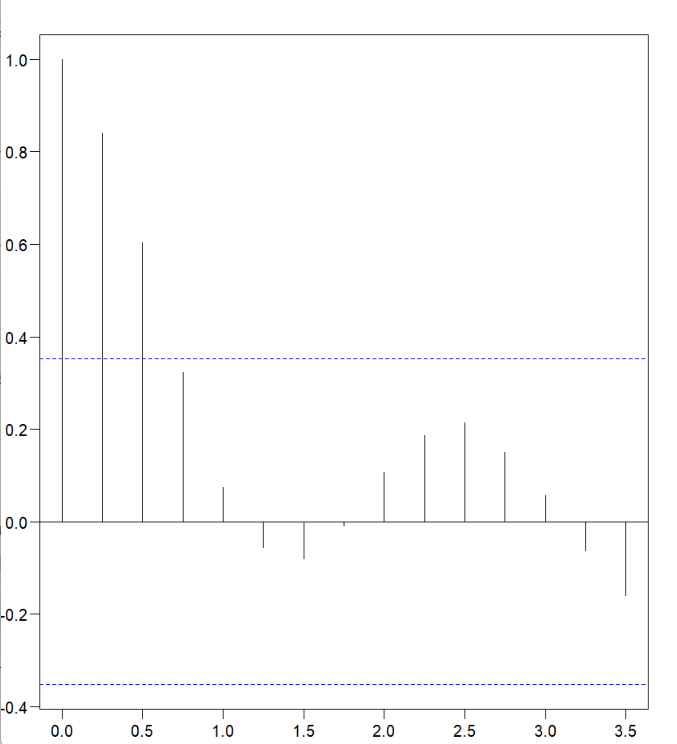
1. Arima ręczna na szeregu testowym

Aby dobrać opóźnienie badamy wykresy acf i pacf:

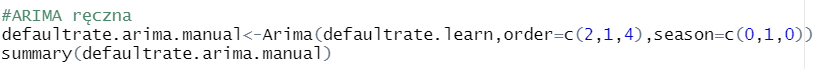


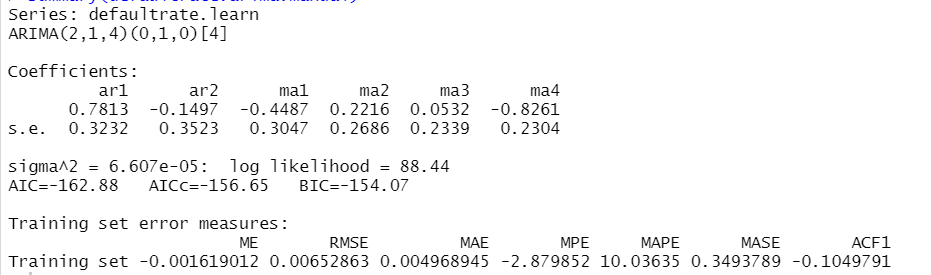




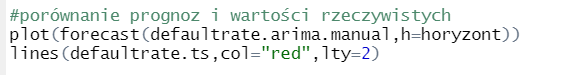


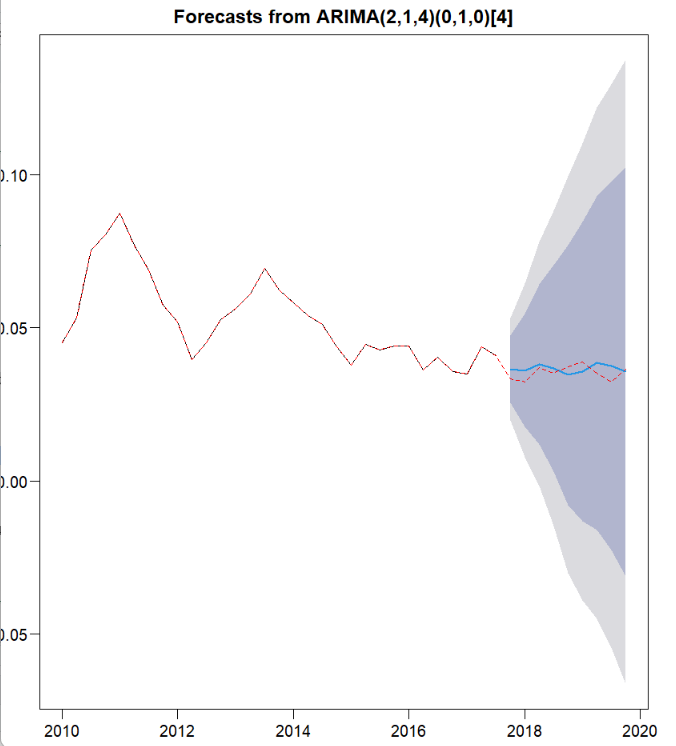
Na podstawie “pików” dobieramy opóźnienia w modelu.



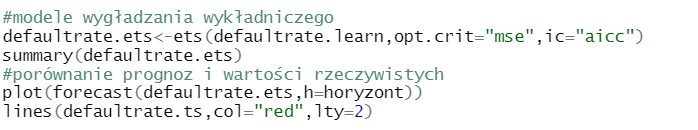


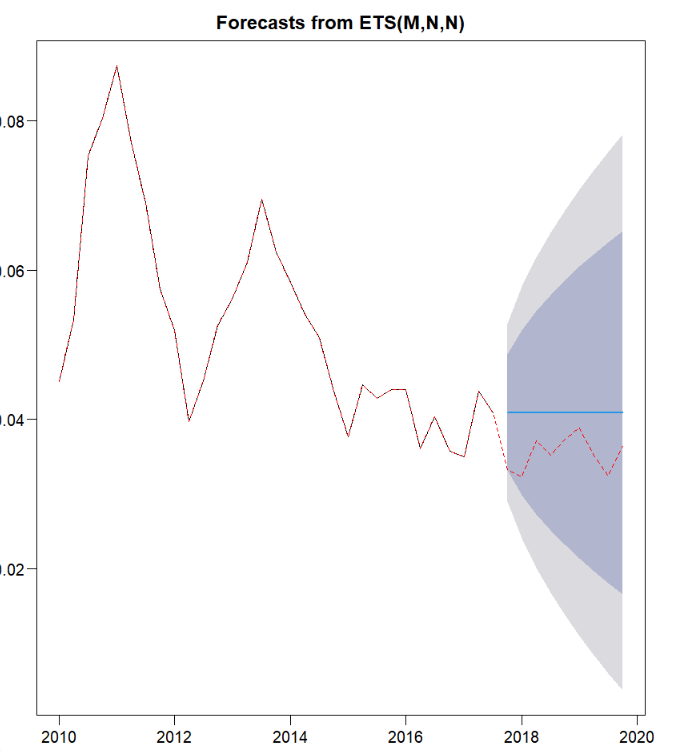
Po porównaniu błędów uważamy, że model defaultrate.arima.manual jest lepszym modelem od modelu automatycznego.

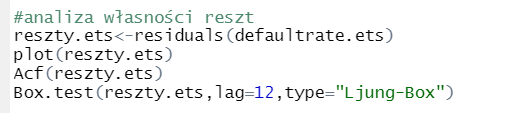


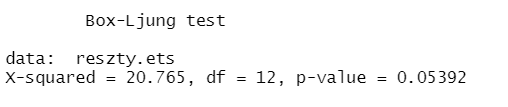


1. Wygładzanie wykładnicze:





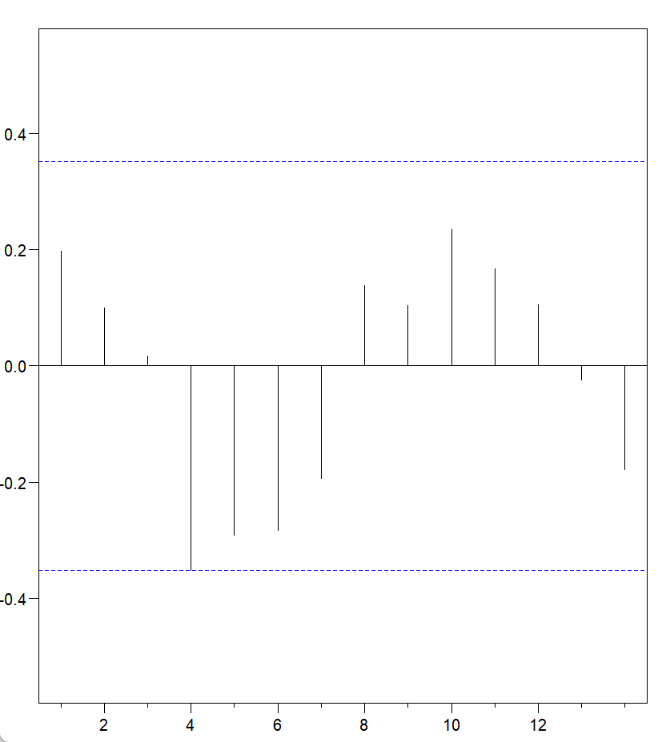




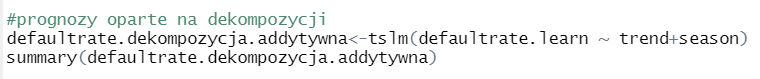
Test Ljung-Box

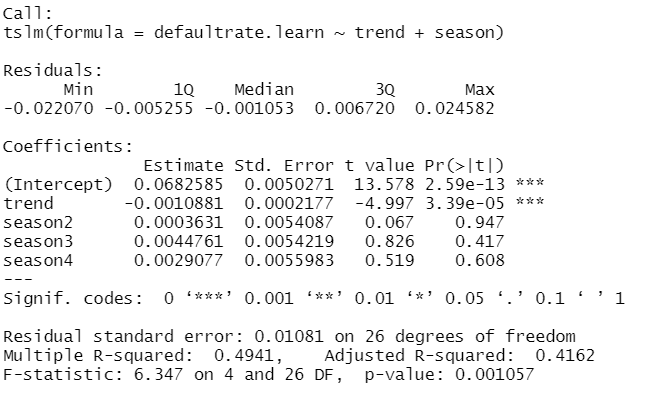
H0-brak autokorelacji (stacjonarność)

p-value jest większe niż 5%, co oznacza brak podstaw do odrzucenia hipotezy zerowej, więc nie występuje autokorelacja, a reszty są stacjonarne

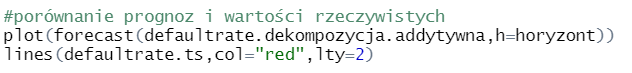


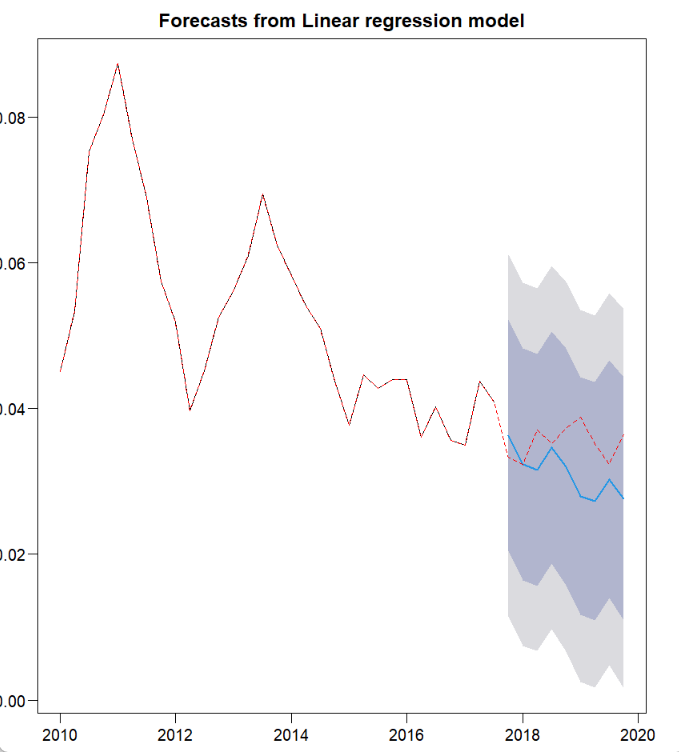
1. Dekompozycja addytywna:

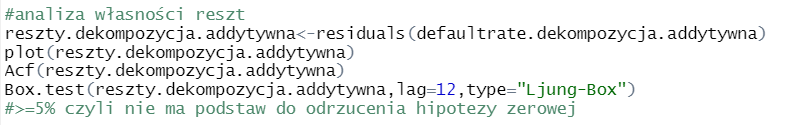


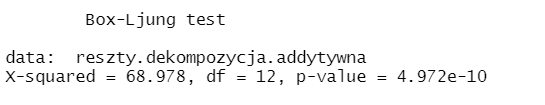


Na podstawie powyższych wyników wnioskujemy, że trend jest istotny statystycznie w tym modelu.

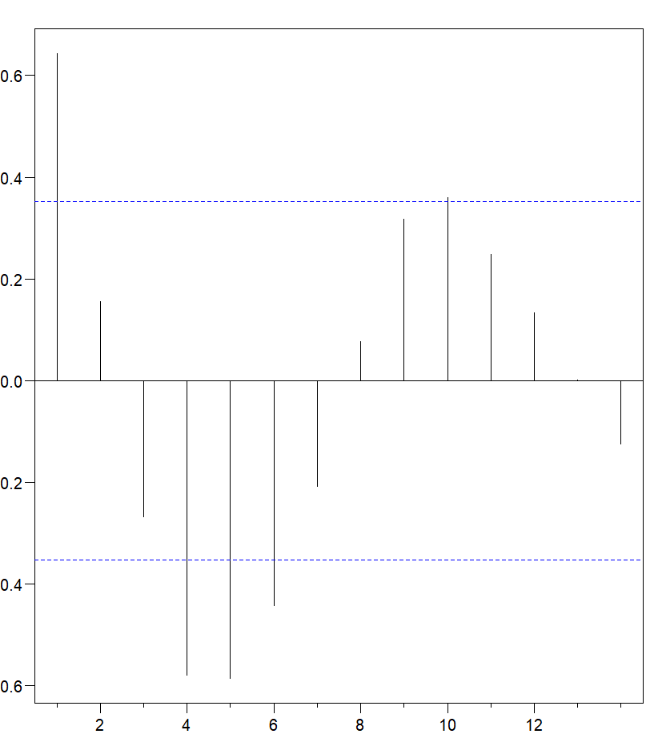








Reszty w tym modelu są niestacjonarne.



**Prognozy na modelu ręcznym z sezonowością i ze zmiennymi makroekonomicznymi**

**Obraz zawierający tekst

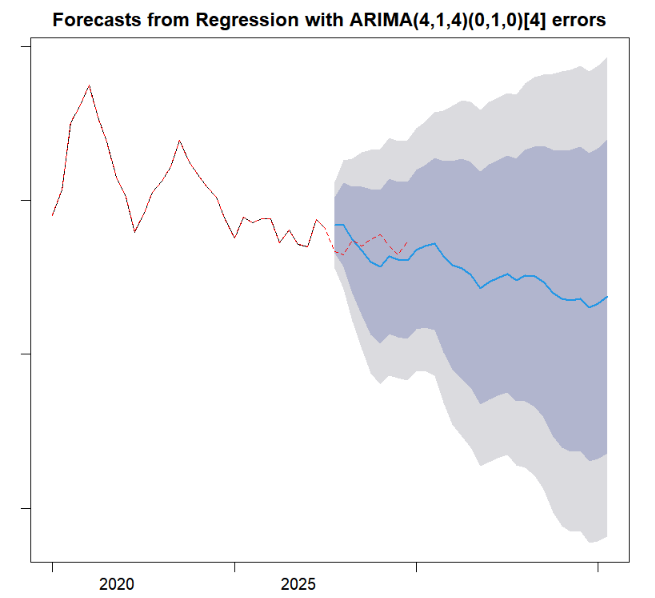
Opis wygenerowany automatycznie**

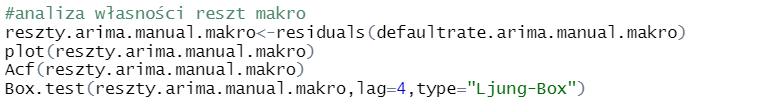
**Obraz zawierający tekst, paragon, zrzut ekranu

Opis wygenerowany automatycznie**

Obraz zawierający tekst

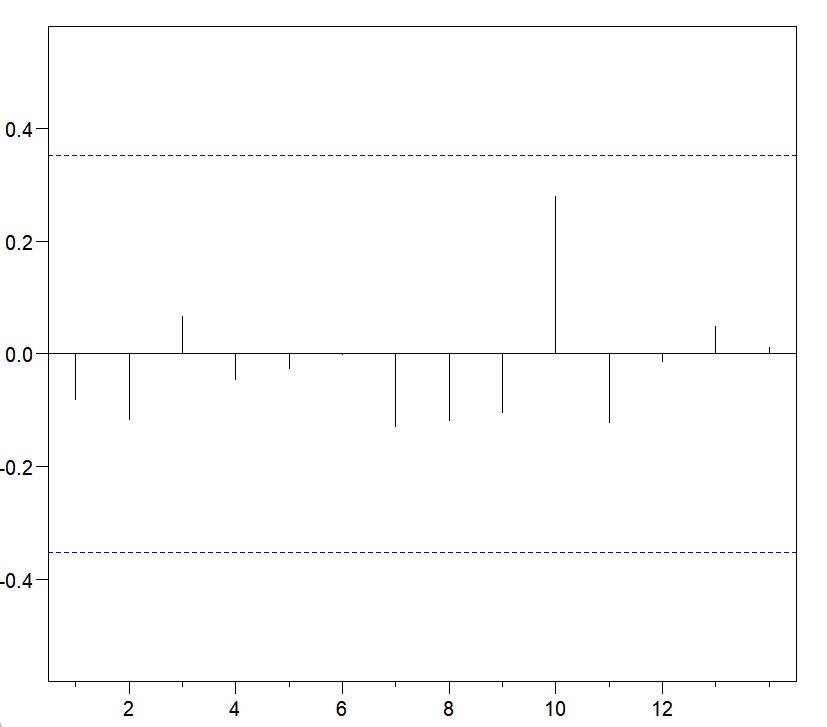
Opis wygenerowany automatycznie



****

**Obraz zawierający tekst

Opis wygenerowany automatycznie**

****

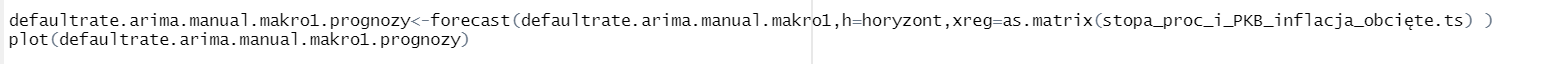
Reszty w tym modelu są stacjonarne.

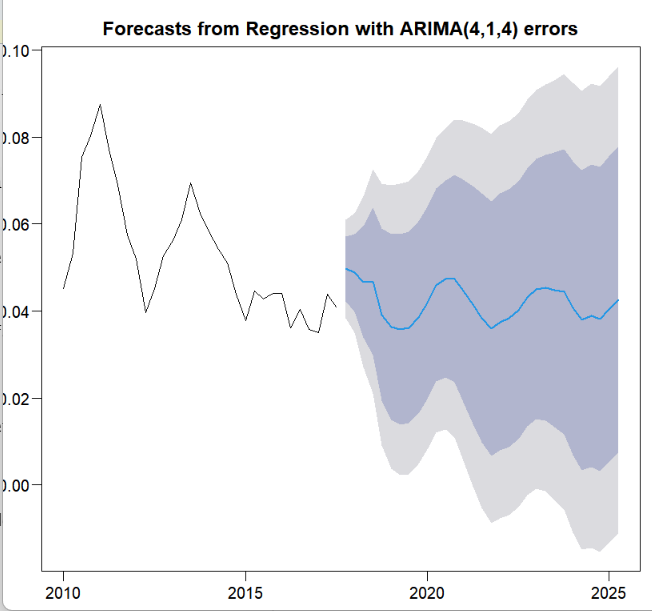
**Prognozy na modelu ręcznym bez sezonowości i ze zmiennymi makroekonomicznymi**



Obraz zawierający stół

Opis wygenerowany automatycznie



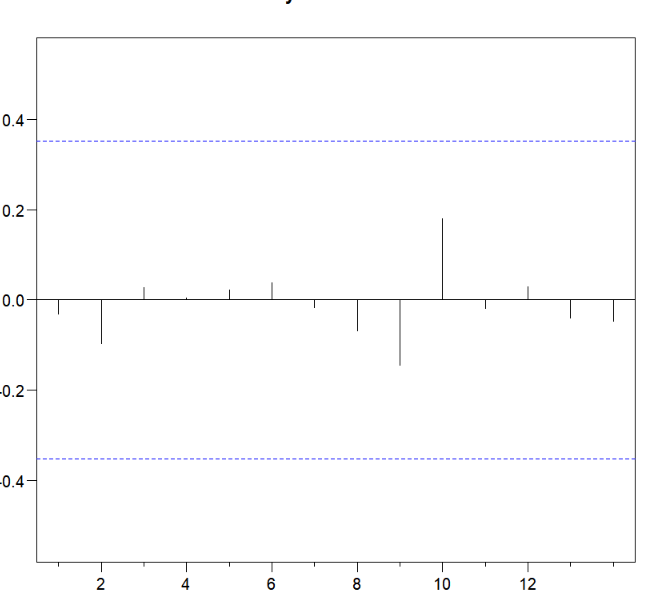


**Obraz zawierający tekst

Opis wygenerowany automatycznie**

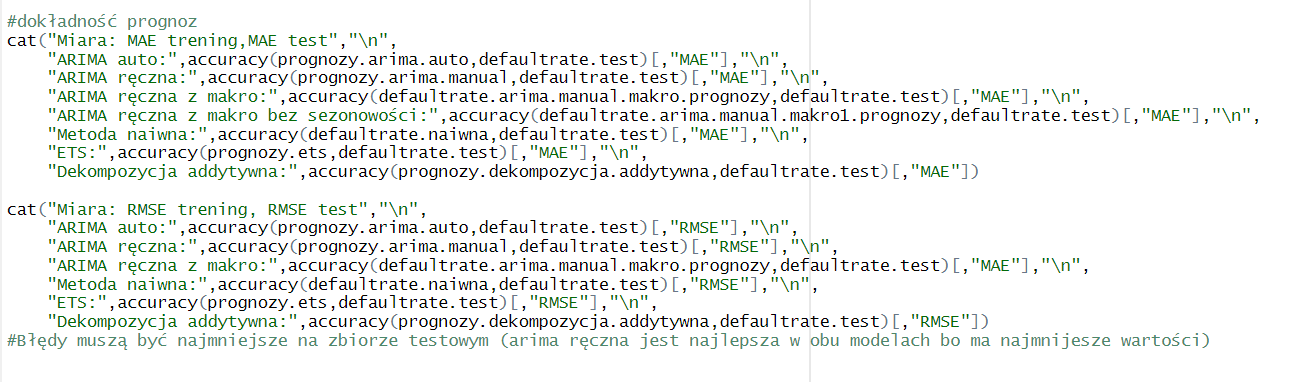
**Obraz zawierający tekst

Opis wygenerowany automatycznie**

****

Reszty w tym modelu są stacjonarne.

**Badanie dokładności prognoz**



Obraz zawierający tekst

Opis wygenerowany automatycznie

Obraz zawierający tekst

Opis wygenerowany automatycznie

Obraz zawierający tekst

Opis wygenerowany automatycznie

Na podstawie dokładności prognoz wnioskujemy, że model ARIMA Ręczna jest najlepiej dopasowanym modelem do prognozy, MAE test i RMSE test przyjmują najniższe wartości dla tego modelu.