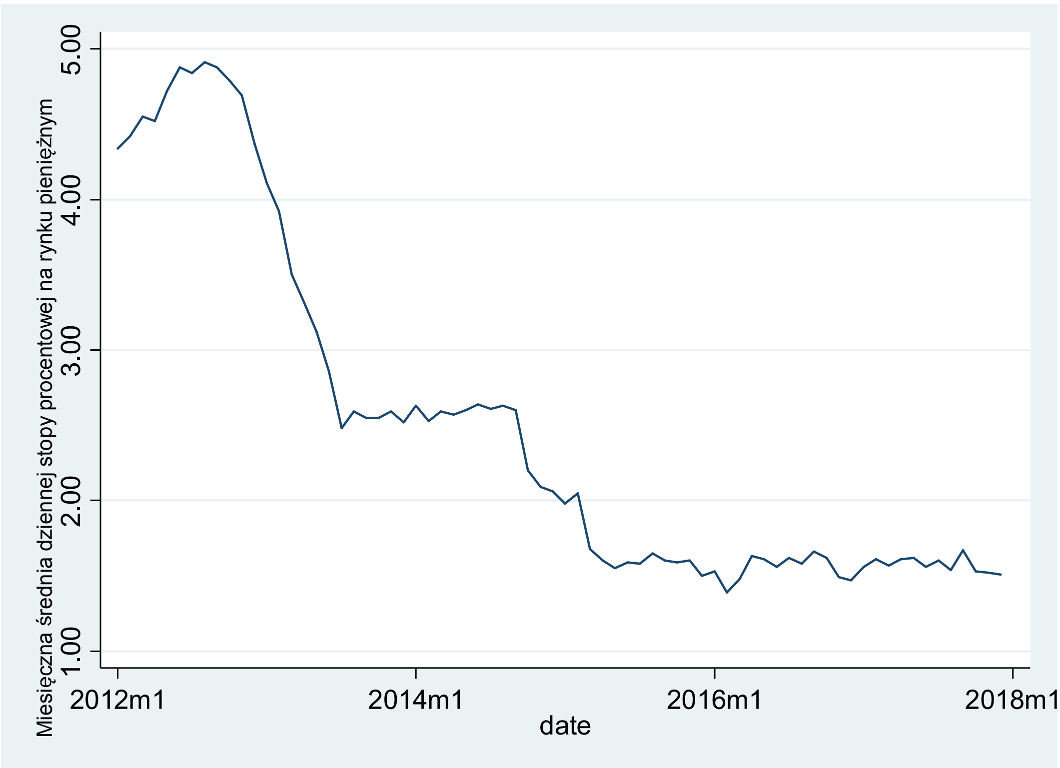
Spis treści

I. Szereg niesezonowy

1.1.Opis danych

Szereg danych niesezonowych pochodzi z bazy danych Eurostatu i dotyczy dziennej stopy procentowej na rynku pieniężnym. Jest to referencyjna stopa procentowa odnosząca się do krótkoterminowej stopy procentowej na rynku finansowym dla pożyczek lub depozytów. Do analizy wykorzystane zostały wydawane przez Eurostat uśrednione, miesięczne wartości tej stopy. Szereg zawiera obserwacje od stycznia 2012 roku do grudnia 2017 roku. Szereg został zaprezentowany na wykresie 1.1.1. Można zauważyć, że wartość stopy spadła o prawie 4 punkty procentowe od 2012. Drastyczny spadek nastąpił na przełomie lat 2012 oraz 2013.

Wykres 1.1.1 Miesięczne średnie dziennej stopy procentowej na rynku pieniężnym



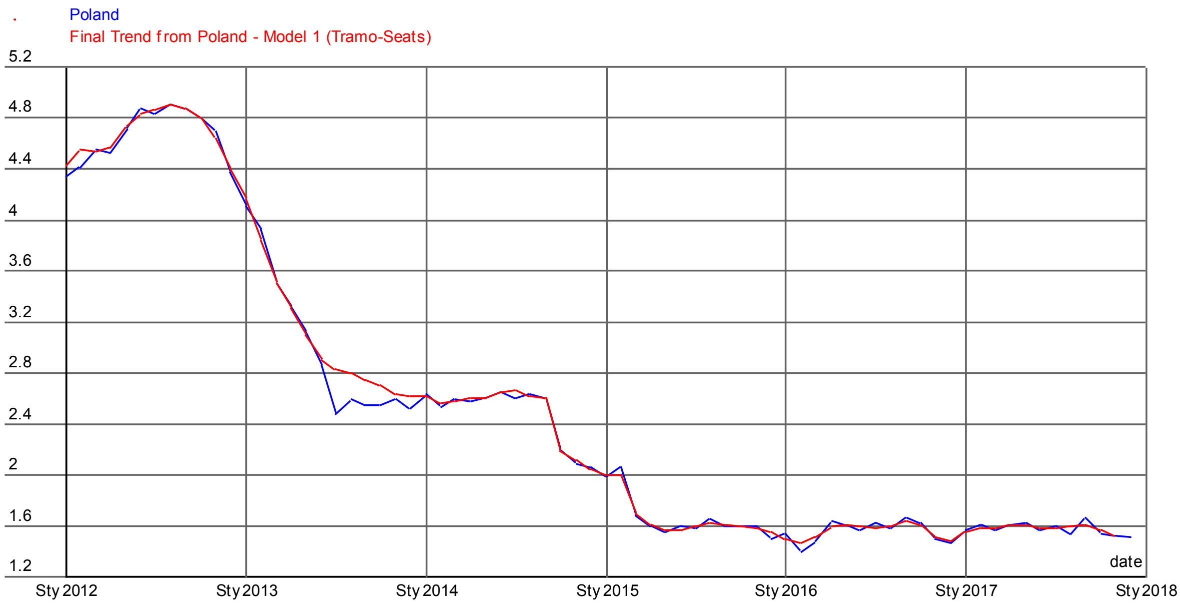
Źródło: Wykres z programu STATA na podstawie danych z Eurostatu

1.2.Dekompozycja szeregu w programie Demetra

Pierwszej analizy szeregu dokonano w programie Demetra. Dekompozycji szeregu dokonano za pomocą metody TRAMO/Seats. Model przeszedł wszystkie testy diagnostyczne oraz została potwierdzona jego niesezonowość.

Demetra wygładziła szereg, można zauważyć więc, że ma on trend malejący, przy czym w drugim kwartale 2015 roku jego wartość stabilizuje się na poziomie 1,2% - 1,6%. (wykres 1.2.1.).

Wykres 1.2.1. Dekompozycja analizowanego szeregu w programie Demetra

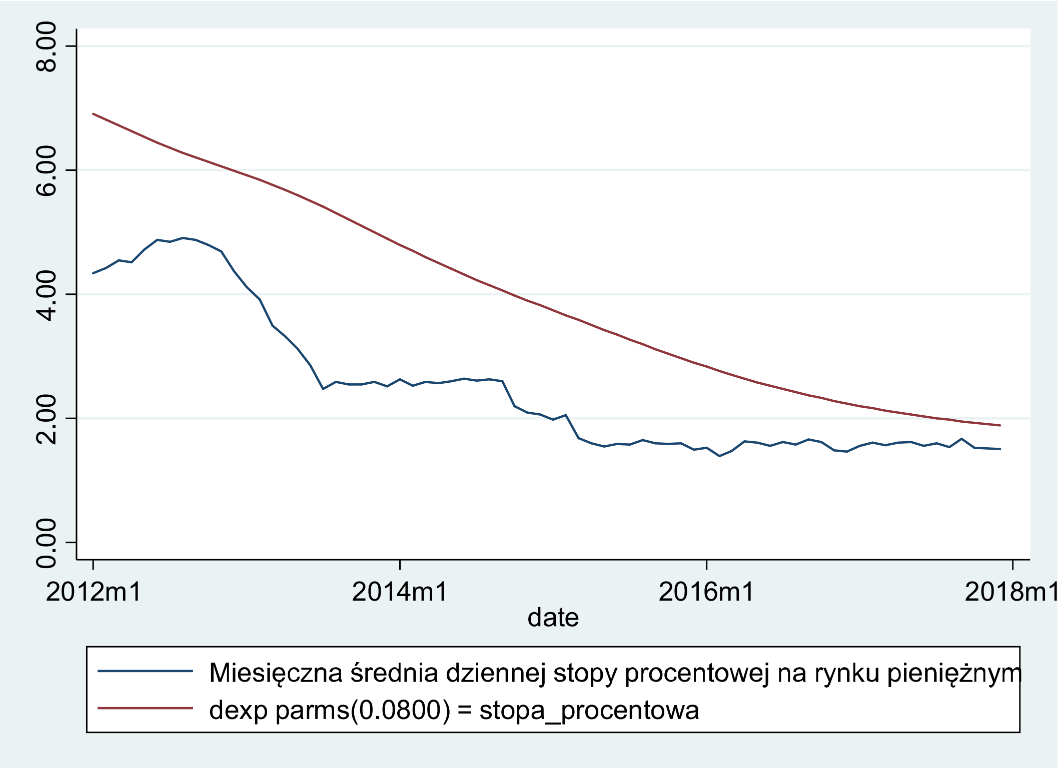


Źródło: Wykres z programu Demetra na podstawie danych z Eurostatu

1.3. Analiza szeregu w STATA

Do wygładzenia szeregu w programie STATA użyto podwójnego wygładzenia wykładniczego. Wybrano odpowiednio niewielki parametr dla tego zabiegu w celu znalezienia trendu dla danych. Wyniki zaprezentowano na wykresie 1.3.1. Widać tutaj, że trend jest malejący dla całego analizowanego okresu. W przeciwieństwie do trendu zaprezentowanego w programie Demetra, nie widać, aby od 2016 roku nastąpiła stabilizacja.

Wykres 1.3.1. Szereg niesezonowy a jego podwójne wygładzenie wykładnicze



Źródło: Wykres z programu STATA na podstawie danych z Eurostatu

1.4.Model ekstrapolacyjny

Aby uzyskać prognozę modelu ekstrapolacyjnego, użyto modelu Holta-Wintersa dla danych niesezonowych. Jako, że wyniki prognozy zostaną porównane z wynikami prognozy modelem ARIMA, który jest krótkookresowy, za okres *out-of-sample* wzięto trzy ostatnie obserwacje, a więc miesiące ostatniego kwartału roku 2017. Na ten okres będzie przeprowadzona prognoza. Tabela 1.4.1. przedstawia obliczone przez program STATA parametry dla modelu oraz średnie błędy prognoz.

Tabela 1.4.1. Wartości parametrów, suma resz tkwadratów oraz błędów dla prognozy przy użyciu modelu esktrapolacyjnego Holta-Wintersa dla danych niesezonowych

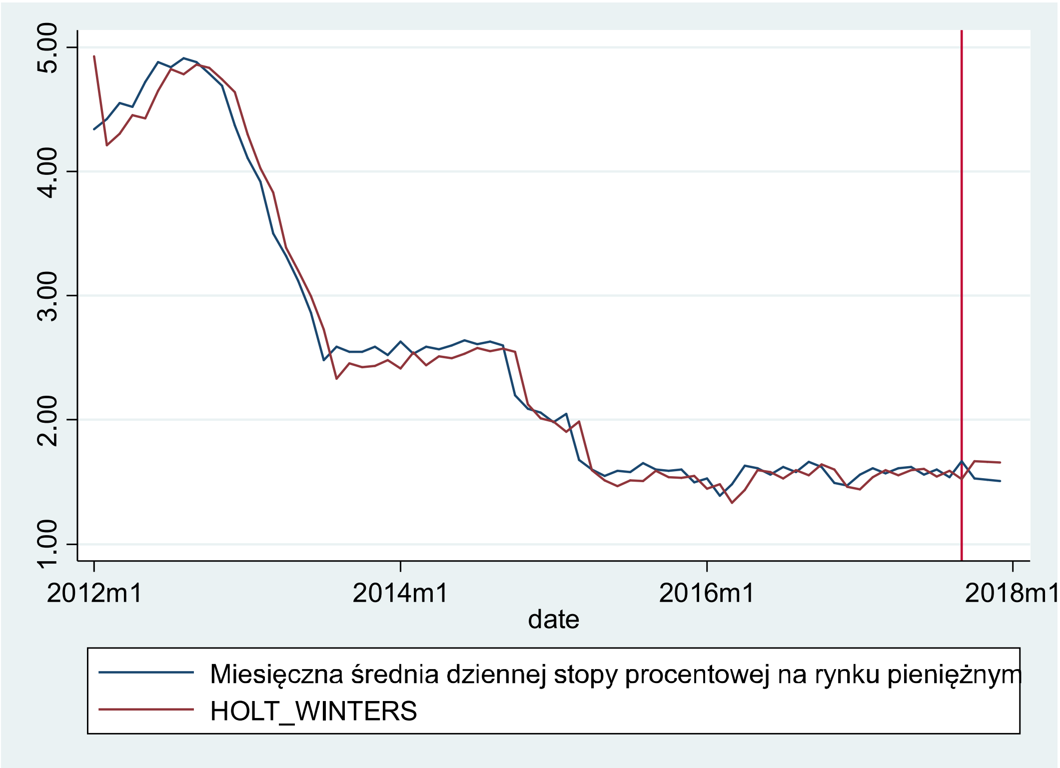
alpha	1
beta	0,0677
suma kwadratów reszt	1,598627
średnia suma kwadratów reszt	0,152212
MAE	0,1403103
MSE	0,0197047
MAPE	0,092369

AMAPE	0,0441241

Źródło: Wyniki estymacji w programie STATA na podstawie danych z Eurostatu

Wykres 1.4.1. przedstawia wyniki estymacji dla modelu Holta-Wintersa. Można zauważyć, dla okresu *out-of-sample* prognoza nie sprawdza się za dobrze. Gdy w rzeczywistości wartość stopy procentowej spada, wartości prognozy rosną.

Wykres 1.4.1. Wyniki prognozy modelu Holta-Wintersa a realne wartości



Źródło: Wykres z programu STATA na podstawie danych z Eurostatu

1.5. Model ARIMA

Kolejno, przeprowadzona zostanie estymacja dla modelu ARIMA.

Aby model mógł być zastosowany, analizowany szereg musi być stacjonarny. Poszukując stacjonarności, analizie poddany został szereg bez przekształceń. Stosując rozszerzony test Dickey-Fullera ze stałą, okazało się, że szereg jest niestacjonarny dla drugiego opóźnienia. Musiała zostać zastosowana rozszerzona wersja tego testu, ponieważ według testu Breauscha-Godfreya występowała autokorelacja. W celu uzyskania stacjonarności przeprowadzono test dla zróżnicowanego szeregu. W tym przypadku również użyty musiał być rozszerzony test Dickey-Fullera na stacjonarność, ponieważ założenie o braku autokorelacji nie było spełnione. Wyniki rozszerzonego testu Dickey-Fullera dla jednego opóźnienia (można było przyjąć hipotezę o braku autokorelacji – tabela 1.5.1.) przedstawione są w tabeli 1.5.2. Wynika z nich, że możemy odrzucić hipotezę o niestacjonarności szeregu. Aby mógł być on użyty dla prognozy modelu ARIMA należy jeszcze upewnić się, że opóźniony szereg nie jest białym szumem.

Tabela 1.5.1. Wyniki testu Breuscha-Godfreya dla analizowanego szeregu dla jednego opóźnienia dla zróżnicowanego szeregu

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	1.587	1	0.2078
2	1.842	2	0.3981
3	1.971	3	0.5784
4	2.894	4	0.5757
5	4.675	5	0.4568
6	4.796	6	0.5703

H0: no serial correlation

Źródło: Wydruk ze STATY na podstawie danych Eurostatu

Tabela 1.5.2. Wyniki rozszerzonego testy Dickey-Fullera dla analizowanego szeregu zróżnicowanego

Augmented Dickey-Fuller test for unit root   Number of obs = 69

	Test Statistic	Interpolated Dickey-Fuller: 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-3.769	-3.553	-2.915	-2.592

Źródło: Wydruk z programu STATA na podstawie danych Eurostatu

Wyniki testu Ljunga-Boxa pozwalają nam odrzucić hipotezę zerową mówiącą o tym, że badana w teście zmienna jest białym szumem (tabela 1.5.3.). Wszystko wskazuje więc na to, że szereg może być użyty do estymacji modelem ARIMA, dla *d*=1. Wykres 1.5.1. przedstawia zróżnicowany szereg czasowy analizowanych stóp procentowych.

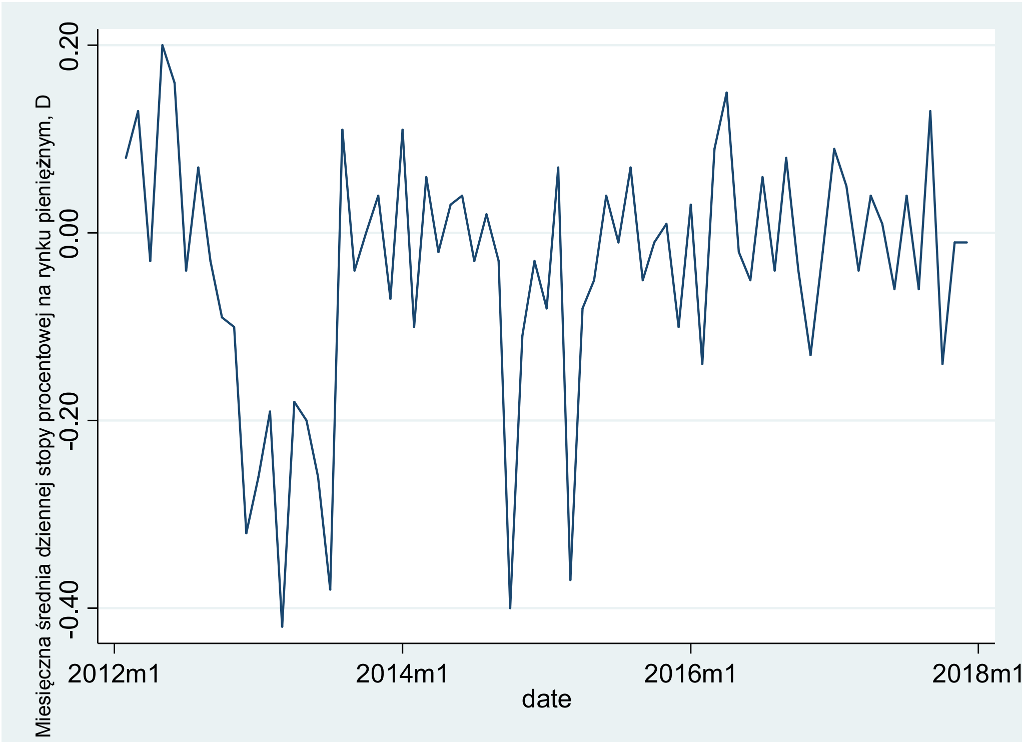
Tabela 1.5.3. Wyniki testu na biały szum dla zróżnicowanego szeregu

Portmanteau test for white noise

Portmanteau (Q) statistic =	544.1306
Prob > chi2(34) =	0.0000

Źródło: Wydruk ze STATY na podstawie danych Eurostatu

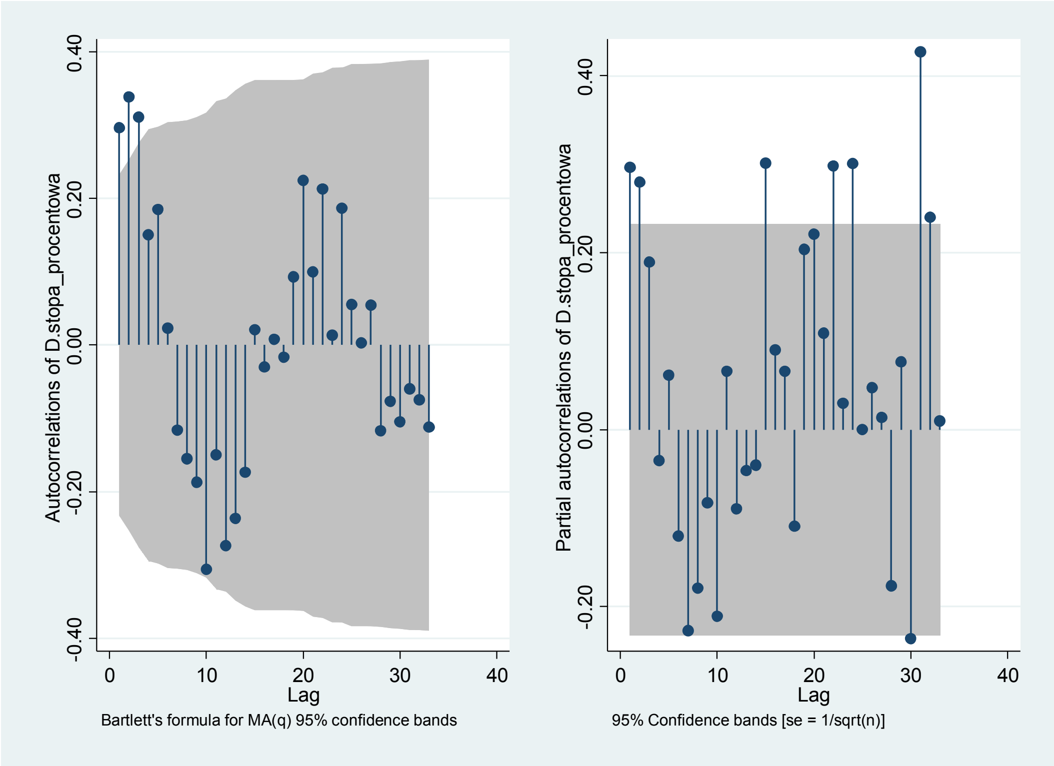
Wykres 1.5.1. Zróżnicowany szereg danych niesezonowych



Źródło: Wykres z programu STATA na podstawie danych z Eurostatu

Kolejnym krokiem do stworzenia modelu ARIMA jest wyznaczenie jego rzędów p i q, ponieważ z analizy stacjonarności otrzymaliśmy już d=1. W tym celu należy przeanalizować wykresy ACF oraz PACF. Są one przedstawione poniżej. Dzięki nim możemy wybrać wyjściowy model ARIMA, na który będziemy później nakładać ograniczenia. Nasz wyjściowy model to ARIMA(2,1,3) – widać bowiem, że wykres PACF ma istotne dwie pierwsze wypustki, a więc p=2, a wykres ACF ma istotne trzy pierwsze wypustki, a więc q=3.

Wykres 1.5.2. Wykresy PACF oraz ACF dla zróżnicowanego szeregu



Źródło: Wykres z programu STATA na podstawie danych z Eurostatu

Kolejno, na model zostały narzucone ograniczenia:

1. $H_0 : \beta_5=0$ $H_1 : \beta_5 \neq 0$
2. $H_0 : \beta_5=\beta_4=0$ $H_1 : \beta_5 \neq \beta_4 \neq 0$
3. $H_0 : \beta_5=\beta_4=\beta_3=0$ $H_1 : \beta_5 \neq \beta_4 \neq \beta_3 \neq 0$
4. $H_0 : \beta_2=\beta_5=0$ $H_1 : \beta_2 \neq \beta_5 \neq 0$
5. $H_0 : \beta_2=\beta_5=\beta_4=0$ $H_1 : \beta_2 \neq \beta_5 \neq \beta_4 \neq 0$
6. $H_0 : \beta_2=\beta_5=\beta_4=\beta_3=0$ $H_1 : \beta_2 \neq \beta_5 \neq \beta_4 \neq \beta_3 \neq 0$

Istotność narzucenia ograniczeń została poddana teście ilorazu wiarygodności. Na poziomie istotności 1%, hipotezę należało przyjąć dla ograniczeń $\beta_2=\beta_5=\beta_4=0$. Aby wybrać jak najlepsze rzędy dla modelu, należy również porównać modele za pomocą kryteriów informacyjnych. Jak widać, ARIMA(2,1,3) charakteryzuje się najmniejszą wartością AIC, ale ARIMA(1,1,1) posiada najmniejsze BIC (tabela 1.5.4.). Aby podjąć decyzję, sprawdzono również wyniki estymacji dla każdego z modeli, aby sprawdzić, czy wszystkie opóźnienia są w nich istotne. Okazuje się, że dla estymacji modelu, ARIMA(1,1,1) wszystkie opóźnienia są istotne, czego nie można powiedzieć, o modelu ARIMA(2,1,3). Do estymacji wybrany zostaje więc model ARIMA (1,1,1) (wyniki estymacji w tabeli 1.5.5.)

Tabela 1.5.4. Kryteria informacyjne dla poszczególnych modeli ARIMA

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
arma23	68	.	52.66675	7	-91.3335	-75.79695
arma22	68	.	48.26713	5	-86.53425	-75.43671
arma21	68	.	47.78404	5	-85.56807	-74.47053
arma20	68	.	46.98639	4	-85.97279	-77.09476
arma12	68	.	48.05552	5	-86.11104	-75.0135
arma11	68	.	47.11142	4	-86.22284	-77.34481
arma10	68	.	44.37191	3	-82.74383	-76.08531

Źródło: Wydruk z programu STATA na podstawie danych Eurostatu

Tabela 1.5.5. Estymacja dla modelu ARIMA(1,1,1)

D. stopa_procentowa	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
stopa_procentowa						
_cons	-.0305442	.0418249	-0.73	0.465	-.1125195	.051431
ARMA						
ar						
L1.	.8075501	.1497651	5.39	0.000	.514016	1.101084
ma						
L1.	-.5323333	.2426387	-2.19	0.028	-1.007896	-.0567703
/sigma	.1207897	.0089139	13.55	0.000	.1033188	.1382606

Źródło: Wydruk z programu STATA na podstawie danych Eurostatu

Dla wybranego modelu wiemy również, że możemy analizować jego wyniki, ponieważ reszty z estymowanego modelu nie są białym szumem, co zostało sprawdzone testem Ljunga-Boxa – nie było podstaw do odrzucenia hipotezy mówiącej, że badana zmienna jest białym szumem (tabela 1.5.6.).

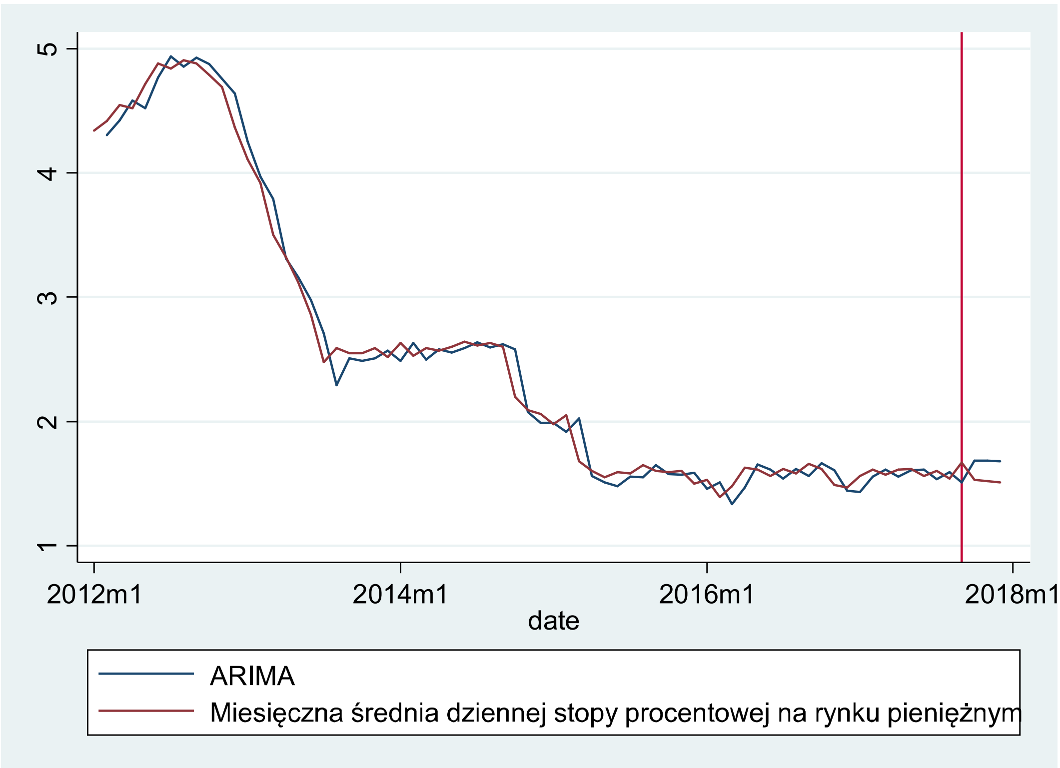
Tabela 1.5.6. Wyniki testu na biały szum dla reszt modelu ARIMA (1,1,1)

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    34.9816
 Prob > chi2(32)           =     0.3283
```

Źródło: Wydruk z programu STATA na podstawie danych Eurostatu

Mając spełnione wymogi do przeprowadzenia prognozy ARIMA, wykres 1.5.3. przedstawia wyniki prognozy dynamicznej modelu ARIMA dla 3 ostatnich okresów. Można zauważyć, że prognoza ta jest podobna do estymacji modelem Holta-Wintersa – prognoza ARIMA również zakłada wzrost wartości, gdy w rzeczywistości następuje jej spadek. Dodatkowo, tabela 1.5.7. przedstawia średnie błędy dla prognoz modelu ARIMA.

Wykres 1.5.3. Wyniki estymacji modelu ARIMA(1,1,1) dla zmiennej *lnHDD*



Źródło: Wykres z programu STATA na podstawie danych z Eurostatu

Tabela 1.5.7. Średnie błędy dla prognozy ARIMA

Błąd	**Wartość**
MSE	0,0264735
MAE	0,1625227

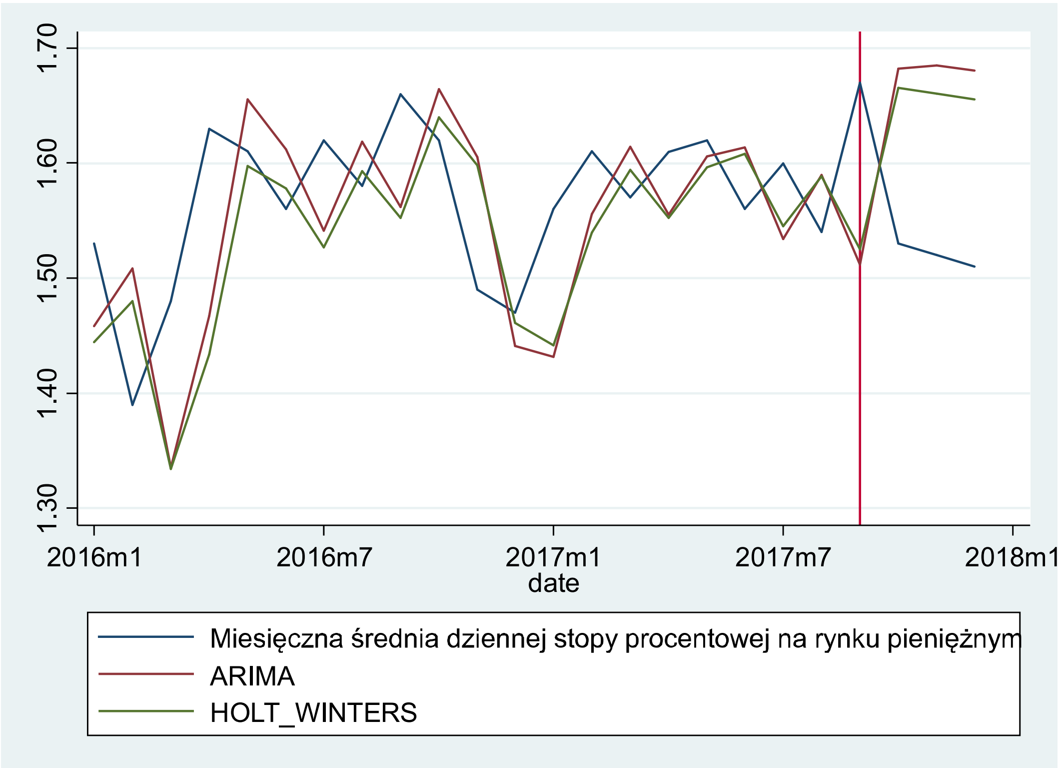
MAPE	0,1069525
AMAPE	0,0507549

Źródło: Opracowanie własne wyników z programu STATA na podstawie danych Eurostatu

1.6.Porównanie modeli

Jak zauważono w poprzednim rozdziale, predykcja modelem ARIMA(1,1,1) daje podobne rezultaty jak ta modelem ekstrapolacyjnym. Wykres 1.6.1. przedstawia obie prognozy wraz z rzeczywistymi danymi (aby lepiej widzieć prognozę, wykres przedstawia jedynie obserwacje od 2016 roku). Z wykresu można zauważyć, że bliżej rzeczywistych wartości jest prognoza modelem Holta-Wintersa. Dla pewności, w tabeli 1.6.1. porównano średnie błędy prognoz dla obu modeli. Rzeczywiście, średnia każdego z błędów jest mniejszyadla modelu Holta-Wintersa. Okazuje się więc, że dla badanych danych, to model ekstrapolacyjny sprawdza się lepiej w prognozach, w porównaniu do modelu ARIMA.

Wykres 1.6.1. Wartości modelu ARIMA a wartości modelu ekstrapolacyjnego a realne wartości indeksu



Źródło: Wykres z programu STATA na podstawie danych z Eurostatu

Tabela 1.6.1. Średnie błędy prognoz dla modelu ARIMA (1,1,1) oraz modelu ekstrapolacyjnego Holta-Wintersa

Średni błąd	**ARIMA(1,1,1)**	**Holt-Winters**
MSE	0,026474	0,019705
MAE	0,162523	0,14031
MAPE	0,106953	0,092327
AMAPE	0,050755	0,044124

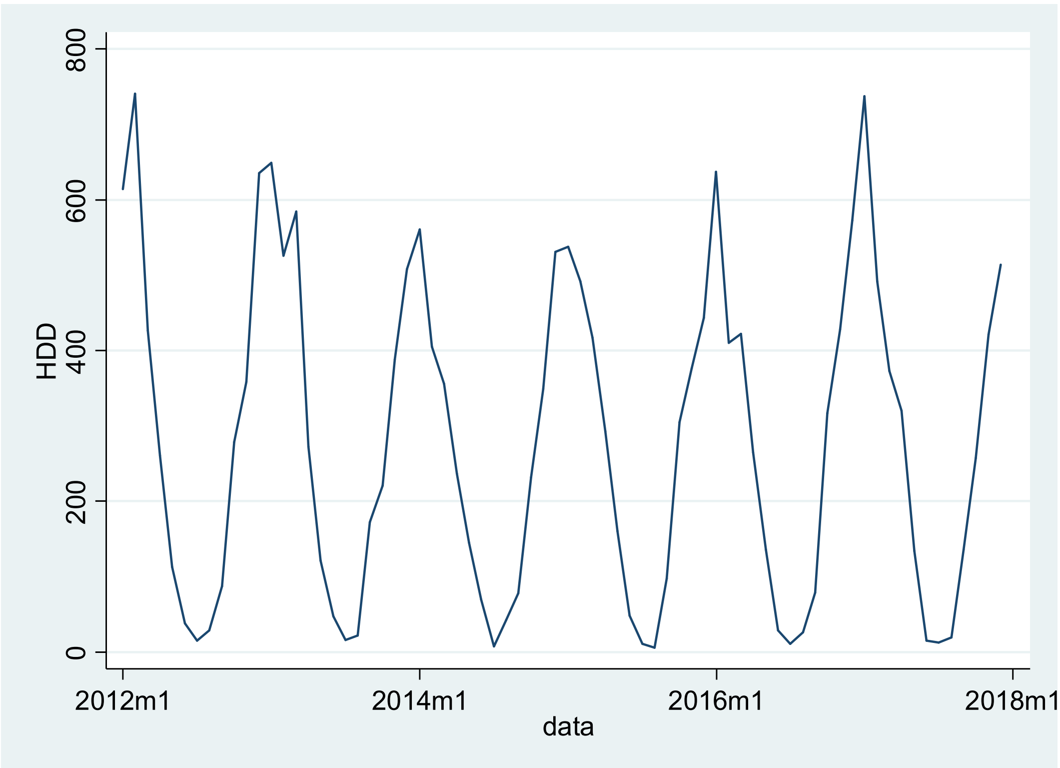
Źródło: Opracowanie własne wyników z programu STATA na podstawie danych Eurostatu

II Szereg sezonowy

2.1. Opis danych

Dane sezonowe analizowane w pracy to dane miesięczne dotyczące wskaźnika *Heating Degreee Days (HDD)* dla małopolski. Wskaźnik *HDD* jest to techniczny indeks opisujący potrzebę energii grzewczej wymaganej w budynkach dla konkretnego obszaru dla każdego miesiąca roku. Obliczany jest na podstawie obserwacji meteorologicznych. Analizowane dane mają charakter miesięczny, a okres, którego dotyczą to: styczeń 2012 – grudzień 2017 rok. Dane pochodzą z Eurostatu. Wykres 2.1.1. przedstawia „surowy" wykres szeregu.

Wykres 2.1.1. Szereg sezonowy wartości HDD dla okresu 01.2012-12.2017

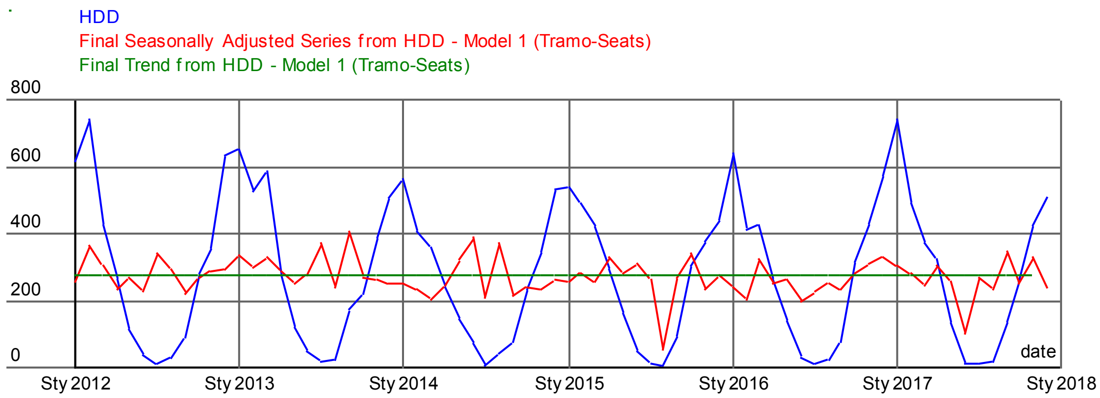


Źródło: Wykres wygenerowany z programu STATA a podstawie danych pochodzących z Eurostatu

2.2. Dekompozycja w Demetrze

Dane poddane analizie w Demetrze wskazują na sezonowość. Wykres 2.2.1. przedstawia dekompozycję szeregu w programie Demetra przy użyciu metody Tramo-Seats z siedmioma regresorami. Trend po odsezonowaniu jest stały. Może być to spowodowane faktem, że w rejonie, w którym żyjemy nie nastąpiły na przestrzeni analizowanego okresu zmiany klimatu, które doprowadziłyby do zwiększenia lub zmniejszenia czasu ogrzewania budynków.

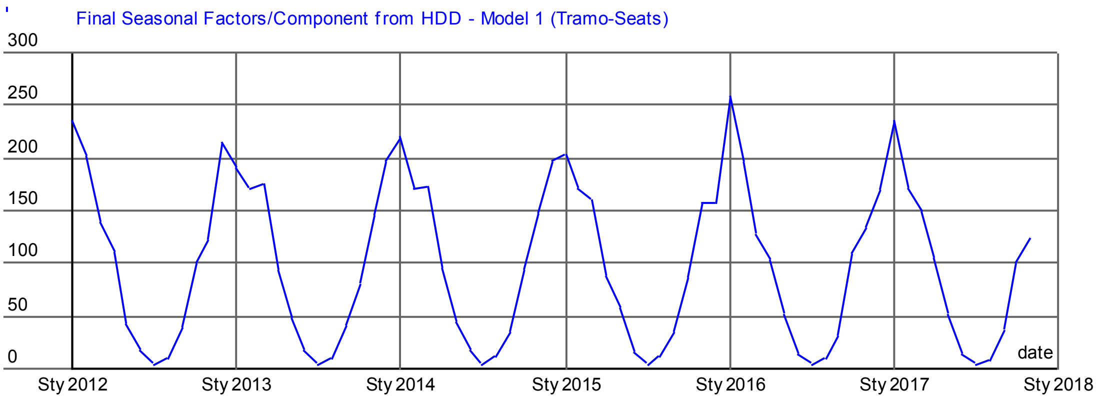
Wykres 2.2.1. Dekompozycja szeregu czasowego dot. wartości wskaźnika *HDD*



Źródło: wydruk z programu Demetra na podstawie danych pochodzących z Eurostatu

Wykres 2.2.2. przedstawia wyodrębniony komponent sezonowy. Amplituda wahań jest względnie stała, wygląda więc na to, że szereg charakteryzuje się sezonowością addytywną. Tak jak wskazuje trend, sezonowość jest dość regularna, co jest zrozumiałe, ponieważ okres ogrzewania budynków w roku jest stały, szczególnie dla mieszkań w blokach, gdzie okres ogrzewania jest z góry określony.

Wykres 2.2.2. Komponent sezonowy dla szeregu dot. danych nt. wskaźnika HDD

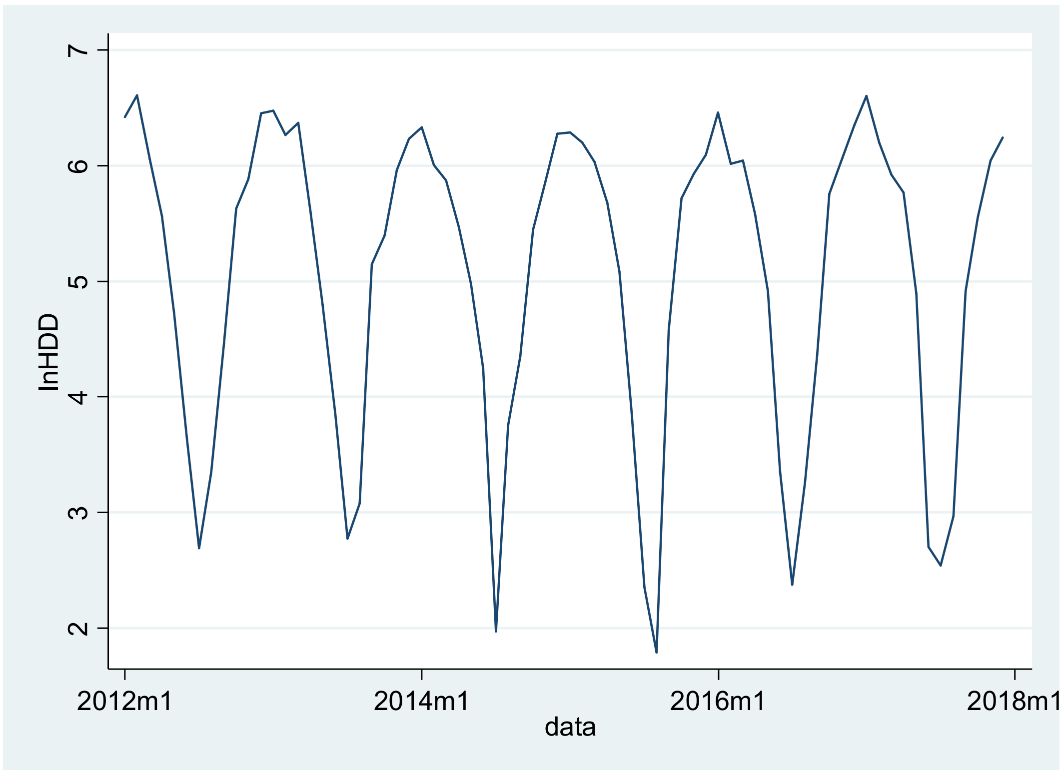


Źródło: Wydruk z programu Demetra na podstawie danych pochodzących z Eurostatu

2.3. Analiza szeregu w STATA

Przedstawiony we wcześniejszych podrozdziałach szereg został zlogarytmowany w celu zmniejszenia wartości wariancji. Od tej pory będzie on analizowany w takiej formie. Wykres 2.3.1. przedstawia „surowy" szereg zlogarytmowany.

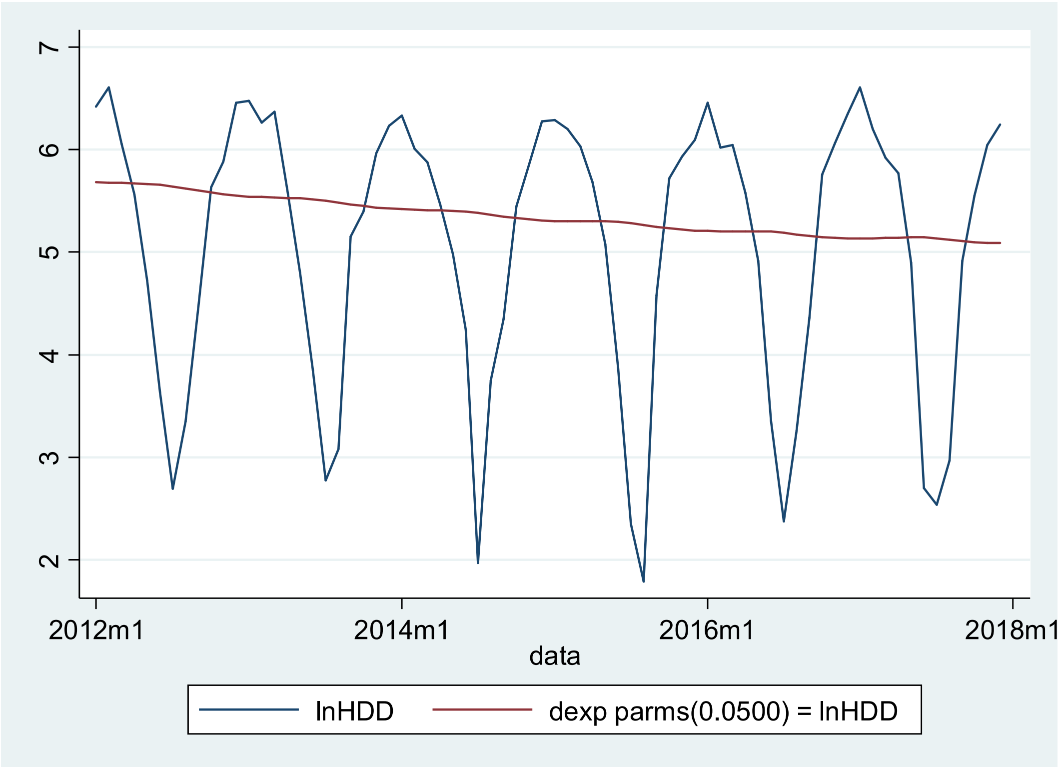
2.3.1. Zlogarytmowany szereg sezonowy – wartość wskaźnika HDD



Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Kolejno, w celu przedstawienia trendu w programie STATA, został on poddany podwójnemu wygładzeniu wykładniczemu. Wynik tego zabiegu przedstawiony jest na wykresie 2.3.2. Można zauważyć, że otrzymany trend również jest w miarę stały, jednak w przeciwieństwie do tego wygenerowanego przez program Demetra, zauważyć można, że maleje.

Wykres 2.3.2. Szereg „surowy" oraz wygładzony podwójnie wykładniczo



Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

2.4. Model ekstrapolacyjny

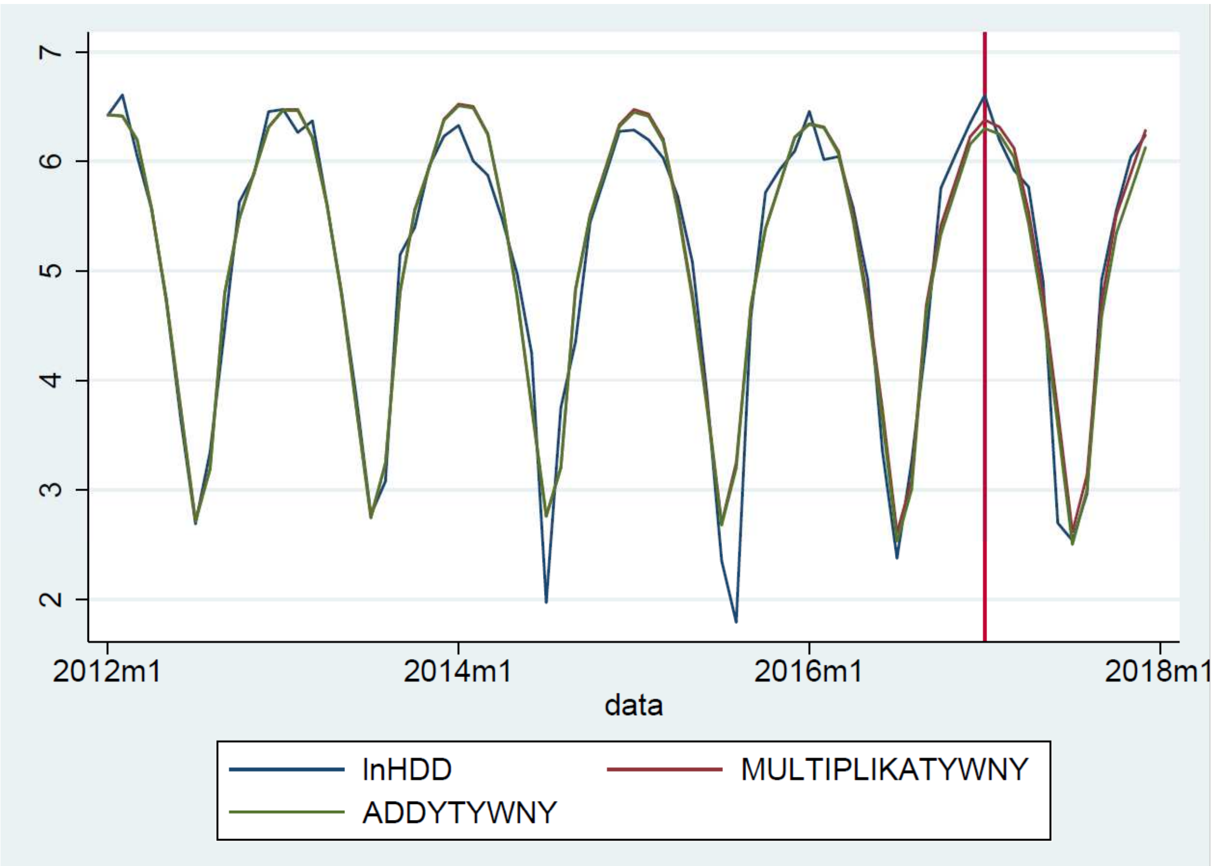
Pierwszym modelem, którego użyjemy do prognozy jest model ekstrapolacyjny Holta-Wintersa dla szeregów sezonowych. Pierwszym krokiem, który należy wykonać jest wybór postaci modelu ekstrapolacyjnego – addytywnego, czy multiplikatywnego. Z samej analizy wykresu można założyć, że byłaby to forma addytywna, jednak należy się upewnić analizując oba modele. Jako próbę *in-sample* wybrano okres od 2012 do 2016 roku, próba *out-of-sample* to 12 ostatnich miesięcy analizowanego okresu, a więc cały rok 2017. Tabela 2.4.1 przedstawia średnie błędy prognoz dla obu modeli. Z przedstawionych wartości widać, że dla większości wartości, są one mniejsze dla modelu multiplikatywnego, dlatego też to on zostaje wybrany do prognoz. Obliczone wartości parametrów dla modelu to *alpha*= 0,0482, *beta*=0,001, *gamma*=0,0712, a średnia suma kwadratów reszt wynosi 0,308589. Dodatkowo, na wykresie przedstawiającym wyniki z estymacji obu modeli wraz z badanym szeregiem, można zauważyć, że szczególnie dla ostatnich obserwacji, model multiplikatywny jest bliższy realnym wartościom (wykres 2.4.1.).

Tabela 2.4.1. Średnie błędy prognoz dla modeli ekstrapolacyjnych dla zmiennej *lnHDD*

	Model addytywny	**Model multiplikatywny**
MAE	0,249416	0,2171424
MSE	0,1118675	0,1104469
MAPE	0,0596538	0,0583166
AMAPE	0,0282573	0,0266377

Źródło: Opracowanie własne wyników otrzymanych w STATA na postawie danych z Eurostatu

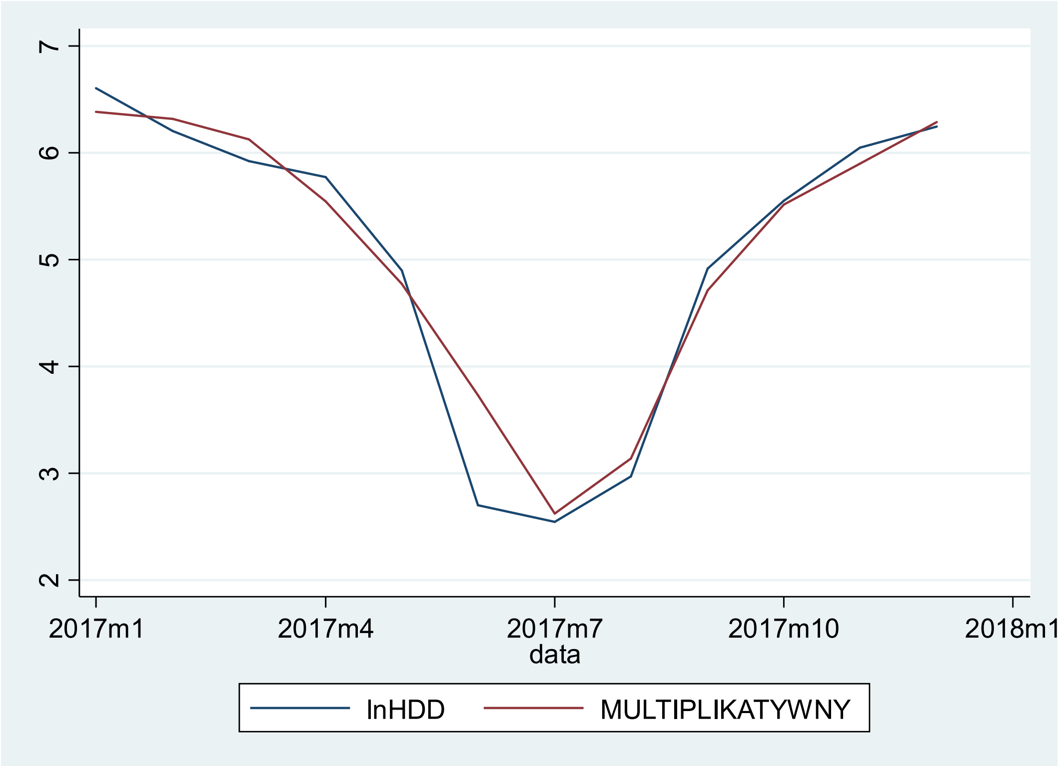
Wykres 2.4.1. Modele ekstrapolacyjne a szereg lnHDD



Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Chcąc zweryfikować prognozę wybranym modelem ekstrapolacyjnym z realnymi wartościami, można zauważyć, że okresem,w którym wartości najmniej się pokrywają jest drugi kwartał 2017 roku (wykres 2.4.1.)

Wykres 2.4.1. Modele ekstrapolacyjne a szereg lnHDD



Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

2.5. Model SARIMA

Model SARIMA jest modelem długookresowym, dlatego też prognoza będzie obejmowała okres *out-of-sample*, a więc podobnie jak w modelu ekstrapolacyjnym, będzie to 12 ostatnich miesięcy (cały rok 2017). Aby model mógł być zastosowany do danych, należy sprawdzić ich stacjonarność, ponieważ nie można go stosować dla danych niestacjonarnych. Testując stacjonarność szeregu rozszerzonym testem Dickey-Fullera (należało go zastosować z powodu występowania autokorelacji), okazała się, że szereg danych *lnHDD* nie jest stacjonarny (tabela 2.5.1.). Dopiero po zróżnicowaniu zmiennej lnHDD, wyniki test Dickey-Fullera pozwoliły odrzucić hipotezę o niestacjonarności zmiennej (tabela 2.5.2.). Po zróżnicowaniu nie wykryto również autokorelacji testem Breuscha-Godfreya, dzięki czemu nie trzeba było stosować testu rozszerzonego (wyniki również zawarte w tabeli 2.5.2.).

Tabela 2.5.1. Wyniki rozszerzonego testu Dickey-Fullera dla lnHDD

Augmented Dickey-Fuller test for unit root Number of obs = 70

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-1.170	-2.612	-1.950	-1.610

Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Tabela 2.5.2. Wyniki testu Dickey-Fullera dla zróżnicowanego szeregu lnHDD oraz wyniki testu Breuscha-Godfreya

Dickey-Fuller test for unit root Number of obs = 70

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-5.415	-2.612	-1.950	-1.610

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	1.184	1	0.2766
2	5.380	2	0.0679
3	6.339	3	0.0962
4	8.698	4	0.0691
5	8.715	5	0.1210
6	9.612	6	0.1420
7	11.600	7	0.1145
8	17.051	8	0.0296

H0: no serial correlation

Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Dzięki pojedynczemu zróżnicowaniu szeregu, wiemy, że w modelu SARIMA d=1. Dodatkowo, po jednokrotnym różnicowaniu sezonowym, sezonowość zanika, a więc również D=1.

Po zróżnicowaniu, należy sprawdzić, czy zróżnicowany szereg nie stał się białym szumem. Wyniki testu Ljunga-Boxa wskazują, że należy odrzucić hipotezę mówiącą, że szereg jest białym szumem (tabela 2.5.3.).

Tabela 2.5.3. Wyniki testu na biały szum dla zróżnicowanej zmiennej *lnHDD*

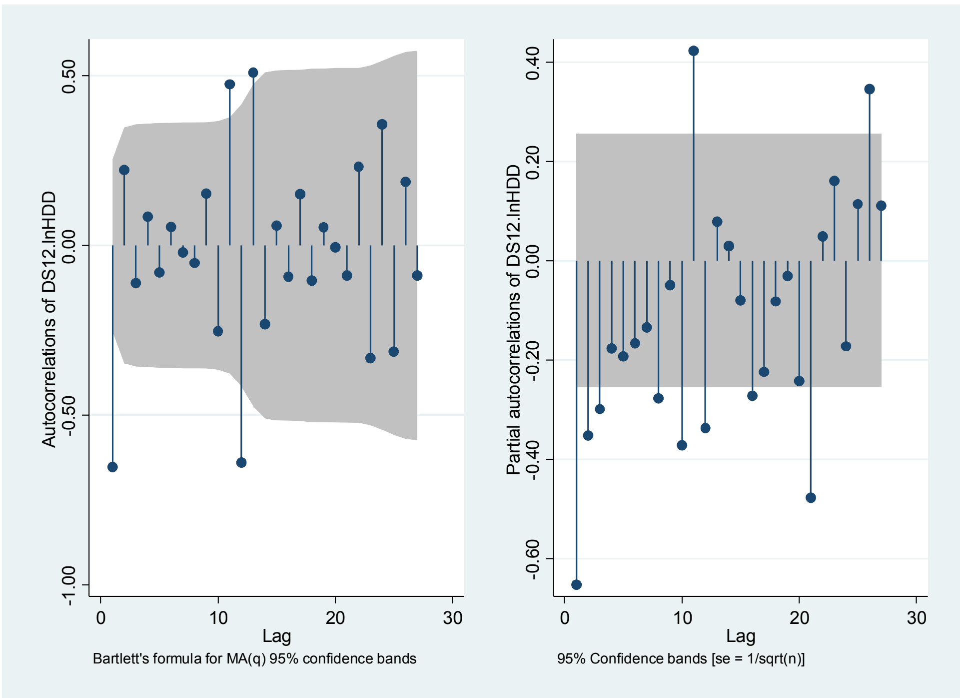
```
Portmanteau test for white noise
_______________________________________________

 Portmanteau (Q) statistic =   231.2813
 Prob > chi2(33)           =     0.0000
```

Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Następnie poszukujemy rzędów *p, q, P, Q*. Wstępnie można wyznaczyć je dzięki analizie wykresów funkcji autokorelacji oraz autokorelacji cząstkowej dla odsezonowanego, zróżnicowanego szeregu (wykres 2.5.1.). Rzędy *p* oraz *q* wyznacza się tak jak w modelu ARIMA, dlatego też po analizie wykresów, wstępnie zakładamy, że *p*=3 (trzy pierwsze wypustki istotne dla wykresu PACF), a *q*=1 (tylko pierwsza wypustka istotna dla wykresu ACF). Kolejno, nalezy dobrać rzędy dla SMA oraz SRA. Analizując wykres ACF, jedynie pierwsza dwunasta wypustka jest istotna, więc w naszym modelu wyjściowym mamy *Q*=1. Podobna sytuacja zachodzi dla wykresy PACF, a więc *P*=1. Po analizie wykresów funkcji autokorelacji oraz autokorelacji cząstkowej dla odsezonowanej, zróżnicowanej zmiennej, wychodzi nam, że wyjściowym modelem dla szeregu sezonowego będzie model SARIMA(3,1,1)(1,1,1,12).

Wykres 2.5.1. Wykresy funkcji autokorelacji oraz autokorelacji cząstkowej dla odesenowanego, zróżnicowanego szeregu *lnHDD*



Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Mając model wyjściowy, powinniśmy zbadać wartości kryterium AIC oraz BIC dla modelu wyjściowego oraz modeli z ograniczeniami. Niestety, modele o najniższych wartościach AIC oraz BIC charakteryzują się nieistotnością opóźnień dla części ARIMA lub SARIMA (tabela 2.5.4.). Do prognoz wybrano więc model z najmniejszą wartością AIC oraz BIC, dla którego istotne są części ARIMA oraz SARIMA. Ostatecznie, jest to model SARIMA(2,1,0)(0,1,1,12). Aby upewnić się, że wybrany model jest poprawny, należy sprawdzić, czy reszty z modelu są białym szumem. Jak wskazują wyniki testu Ljunga-Boxa, hipoteza mówiąca o tym, że badana zmienna jest białym szumem nie została odrzucona (tabela 2.5.5.).

Tablea 2.5.4. Kryteria informacyjne dla wybranych modeli SARIMA

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
sarma3010	59	.	-33.3394	6	78.67881	91.14403
sarma2010	59	.	-35.77345	5	81.5469	91.93459
sarma1010	59	.	-38.95839	4	85.91678	94.22693
sarma2110	59	.	-26.54813	5	63.09627	73.48396
sarma0110	59	.	-26.89171	4	61.78342	70.09357
sarma1110	59	.	-26.70026	4	61.40052	69.71067
sarma3110	59	.	-26.26004	6	64.52007	76.9853
sarma3011	59	.	-31.36131	7	76.72262	91.26538
sarma2011	59	.	-33.70994	6	79.41988	91.8851
sarma1011	59	.	-36.69508	5	83.39016	93.77785
sarma2111	59	.	-24.82427	5	59.64854	70.03623
sarma0111	59	.	-25.40945	4	58.81891	67.12906
sarma1111	59	.	-24.86162	4	57.72324	66.03339
sarma3111	59	.	-24.47653	7	62.95307	77.49583
sarma3001	59	.	-31.37844	6	74.75689	87.22211
sarma2001	59	.	-33.71248	4	75.42495	83.7351
sarma1001	59	.	-36.69509	3	79.39018	85.62279
sarma2101	59	.	-24.82451	5	59.64903	70.03672
sarma0101	59	.	-25.49376	3	56.98751	63.22013
sarma1101	59	.	-24.86468	5	59.72936	70.11705
sarma3101	59	.	-24.47722	6	60.95444	73.41967

Note: N=Obs used in calculating BIC; see [R] BIC note.

Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

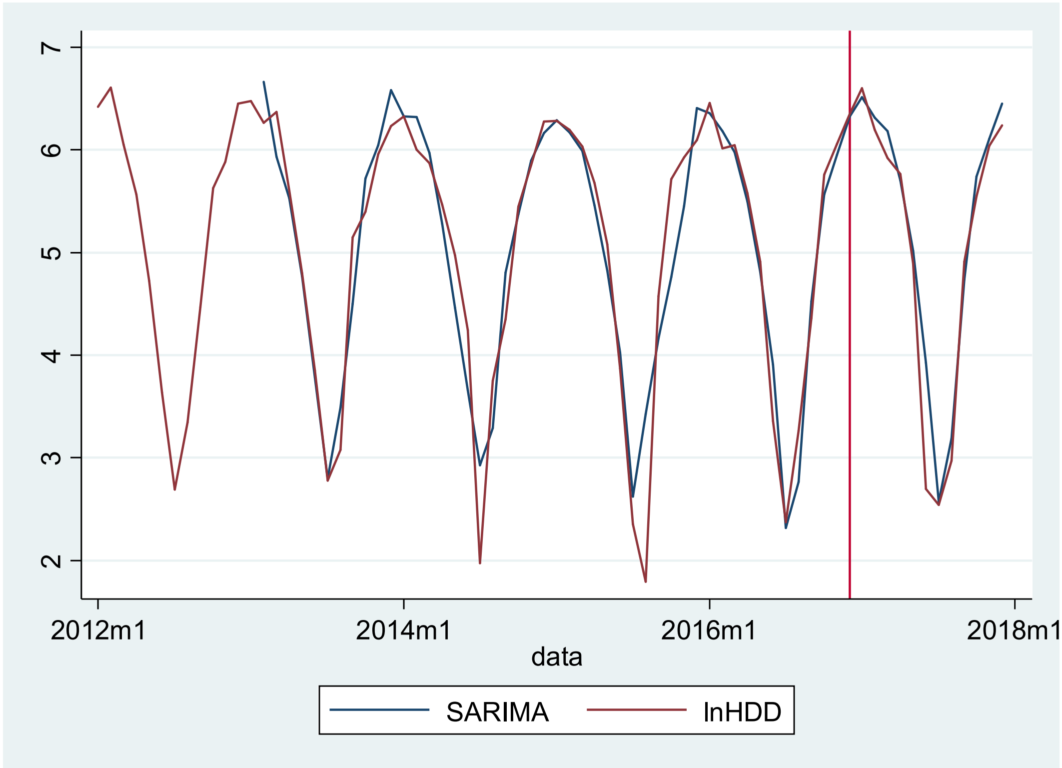
Tabela 2.5.5. Wyniki testu Ljunga-Boxa dla białegu szumu z modelu SARIMA

```
Portmanteau test for white noise
-----------------------------------------------
 Portmanteau (Q) statistic =    33.6866
 Prob > chi2(27)           =     0.1753
```

Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Wartości dla modelu SARIMA przedstawione zostały na wykresie 2.5.2. Jak można zauważyć, prognoza dla okresu *out-of-sample* pokrywa się z wartościami realnymi.

Wykres 2.5.2. Realne wartości szeregu lnHDD a estymacja modelu SARIMA(2,1,0)(0,1,1,12)



Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Tabela 2.5.6. Średnie błędy dla prognozy SARIMA(2,1,0)(0,1,1,12)

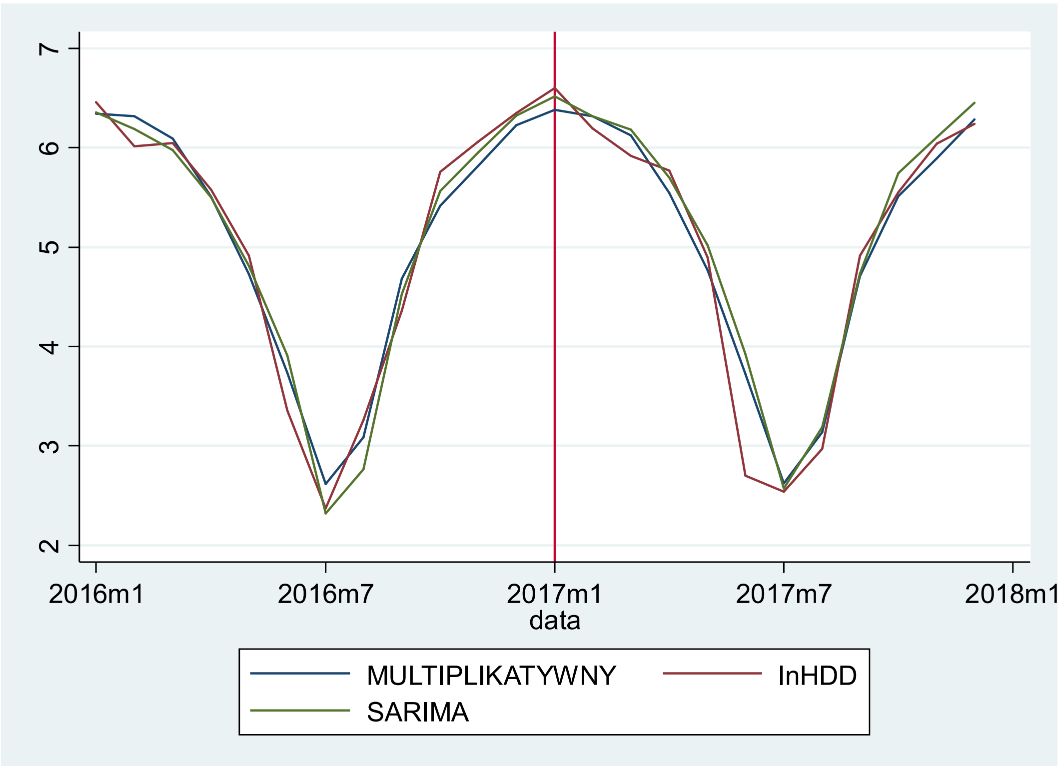
Średnie błędy	**SARIMA(2,1,0)(0,1,1,12)**
MAE	0,2317701
MSE	0,1479902
MAPE	0,0583166
AMAPE	0,0266377

Źródło: Opracowanie własne wyników otrzymanych w STATA na postawie danych z Eurostatu

2.6. Porównanie modeli

Podczas estymacji modelami ekstrapolacyjnymi, jako lepszy wybrany został model multiplikatywny. Teraz porównamy jego wyniki z wynikami modelu SARIMA(2,1,0)(0,1,1,12). Aby sprawdzić dla którego modelu mamy lepszą prognozę, należy przenalizować wykresy estymacji obu modeli oraz wartości rzeczywistych. Z wykresu 2.6.1. można wnioskować, że to model multiplikatywny lepiej estymuje wartości rzeczywiste. Dla pewności, należy jeszcze porównać wartości średnich błędów dla prognoz dla każdego z modeli. Są one zamieszczone w tabeli 2.6.1. Wynika z nich, że rzeczywiście, dla modelu multiplikatywnego średnie wartości błędów są mniejsze. Okazuje się więc, że to model ekstrapolacyjny multiplikatywny prognozuje najlepiej dla badanego, zlogarytmowanego szeregu.

Wykres 2.6.1. Estymacja modelem multiplikatywnym, modelem SARIMA a realne wartości szeregu lnHDD



Źródło: Wydruk programu STATA na podstawie danych pochodzących z Eurostatu

Tabela 2.6.1. Średnie błędy prognoz dla modelu SARIMA oraz modelu multiplikatywnego

	SARIMA(2,1,0)(0,1,1,12)	**Model multiplikatywny**
MAE	0,2317701	0,2171424
MSE	0,1479902	0,1104469
MAPE	0,0583166	0,0583166
AMAPE	0,0266377	0,0266377

Źródło: Opracowanie własne wyników otrzymanych w STATA na postawie danych z Eurostatu

PODSUMOWANIE

W pracy została przeprowadzona analiza dwóch szeregów czasowych - sezonowego oraz niesezonowego. Dokonana została dekompozycja szeregu, prognoza dla poszczególnego szeregu wybranym modelem ekstrapolacyjnym, dla każdego szeregu dopasowany został odpowiedni model z klasy ARIMA/SARIMA, z którego również wykonano prognozę. Następnie porównano wyniki prognoz z wybranych modeli. Zarówno dla szeregu sezonowego, jak i niesezonowego, model ekstrapolacyjny charakteryzował się lepszymi prognozami (w przypadku danych niesezonowych był to model ekstrapolacyjny dla szeregu niesezonowego Holta-Wintersa, a dla danych sezonowych model ekstrapolacyjny multiplikatywny Holta-Wintersa).