

Uniwersytet Gdański  
Wydział zarządzania  
Informatyka i ekonometria – Big data

Kasper Kurzyński

Grupa: S42-11

Nr albumu: 252932

*„Proгноza PKB dla Finlandii”*



Gdańsk 2021

## Spis treści

1. Podstawowe informacje.....	3
2. Analiza wstępna.....	4
3. Test ADF.....	7
4. Test HEGY.....	10
5. Prognozy.....	11
6. Test CUSUM.....	16

## Podstawowe informacje

PKB jest jednym z najbardziej podstawowych wskaźników ekonomicznych. Według definicji, PKB to „suma wartości dodanej wszystkich dóbr końcowych i usług wytworzonych w gospodarce w ciągu roku”. Inaczej mówiąc, jest to suma wartości dodanej wytworzonej przez wszystkie podmioty gospodarcze danego kraju, niezależnie kto jest ich właścicielem. PKB jest miarą wielkości gospodarki. Jego wzrost lub spadek oraz dynamika tych ruchów determinuje miarę wzrostu gospodarczego.

Praca została poświęcona prognozowaniu PKB dla Finlandii. Jest to kraj znajdujący się na północy Europy, którego stolicą są Helsinki. Całkowita liczba ludności Finlandii wynosi 5 503 000 osób, natomiast powierzchnia kraju to 338 145 km<sup>2</sup>. Gęstość zaludnienia Finlandii kształtuje się na poziomie 16 osób/km<sup>2</sup>. Obowiązującą walutą w kraju jest euro.

Dane wykorzystane w projekcie są to dane kwartalne, które obejmują wartości PKB rozpoczynające się w pierwszym kwartale 1990 roku i kończą się w czwartym kwartale 2020 roku.

## Analiza wstępna

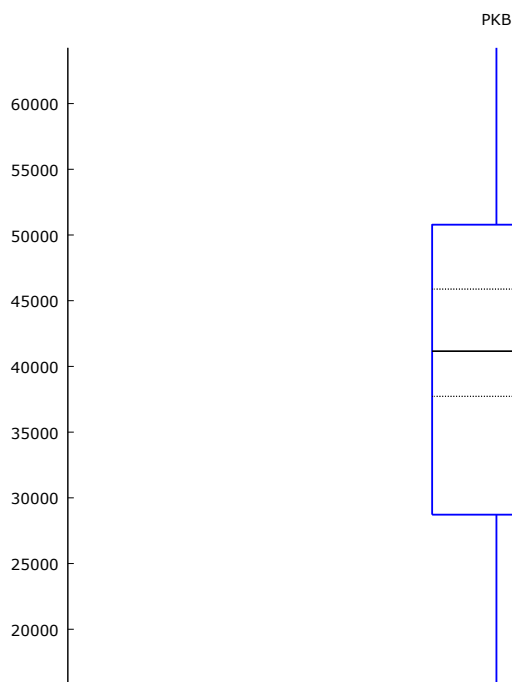
Analiza wstępna ma na celu podstawową interpretację statystyk opisowych, które dotyczą PKB dla Finlandii. W analizie wstępnej umieszczony został również wykres pudełkowy, histogram oraz wykres szeregu czasowego. Dodatkowo, przeprowadzony został test Shapiro - Wilka na normalność rozkładu analizowanej zmiennej.

### Statystyki opisowe:

- Średnia: 40614,
- Mediana: 41161,
- Minimalna: 17573,
- Maksymalna: 62446,
- Odchylenie standardowe: 12401,
- Współczynnik zmienności: 31%
- Współczynnik skośności: -0,092913
- Kurtoza: -1,2065

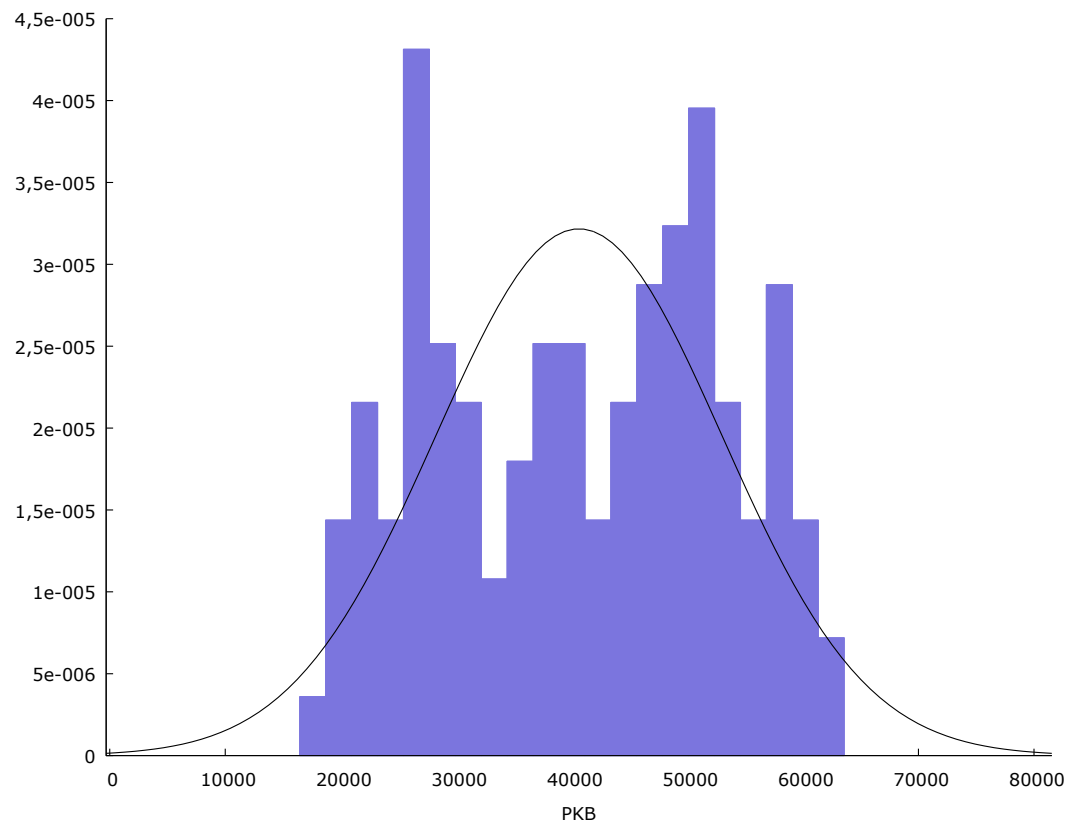
Średnia wartość PKB Finlandii w badanym okresie wyniosła 40614 milionów euro, natomiast mediana 41161 milionów euro. Najmniejsza wartość PKB ukształtowała się na poziomie 17573 milionów euro i wystąpiła w 1 kwartale 1993 roku. Z kolei najwyższa wartość PKB równa 62446 milionów euro została osiągnięta w czwartym kwartale 2019 roku. Odchylenie standardowe dla PKB Finlandii wyniosło 12401 milionów euro, natomiast współczynnik zmienności z dokładnością do 1% wyniósł 31%. Współczynnik skośności ukształtował się na poziomie -0,092913, co oznacza, że rozkład PKB posiada lewostronną asymetrię (lewe ramię rozkładu jest wydłużone). Wartość kurtozy z kolei wyniosła -1,2065, więc jest to rozkład platykurtyczny. Taki wynik kurtozy świadczy o większym spłaszczeniu rozkładu.

## Wykres pudełkowy



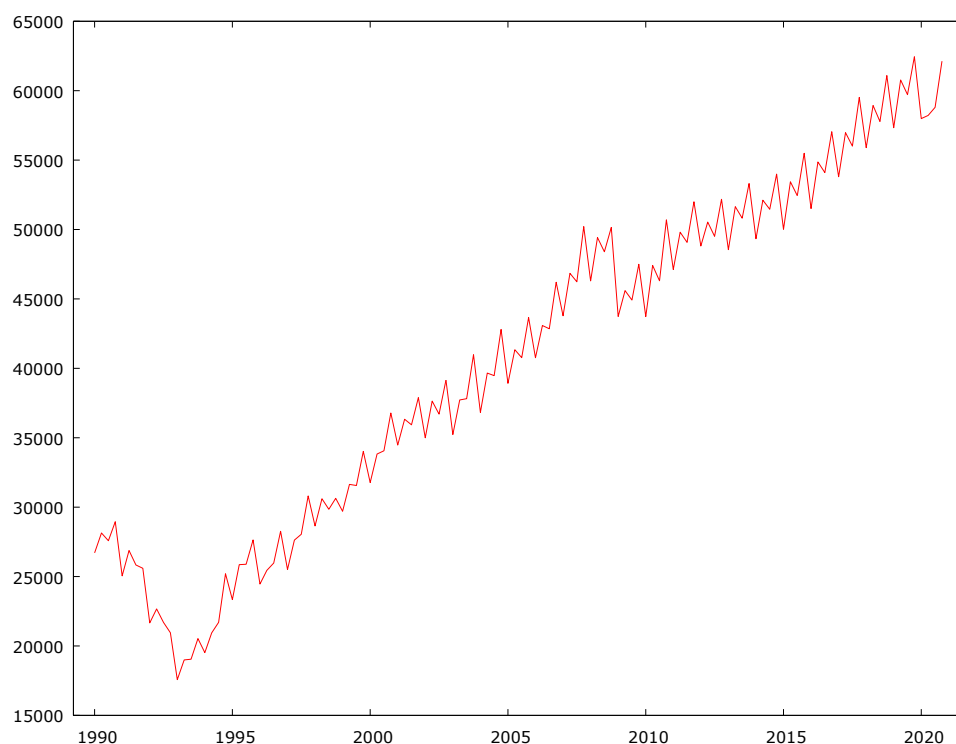
Wykres pudełkowy prezentuje rozproszenie wartości PKB wraz z medianą, która została oznaczona czarną linią.

## Histogram



Na histogramie widoczny jest rozkład częstości dla danych wartości PKB. Dodatkowo, za pomocą czarnej linii została pokazana zgodność z rozkładem normalnym.

### Wykres szeregu czasowego dla PKB Finlandii



Na powyższym wykresie widoczny jest trend liniowy, co może oznaczać, że proces PKB jest szeregiem niestacjonarnym. W celu potwierdzenia tego przypuszczenia wykonany zostanie test ADF.

Test Shapiro-Wilka = 0,954378, z wartością p 0,0003592

Wniosek: Należy odrzucić hipotezę zerową na korzyść hipotezy alternatywnej. PKB dla Finlandii w badanym okresie nie jest zgodne z rozkładem normalnym.

## Test ADF

Celem testu jest sprawdzenie czy szereg czasowy PKB jest stacjonarny oraz czy należy dokonać różnicowania, aby doprowadzić szereg do stacjonarności. Testowano istotność opóźnienia od rzędu 12 dla kryterium AIC. Liczebność próby w teście wynosiła: 111.

*Hipoteza zerowa: występuje pierwiastek jednostkowy*

- I. test bez wyrazu wolnego (const)  
dla opóźnienia rzędu 12 procesu  $(1-L)PKB$   
model:  $(1-L)y = (a-1)*y(-1) + \dots + e$   
estymowana wartość  $(a-1)$  wynosi: 0,00832075  
Statystyka testu:  $\tau_{nc}(1) = 2,93269$   
asymptotyczna wartość  $p = 0,9993$   
Autokorelacja reszt rzędu pierwszego: -0,017  
opóźnione różnice:  $F(12, 98) = 83,615 [0,0000]$
- II. test z wyrazem wolnym (const)  
dla opóźnienia rzędu 12 procesu  $(1-L)PKB$   
model:  $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$   
estymowana wartość  $(a-1)$  wynosi: -0,00840586  
Statystyka testu:  $\tau_c(1) = -1,22043$   
asymptotyczna wartość  $p = 0,668$   
Autokorelacja reszt rzędu pierwszego: -0,044  
opóźnione różnice:  $F(12, 97) = 86,260 [0,0000]$
- III. z wyrazem wolnym i trendem liniowym  
dla opóźnienia rzędu 12 procesu  $(1-L)PKB$   
model:  $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$   
estymowana wartość  $(a-1)$  wynosi: -0,14617  
Statystyka testu:  $\tau_{ct}(1) = -2,43587$   
asymptotyczna wartość  $p = 0,3607$   
Autokorelacja reszt rzędu pierwszego: -0,048  
opóźnione różnice:  $F(12, 96) = 56,802 [0,0000]$

### Interpretacja wyników:

W każdym z trzech wariantów testu (bez wyrazu wolnego, z wyrazem wolnym, z wyrazem wolnym i trendem liniowym) wartość **p-value** nie znalazła się poniżej przyjętego poziomu istotności = 5%. Regułą decyzyjną w takim wypadku jest brak podstaw do odrzucenia hipotezy zerowej. Oznacza to, że w danym szeregu czasowym występuje pierwiastek jednostkowy i proces jest niestacjonarny. Należy dokonać różnicowania w celu uzyskania stacjonarności. Zaznaczone na czerwono wartości odnoszą się do **współczynnika autokorelacji reszt**. Ich wartość jest mniejsza niż 0,1, więc w modelu pomocniczym nie występuje statystycznie istotna autokorelacja składnika zakłócającego.

### Test ADF dla pierwszych różnic procesu PKB:

*Hipoteza zerowa: występuje pierwiastek jednostkowy*

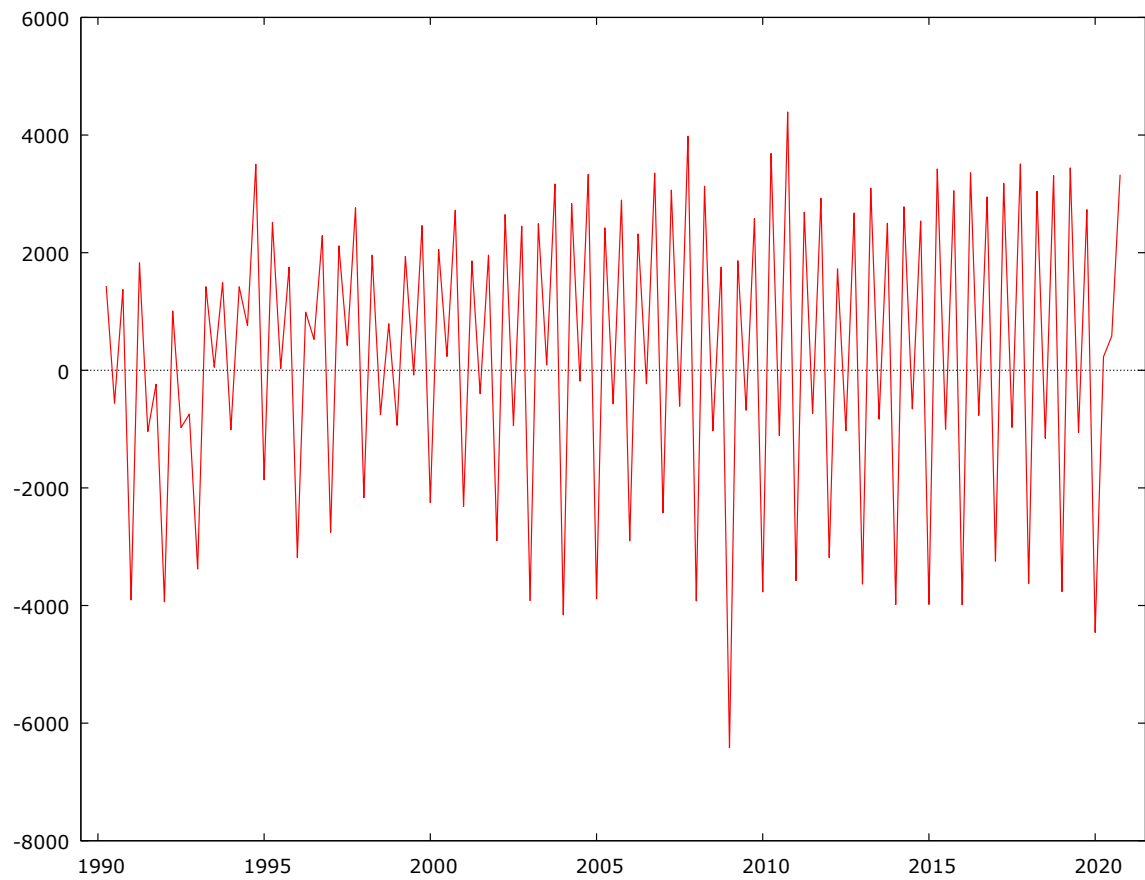
- I. test bez wyrazu wolnego (const)  
dla opóźnienia rzędu 11 procesu  $(1-L)d\_PKB$   
model:  $(1-L)y = (a-1)*y(-1) + \dots + e$   
estymowana wartość  $(a-1)$  wynosi: -0,432092  
Statystyka testu:  $\tau_{nc}(1) = -2,18437$   
asymptotyczna wartość  $p = 0,02787$   
Autokorelacja reszt rzędu pierwszego: -0,021  
opóźnione różnice:  $F(11, 99) = 34,805 [0,0000]$
- II. test z wyrazem wolnym (const)  
dla opóźnienia rzędu 8 procesu  $(1-L)d\_PKB$   
model:  $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$   
estymowana wartość  $(a-1)$  wynosi: -1,04638  
Statystyka testu:  $\tau_c(1) = -4,69833$   
asymptotyczna wartość  $p = 8,113e-005$   
Autokorelacja reszt rzędu pierwszego: -0,010  
opóźnione różnice:  $F(8, 104) = 45,187 [0,0000]$
- III. z wyrazem wolnym i trendem liniowym  
dla opóźnienia rzędu 11 procesu  $(1-L)d\_PKB$   
model:  $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$   
estymowana wartość  $(a-1)$  wynosi: -1,22075  
Statystyka testu:  $\tau_{ct}(1) = -4,21238$   
asymptotyczna wartość  $p = 0,004219$   
Autokorelacja reszt rzędu pierwszego: -0,042  
opóźnione różnice:  $F(11, 97) = 33,837 [0,0000]$

### Interpretacja wyników:

W przypadku wykonania testu ADF dla pierwszych różnic procesu PKB wartość **p-value** okazała się mniejsza niż przyjęty poziom istotności 5%. Oznacza to, że należy odrzucić hipotezę zerową na korzyść hipotezy alternatywnej. W pierwszych różnicach procesu PKB nie występuje pierwiastek jednostkowy, a więc proces jest szeregiem czasowym stacjonarnym. Zaznaczone na czerwono wartości odnoszą się do **współczynnika autokorelacji reszt**. Ich wartość jest mniejsza niż 0,1, więc w modelu pomocniczym nie występuje statystycznie istotna autokorelacja składnika zakłócającego.



**Wykres szeregu czasowego dla pierwszych różnic PKB:**



Wygląd szeregu czasowego zaprezentowanego na wykresie potwierdza wyniki testu ADF dla pierwszych różnic PKB. Proces jest szeregiem stacjonarnym.

## Test Hegy

Test Hegy został wykonany w celu sprawdzenia istnienia sezonowego pierwiastka jednostkowego. Liczba opóźnień w teście wyniosła 4, ponieważ są to dane kwartalne. Komponentem deterministycznym w teście była stała.

### Wyniki testu:

Statistic	p-value	Ang. Frequency	Period
t1= -0,09	0,94241	zero	infinity
F1= 1,20	0,31200	+pi/2	4
t2= -0,34	0,47797	pi	2
Fs= 0,84	0,51295	All the seasonal cycles	
Ft= 0,63	0,87475	Delta_s (all the seas. + zero freq.)	

### Regresja pomocnicza:

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	314,628	284,598	1,106	0,2714
y0_1	-0,000150257	0,00172528	-0,08709	0,9308
ya1_1	0,0137337	0,0531851	0,2582	0,7967
yb1_1	0,0811829	0,0532863	1,524	0,1306
ypi_1	-0,00593658	0,0176444	-0,3365	0,7372
sd_y_1	0,912866	0,103551	8,816	2,42e-014 ***
sd_y_2	0,0279690	0,135844	0,2059	0,8373
sd_y_3	-0,0471188	0,143868	-0,3275	0,7439
sd_y_4	-0,179774	0,100860	-1,782	0,0775 *

Średn.aryt.zm.zależnej 1153,268 Odch.stand.zm.zależnej 1919,050  
 Suma kwadratów reszt 77338269 Błąd standardowy reszt 850,1691  
 Wsp. determ. R-kwadrat 0,817390 Skorygowany R-kwadrat 0,803737  
 F(8, 107) 59,86862 Wartość p dla testu F 4,88e-36  
 Logarytm wiarygodności -942,3832 Kryt. inform. Akaike'a 1902,766  
 Kryt. bayes. Schwarza 1927,549 Kryt. Hannana-Quinna 1912,827  
 Autokorel.reszt - rho1 -0,001661 Stat. Durbina-Watsona 1,989411

### Interpretacja wyników:

Test Hegy wykazał, że w przypadku  $H_a$ ,  $H_b$  oraz  $H_c$  nie ma podstaw do odrzucenia  $H_0$ . Oznacza to, że w modelu SARIMA duży parametr „D” powinien wynosić 1. Należy tylko raz różnicować, aby doprowadzić szereg do stacjonarności sezonowej.

## Prognozowanie

W celu dobrania odpowiedniego modelu do prognozowania zostało oszacowanych kilka modeli SARIMA. Wśród modeli, które charakteryzowały się najlepszymi właściwościami pozwalającymi na przeprowadzenie dokładniejszej prognozy były modele:

- SARIMA (2 1 2 ; 1 1 2)
- SARIMA (1 1 1 ; 1 1 1)
- SARIMA (2 1 0; 1 1 2)

### ARIMA (2 1 2 ; 1 1 2)

współczynnik	błąd standardowy	z	wartość p
const	13,1310	12,6741	1,036 0,3002
phi_1	-0,335623	0,218854	-1,534 0,1251
phi_2	0,654605	0,219825	2,978 0,0029 ***
Phi_1	-0,691488	0,110522	-6,257 3,94e-010 ***
theta_1	0,422421	0,247928	1,704 0,0884 *
theta_2	-0,388824	0,232847	-1,670 0,0949 *
Theta_1	-0,0511373	0,132520	-0,3859 0,6996
Theta_2	-0,948862	0,128890	-7,362 1,81e-013 ***

Średn.aryt.zm.zależnej 11,25546 Odch.stand.zm.zależnej 950,5088  
 Średnia zaburzeń los. 63,62966 Odch.st. zaburzeń los. 735,1808  
 Logarytm wiarygodności -960,7690 Kryt. inform. Akaike'a 1939,538  
 Kryt. bayes. Schwarza 1964,550 Kryt. Hannana-Quinna 1949,695

	część Rzeczywista	Urojona	Moduł	Częstość
AR				
Pierwiastek 1	-1,0059	0,0000	1,0059	0,5000
Pierwiastek 2	1,5186	0,0000	1,5186	0,0000
AR (sezonowe)				
Pierwiastek 1	-1,4462	0,0000	1,4462	0,5000
MA				
Pierwiastek 1	-1,1500	0,0000	1,1500	0,5000
Pierwiastek 2	2,2364	0,0000	2,2364	0,0000
MA (sezonowe)				
Pierwiastek 1	-1,0539	0,0000	1,0539	0,5000
Pierwiastek 2	1,0000	0,0000	1,0000	0,0000

### SARIMA (1 1 1 ; 1 1 1)

	współczynnik	błąd standardowy	z	wartość p	
const	17,6866	33,4757	0,5283	0,5973	
phi_1	0,607334	0,203472	2,985	0,0028	***
Phi_1	0,214561	0,138030	1,554	0,1201	
theta_1	-0,444971	0,218554	-2,036	0,0418	**
Theta_1	-0,778728	0,0832217	-9,357	8,18e-021	***

Średn.aryt.zm.zależnej 11,25546 Odch.stand.zm.zależnej 950,5088  
Średnia zaburzeń los. 17,20998 Odch.st. zaburzeń los. 814,1303  
Logarytm wiarygodności -967,5995 Kryt. inform. Akaike'a 1947,199  
Kryt. bayes. Schwarza 1963,874 Kryt. Hannana-Quinna 1953,970

	część Rzeczywista	Urojona	Moduł	Częstość
AR				
Pierwiastek 1	1,6465	0,0000	1,6465	0,0000
AR (sezonowe)				
Pierwiastek 1	4,6607	0,0000	4,6607	0,0000
MA				
Pierwiastek 1	2,2473	0,0000	2,2473	0,0000
MA (sezonowe)				
Pierwiastek 1	1,2841	0,0000	1,2841	0,0000

### SARIMA (2 1 0 ; 1 1 2)

	współczynnik	błąd standardowy	z	wartość p	
const	20,5560	30,2786	0,6789	0,4972	
phi_1	0,0901073	0,0942945	0,9556	0,3393	
phi_2	0,203379	0,0991687	2,051	0,0403	**
Phi_1	-0,682715	0,113653	-6,007	1,89e-09	***
Theta_1	0,179697	0,126235	1,424	0,1546	
Theta_2	-0,731317	0,0971270	-7,529	5,09e-014	***

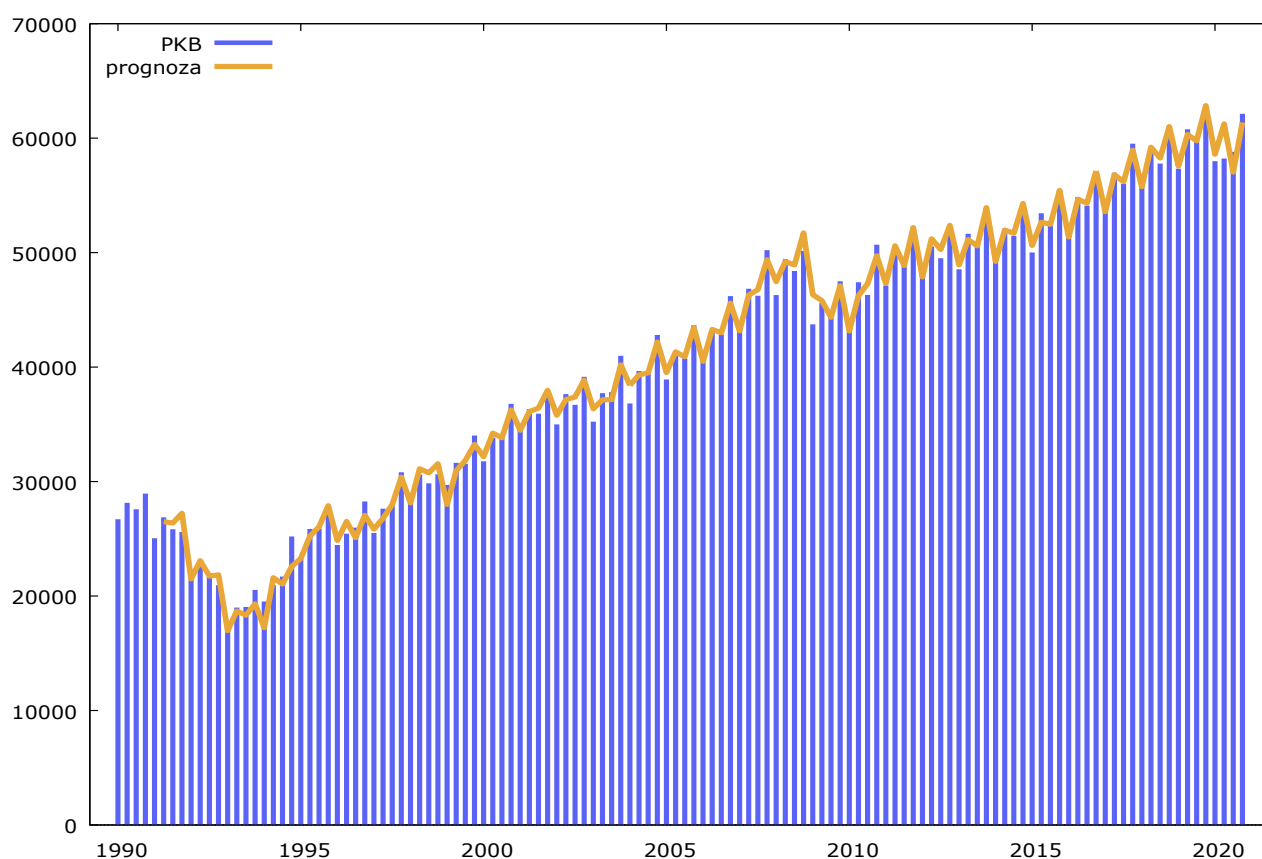
Średn.aryt.zm.zależnej 11,25546 Odch.stand.zm.zależnej 950,5088  
Średnia zaburzeń los. 17,40757 Odch.st. zaburzeń los. 774,8586  
Logarytm wiarygodności -963,4854 Kryt. inform. Akaike'a 1940,971  
Kryt. bayes. Schwarza 1960,425 Kryt. Hannana-Quinna 1948,870

	część Rzeczywista	Urojona	Moduł	Częstość
AR				
Pierwiastek 1	2,0069	0,0000	2,0069	0,0000
Pierwiastek 2	-2,4500	0,0000	2,4500	0,5000
AR (sezonowe)				
Pierwiastek 1	-1,4647	0,0000	1,4647	0,5000
MA (sezonowe)				
Pierwiastek 1	-1,0529	0,0000	1,0529	0,5000
Pierwiastek 2	1,2987	0,0000	1,2987	0,0000

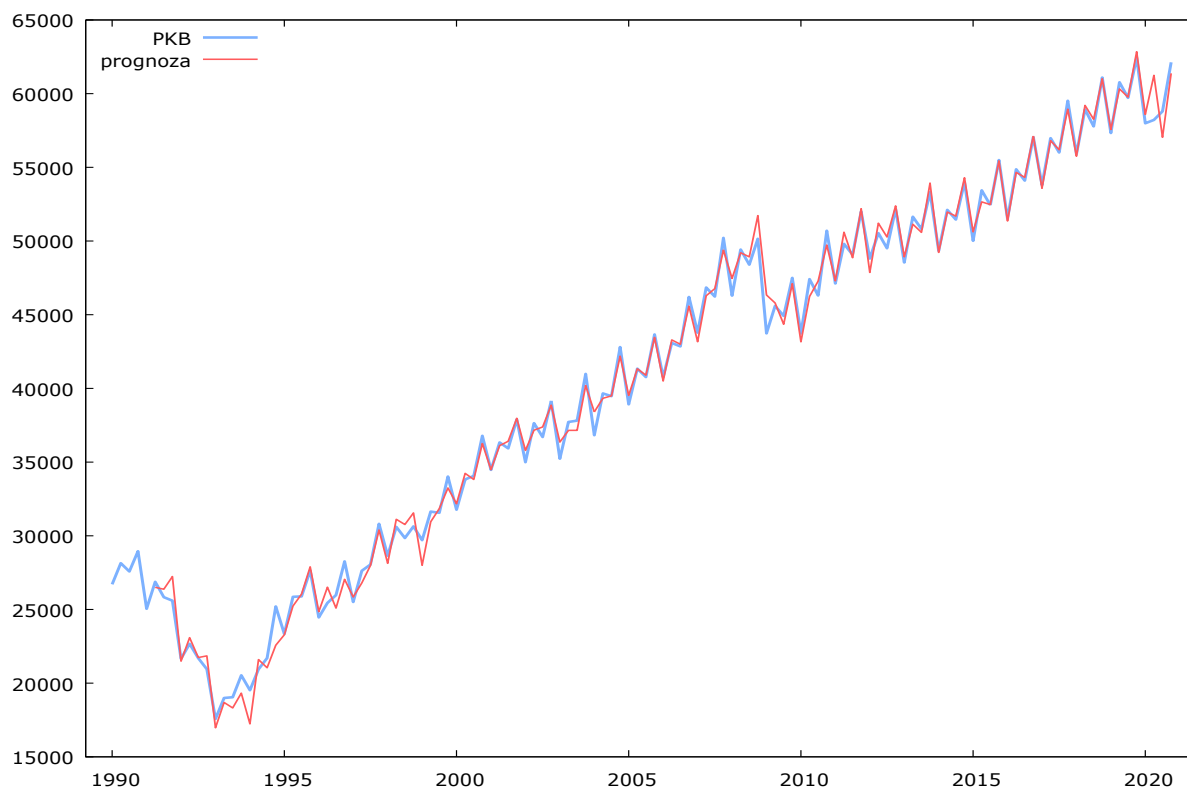
Model  $SARIMA(2\ 1\ 0; 1\ 1\ 2)$  charakteryzował się najniższą wartością kryterium informacyjnego Akaike ( $AIC = 1940,971$ ) w przypadku, gdy wszystkie moduły były większe od 1 i leżały na zewnątrz koła jednostkowego. Model  $SARIMA(1\ 1\ 1; 1\ 1\ 1)$  posiadał najwięcej istotnych współczynników, ponieważ aż 3 współczynniki na 4 możliwe wykazały istotność statystyczną na poziomie istotności 5%. Model  $SARIMA(2\ 1\ 2; 1\ 1\ 2)$  z kolei posiadał najniższą wartość kryterium informacyjnego Akaike ( $AIC = 1939,538$ ), lecz jeden z modułów w tym modelu wyniósł 1, co oznacza, że nie leży on na zewnątrz koła jednostkowego. Ostatecznie do prognozy został wykorzystany model  $SARIMA(2\ 1\ 0; 1\ 1\ 2)$ . Jego niskie kryterium informacyjne Akaike ( $AIC = 1940,971$ ) oraz 3 istotne współczynniki na 5 możliwych zadecydowało o podjęciu takiej decyzji.

### Prognoza na podstawie modelu $SARIMA(2\ 1\ 0; 1\ 1\ 2)$ :

Wykres I:



Wykres II:



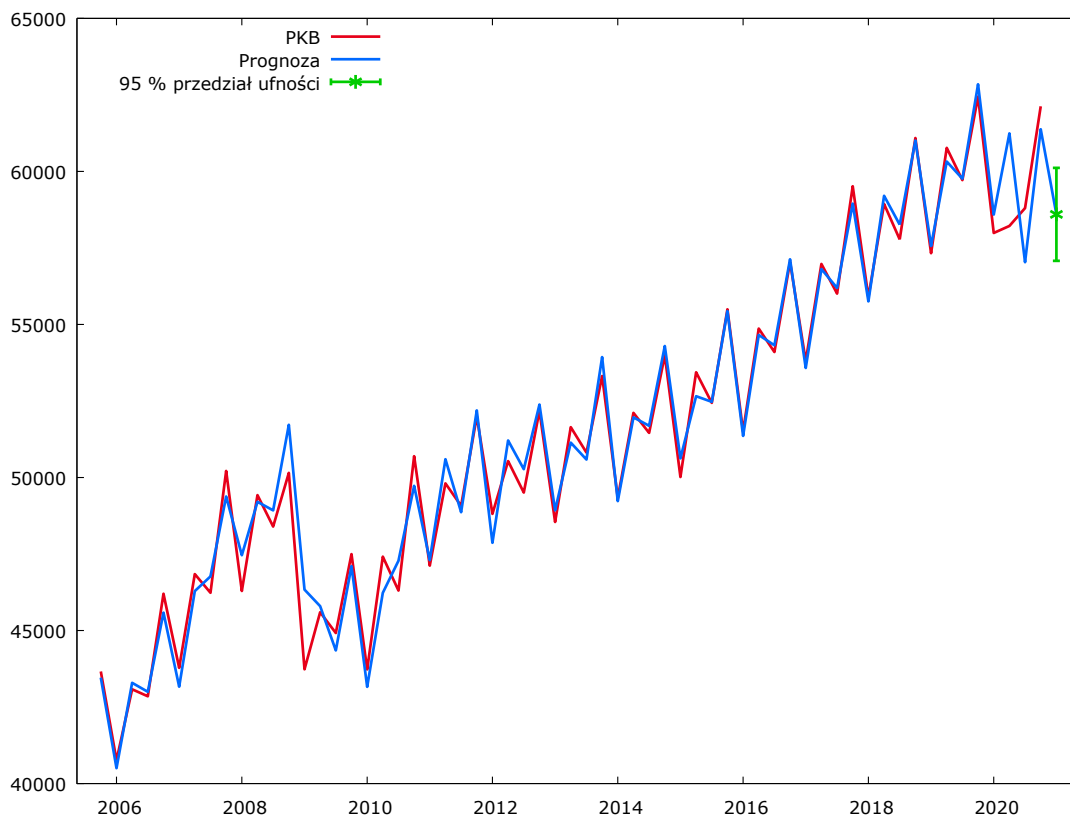
**Miary dokładności prognoz ex post:**

- Średni błąd predykcji  $ME = 17,408$   
Przeciętne odchylenie od prognozy wynosi 17,408 milionów euro.
- Pierwiastek błędu średniokwadr.  $RMSE = 797,71$   
Przeciętnie wartości prognozy PKB dla Finlandii odchylają się od wartości zrealizowanych o 791,71 milionów euro. Wartość została obliczona na podstawie kwadratów odchyleń (MSE).
- Względny błąd prognozy  $V = 1,96\%$   
Nie przekracza 5%, co świadczy o dopuszczalności prognozy.
- Średni błąd absolutny  $MAE = 585,96$
- Średni błąd procentowy  $MPE = 0,16918$
- Średni absolutny błąd procentowy  $MAPE = 1,6729$   
Wartości prognoz odchylają się od wartości zmiennej prognozowanej o +/- 585,96 milionów euro, co stanowi 1,6729% i jest to dopuszczalna wartość dla prognozy (>5%).
- Współczynnik Theila (w %)  $U^2 = 0,39169$   
Współczynnik Theila na poziomie 0,39169 jest niższy niż 1. Świadczy o tym, że prognoza jest lepiej dopasowana niż prognoza naiwna. Błąd prognozy stanowi 40% wartości błędu prognozy naiwnej.

Pierwsza dekompozycja Theila:

- Udział obciążoności predykcji  $I1^2/MSE = 0,00047619$   
Błąd prognozy w 0,04% spowodowany jest obciążeniem prognozy.
- Udział niedost. elastyczności  $I2^2/MSE = 0,034615$   
W 3,5% spowodowany jest niedostosowaniem elastyczności prognoz.
- Udział niezgodności kierunku  $I3^2/MSE = 0,96491$   
W 96,% błąd prognozy spowodowany jest nieprawidłowym prognozowaniem punktów zwrotnych.

#### Prognoza ex ante:

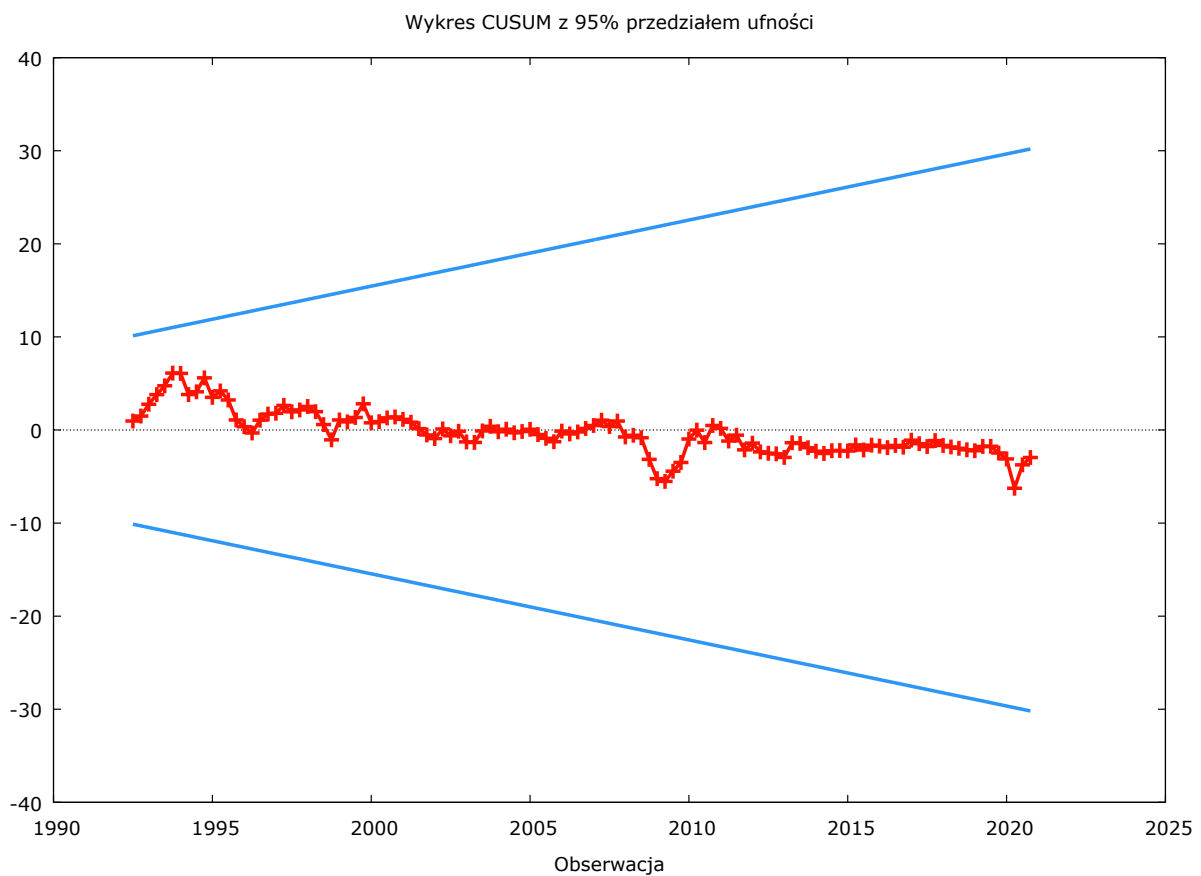


- Prognoza na 2021:01 = 58596,63
- Błąd ex ante = 774,859
- 95% przedział ufności = 57077,94 - 60115,33

Prognoza na pierwszy kwartał 2021 wyniosła 58596,63 mln euro, ze średnim błędem +- 774,859 mln euro. Względny błąd prognozy (V) wynosi:  $774,859/58596,63 * 100 = 1,3\%$ . Dana wartość stanowi 1,3% wartości prognozy i tym samym przekłada się na dopuszczalność powyższej prognozy. Przedział <57077,94 ; 60115,33> z 95% prawdopodobieństwem pokrywa nieznaną wartość PKB w pierwszym kwartale 2021 roku.

## Test CUSUM

Test CUSUM jest testem stabilności parametrów strukturalnych. Hipoteza zerowa w tym teście zakłada, że parametry modelu są stabilne. W przypadku, gdy empiryczny poziom istotności  $p$  statystyki testowej jest mniejszy niż przyjęty poziom istotności  $\alpha = 5\%$ , należy odrzucić hipotezę zerową na korzyść hipotezy alternatywnej zgodnie z którą w modelu nie występuje stabilność parametrów.



➤ Statystyka testu Harvey'a-Colliera  $t(113) = -0,276863$  z wartością  $p = 0,7824$ .

Wniosek: Brak podstaw do odrzucenia hipotezy zerowej. Model posiada stabilne parametry.