Paweł Maśluch

[WDEC] Wspomaganie Decyzji
Laboratorium 11B
Modele prognozowania (część B)

Do sprawozdania dołączam plik Lab11B_v4.txt, w którym znajduje się program SAS-owy realizujący zadania tego laboratorium.
Niniejsze sprawozdanie ogranicza się do przedstawienia wyników i ich omówienia.

Ogólna charakterystyka laboratorium

Celem laboratorium jest budowa modeli prognozowania bazujące na modelach ARIMA (Autoregressive Integrated Moving Average). W drugiej części laboratorium budowane będą modele z wykorzystaniem danych o nieznanej charakterystyce. Analizie podlegają 3 pliki (set1.xlsx, set2.xlsx, set3.xlsx).

Dane – wstępna analiza

1) Oblicz następujące statystyki dla dostępnych danych: średnia, min, max, oraz odchylenie standardowe.

Statystyki danych - set1

Procedura MEANS

Zmienna	Etykieta	Średnia	Minimum	Maksimum	Odch. std.
i	i	50.5000000	1.0000000	100.0000000	29.0114920
u	u	97.3651115	-92.2106537	291.7897183	79.3245826

Statystyki danych - set2

Procedura MEANS

Zmienna	Etykieta	Średnia	Minimum	Maksimum	Odch. std.
i	i	500.5000000	1.0000000	1000.00	288.8194361
u	u	6.6643889	5.9121101	7.4543301	0.2592152

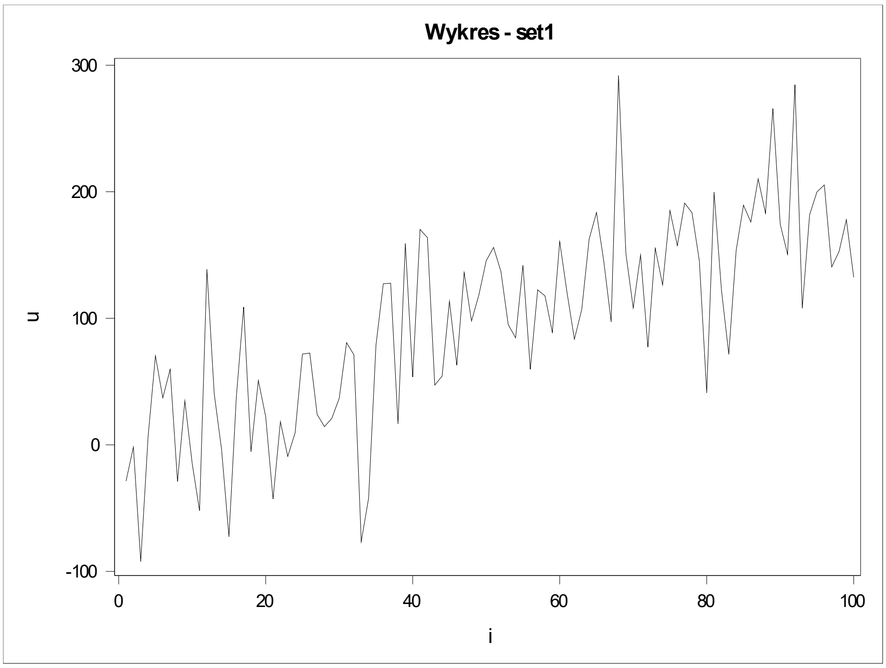
Statystyki danych - set3

Procedura MEANS

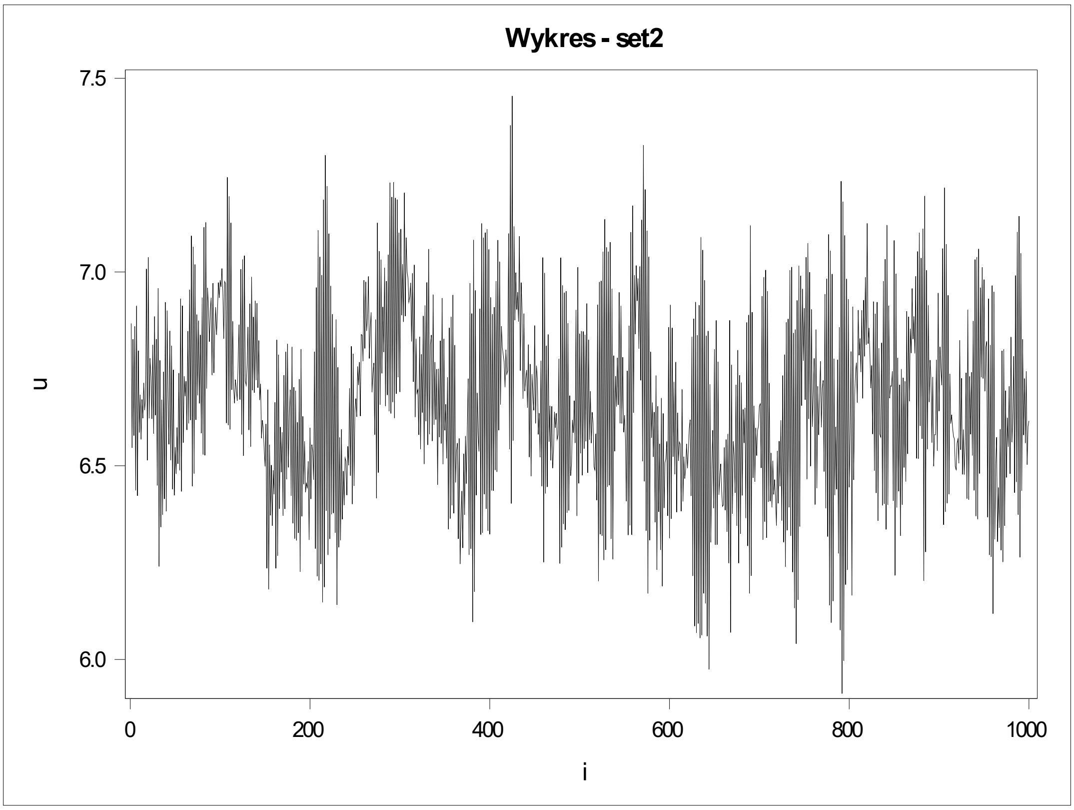
Zmienna analizowana: Total Total			
Średnia	Minimum	Maksimum	Odch. std.
280.2986111	104.0000000	622.0000000	119.9663169

2) <u>Narysuj wykresy dla dostępnych danych.</u>

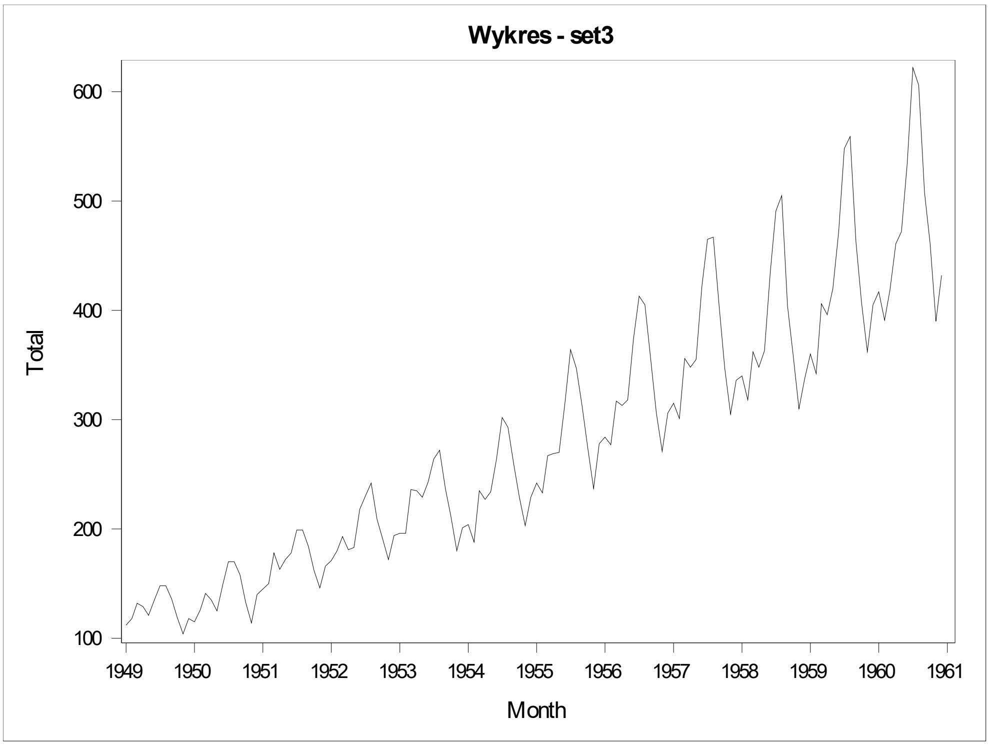
set1



set2



set3



3) Napisz czym charakteryzują się poszczególne zestawy danych. Jakie wnioski możesz wyciągnąć ze wstępnej analizy (w razie potrzeby policz inne charakterystyki danych).

set1
Na wykresie widzimy, że dany zestaw danych charakteryzuje się trendem wzrostowym.
Odchylenie standardowe jest dość duże.

set2
Wykres oscyluje wokół wartości ok. 6.65.
Niewielkie odchylenie standardowe.

set3
Z wykresu widać tendencję wzrostową danych oraz wzrastającą sezonowość danych.
Dość duże odchylenie standardowe.

Obliczenia modelu

1) <u>Dla dostarczonych danych zbuduj trzy modele ARIMA(p,d,q). Parametry p,d,q oraz parametry modelu należy znaleźć korzystając z procedury ARIMA.</u>

Postępujemy według następującego schematu:
a) na początku dokonujemy wstępnej identyfikacji modelu bez okresu różnicowania (d=0)
b) w zależności od wyników analizy z części a), dokonujemy odpowiednich analiz
z konkretnymi parametrami p,d,q (patrzymy na tabelę Testy wyboru porządku próbnego ARMA(p+d,q) oraz przyjmujemy w doborze wartości p>0, d>=0, q>=0),
przy czym ograniczamy się do co najwyżej 2-3 różnych modeli dla każdego zbioru danych.

set1

Testy wyboru porządku próbnego ARMA(p+d,q)	
SCAN	
p+d	q
1	1
5	0

(Poziom istotności 5%)

W dalszych analizach będę rozważał modele ARIMA(p=4,d=1,q=0), ARIMA(p=3,d=2,q=0), ARIMA(p=1,d=0,q=1).

set2

Testy wyboru porządku próbnego ARMA(p+d,q)	
SCAN	
p+d	q
2	2

(Poziom istotności 5%)

W dalszych analizach będę rozważał modele ARIMA(p=1,d=1,q=2), ARIMA(p=2,d=0,q=2).

set3

Testy wyboru porządku próbnego ARMA(p+d,q)	
SCAN	
p+d	q
1	4
3	3
4	2
5	0

(Poziom istotności 5%)

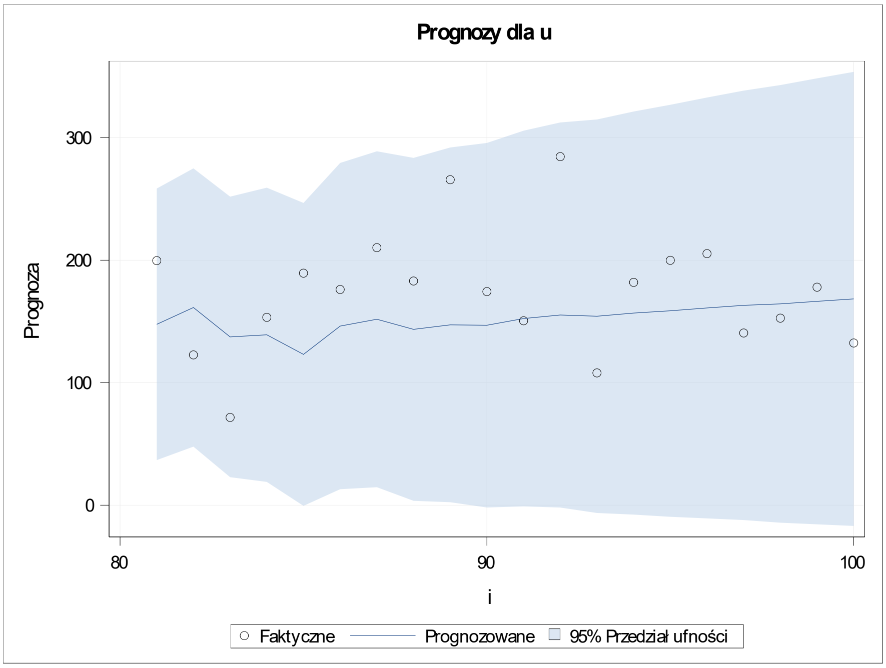
W dalszych analizach będę rozważał modele ARIMA(p=1,d=0,q=4), ARIMA(p=2,d=1,q=3).

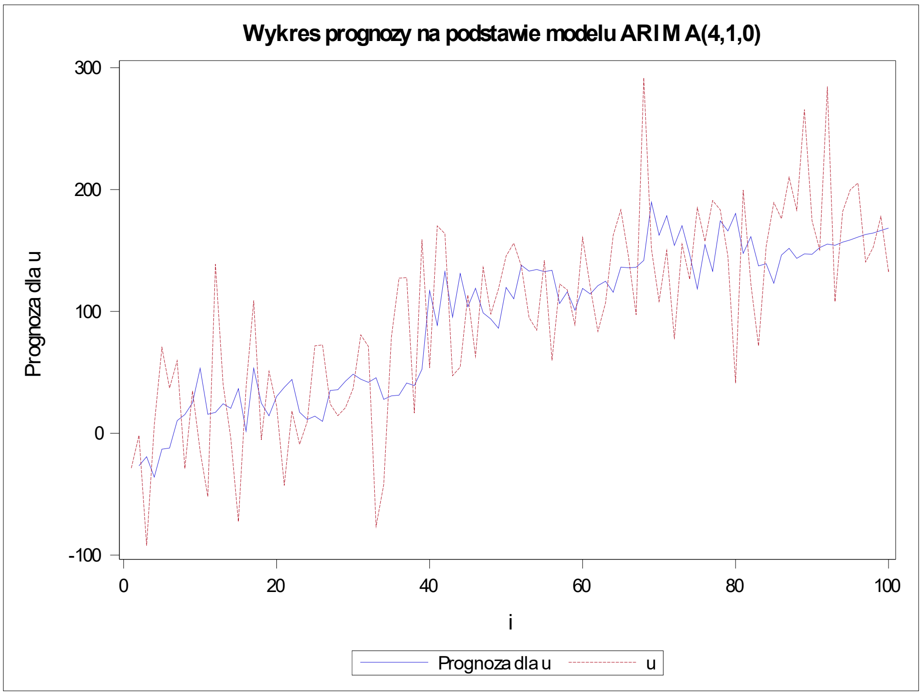
2) <u>Oblicz prognozy na podstawie znalezionych modeli.</u>

set1

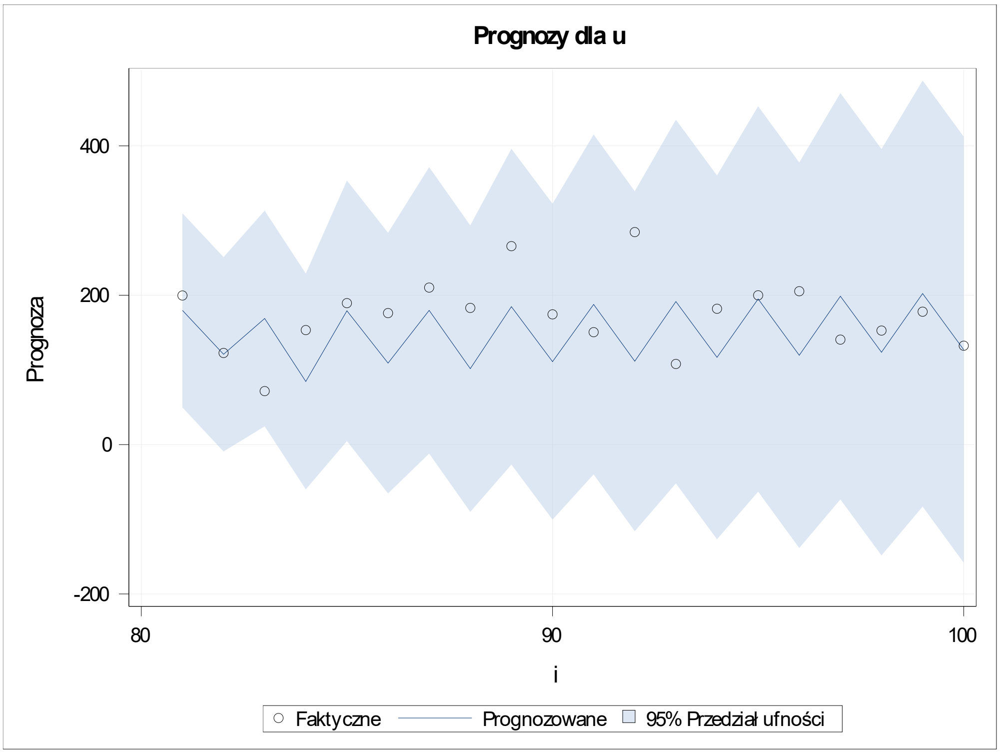
Uwaga. Obliczam prognozy porównawcze dla ostatnich 20 wartości naszych danych.

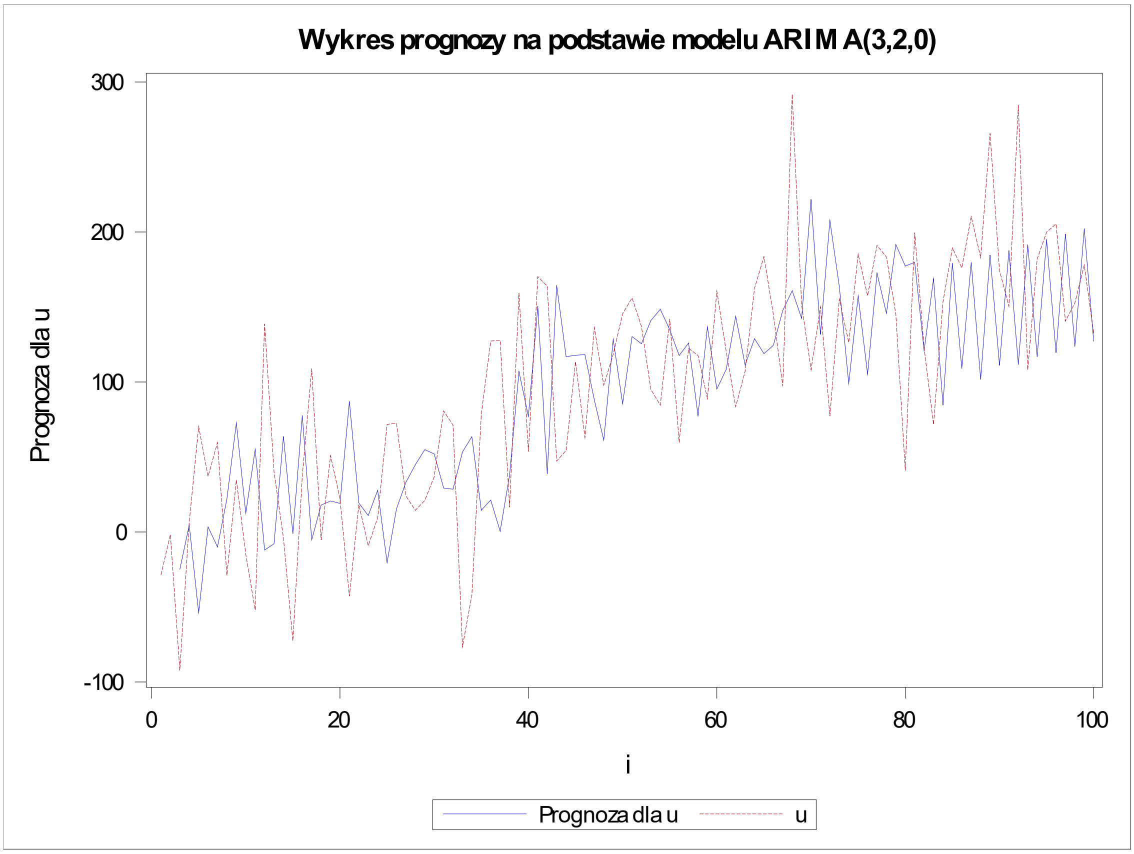
Model ARIMA(p=4,d=1,q=0)



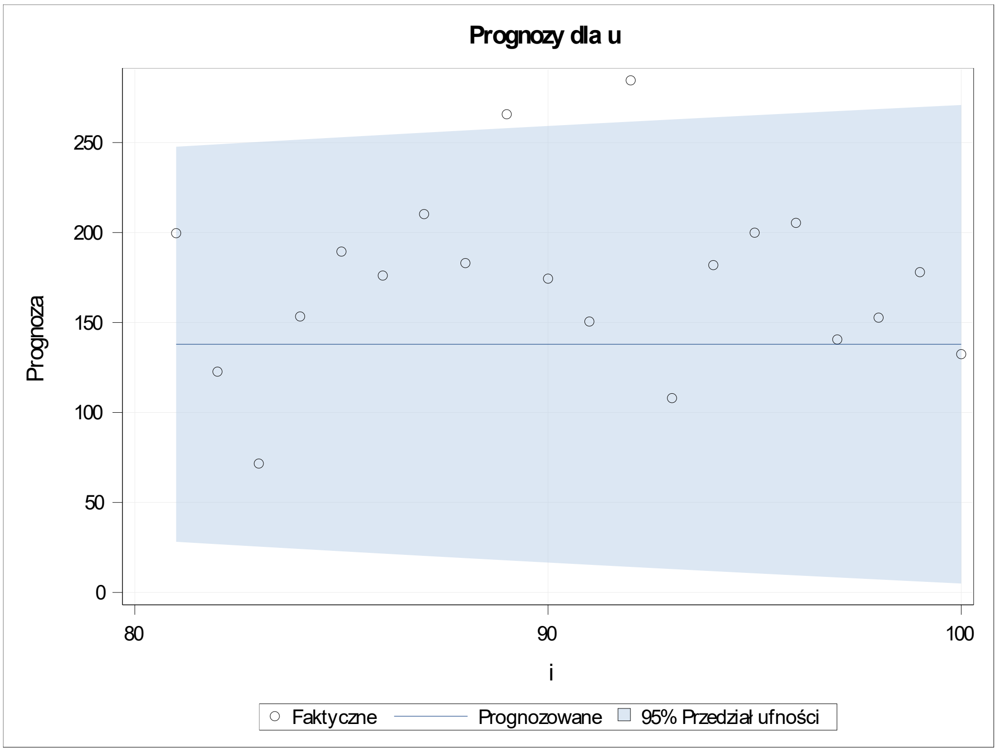


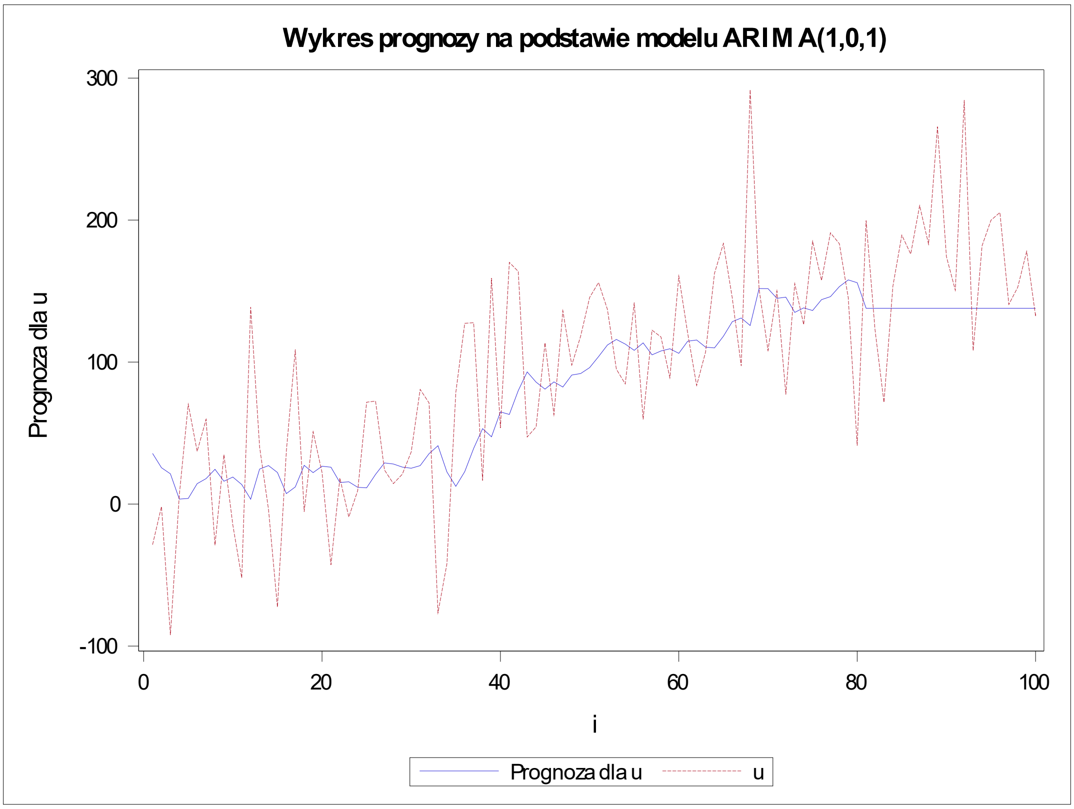
Model ARIMA(p=3,d=2,q=0)





Model ARIMA(p=1,d=0,q=1)

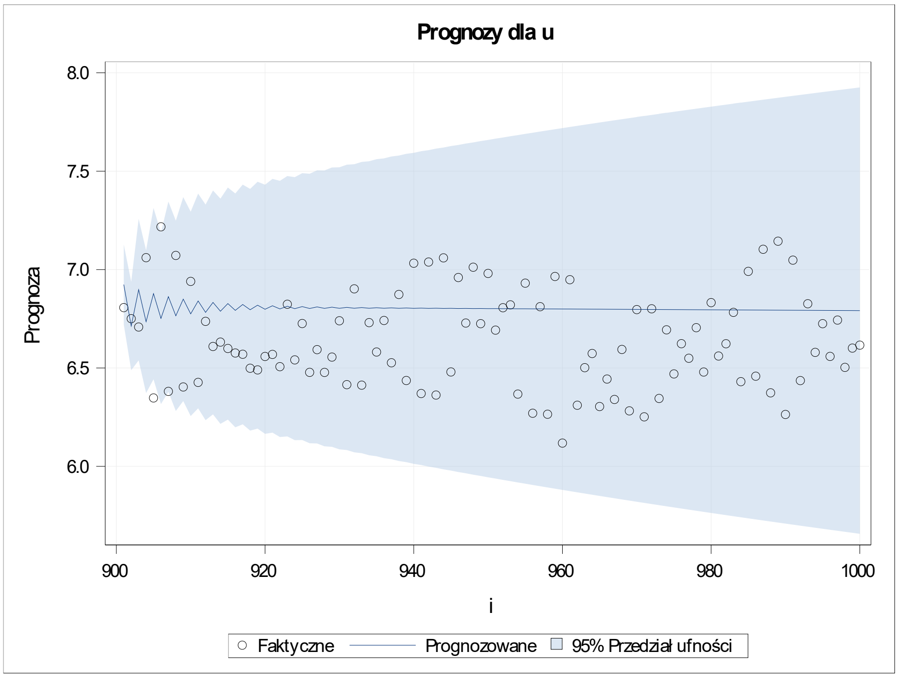


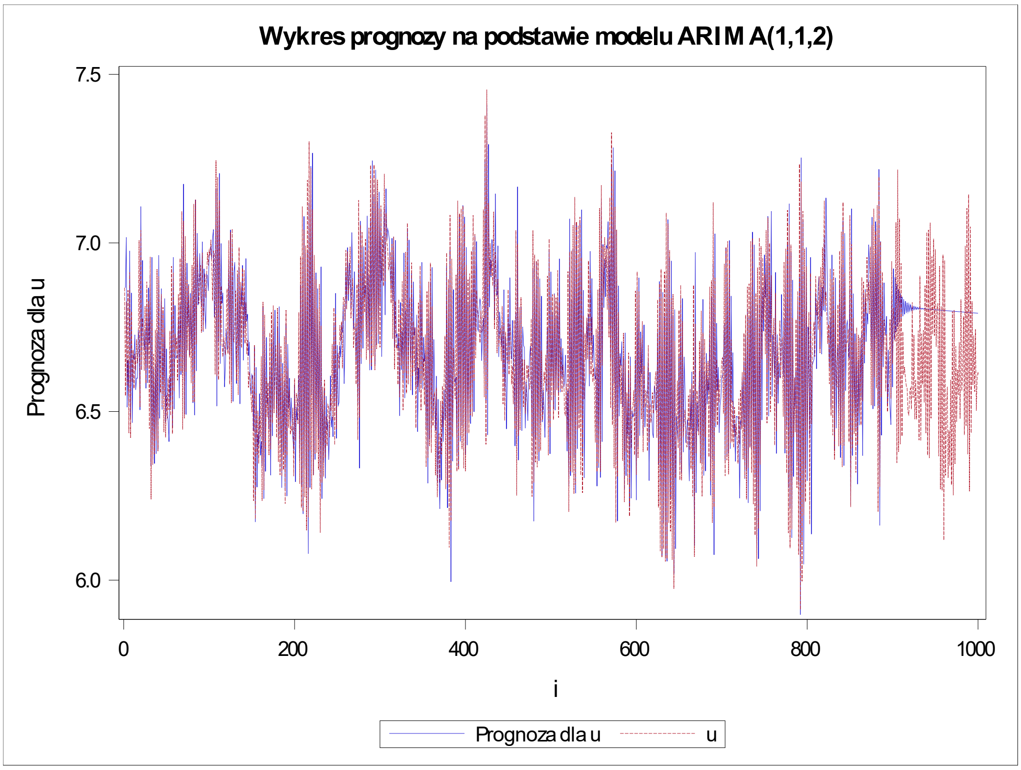


set2

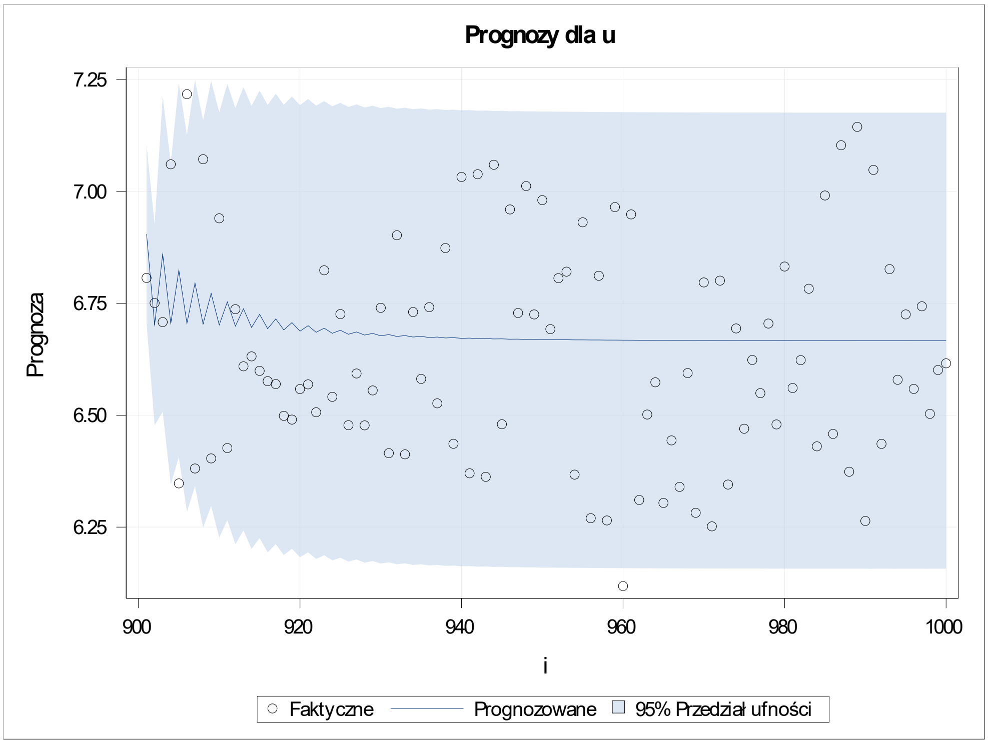
Uwaga. Obliczam prognozy porównawcze dla ostatnich 100 wartości naszych danych.

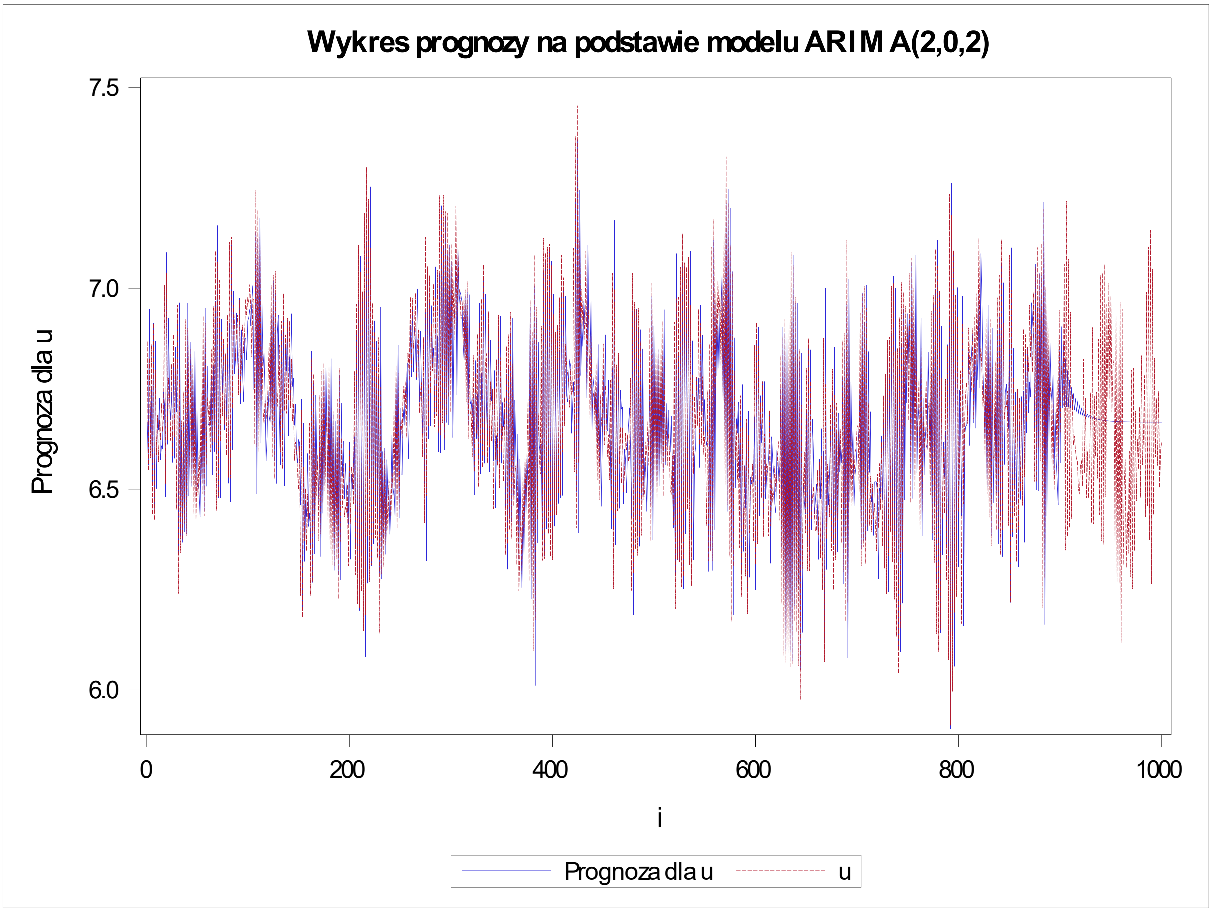
Model ARIMA(p=1,d=1,q=2)





Model ARIMA(p=2, d=0, q=2)

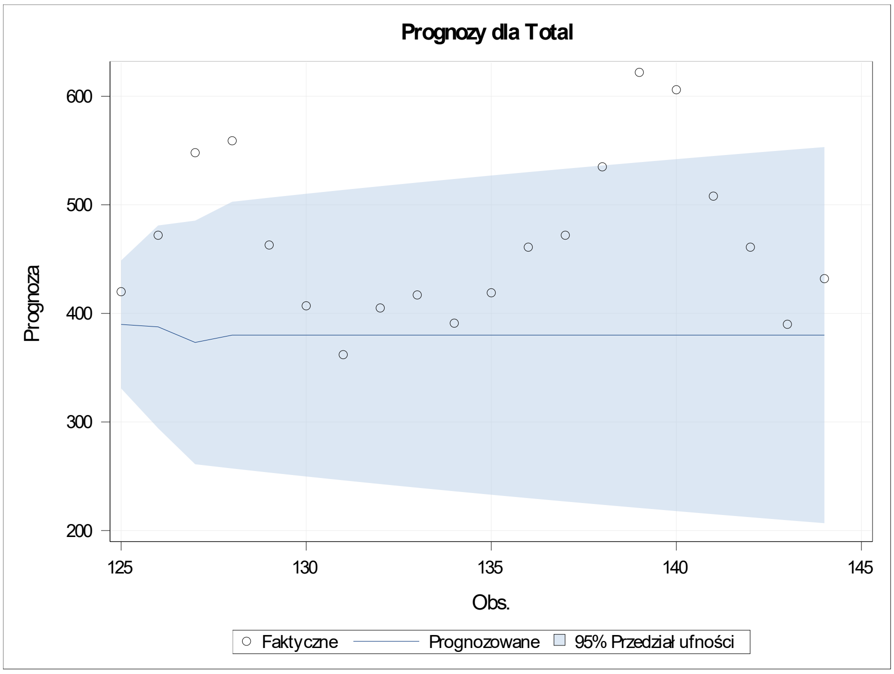


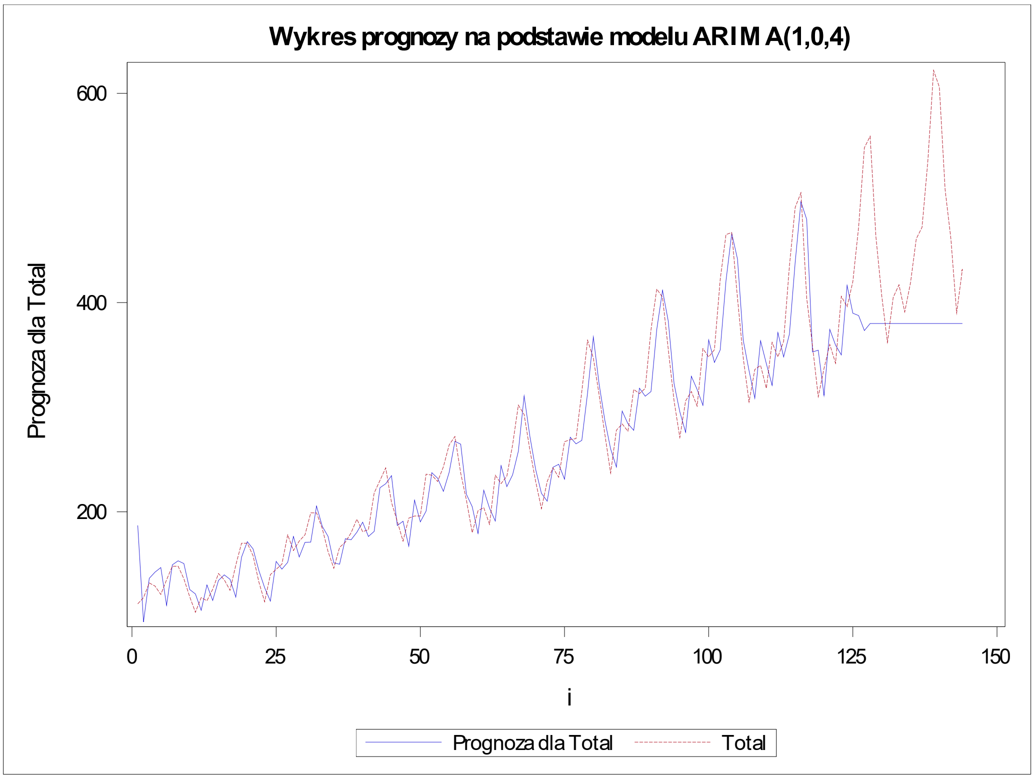


set3

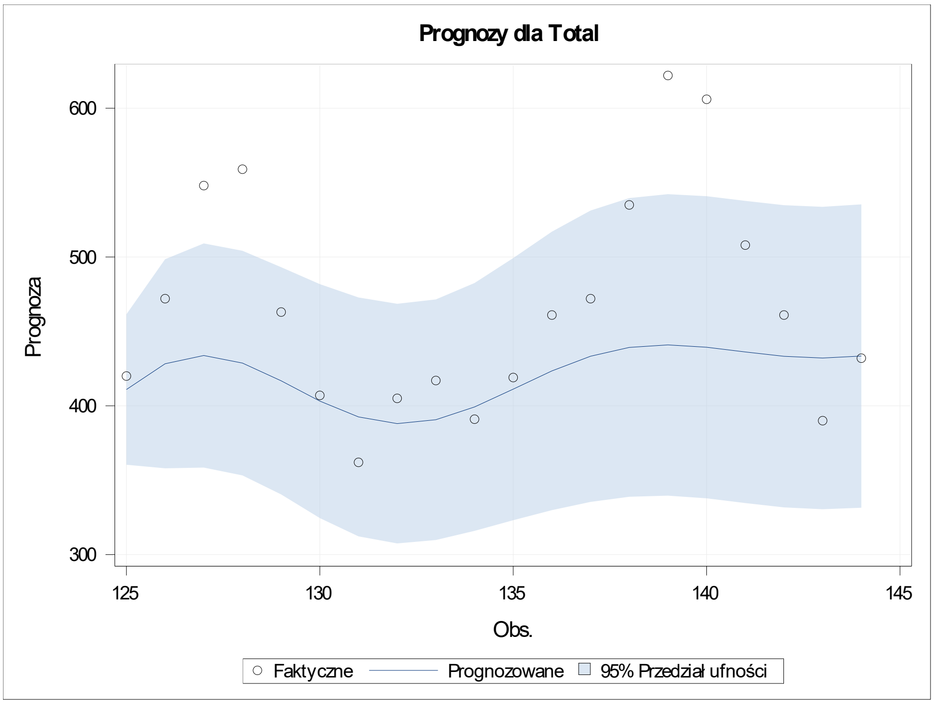
Uwaga. Obliczam prognozy dla ostatnich 20 wartości naszych danych.

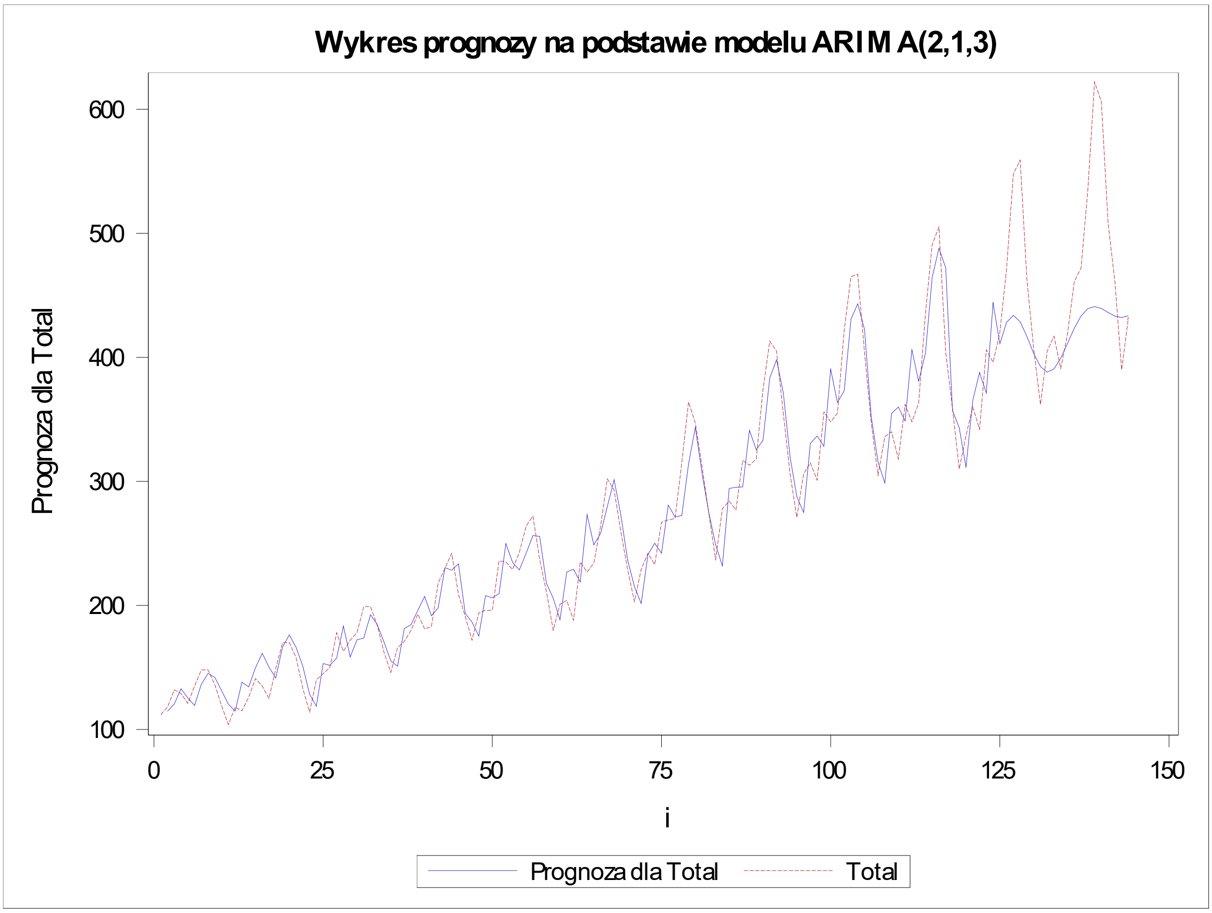
Model ARIMA(p=1,d=0,q=4)





Model ARIMA(p=2,d=1,q=3)

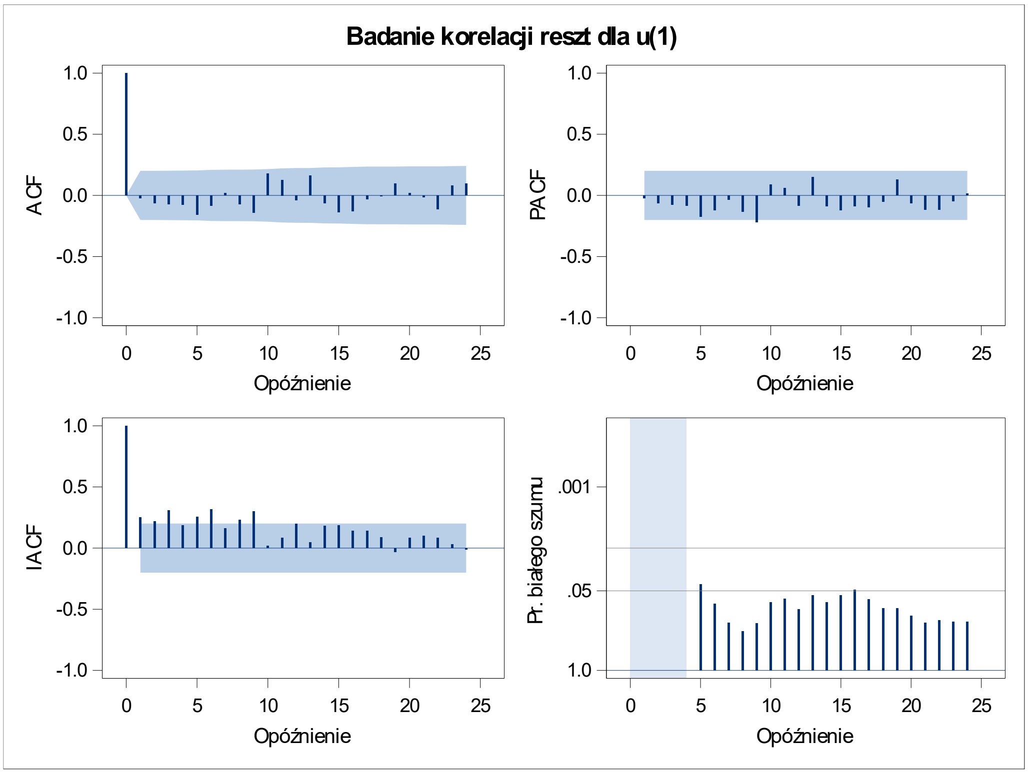


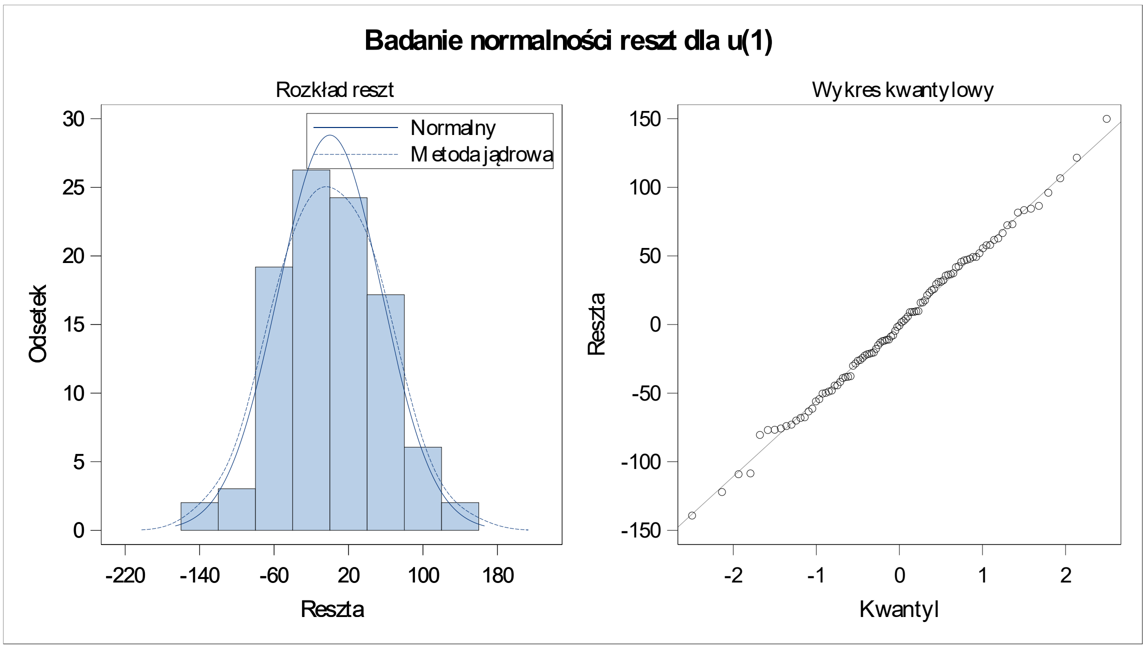


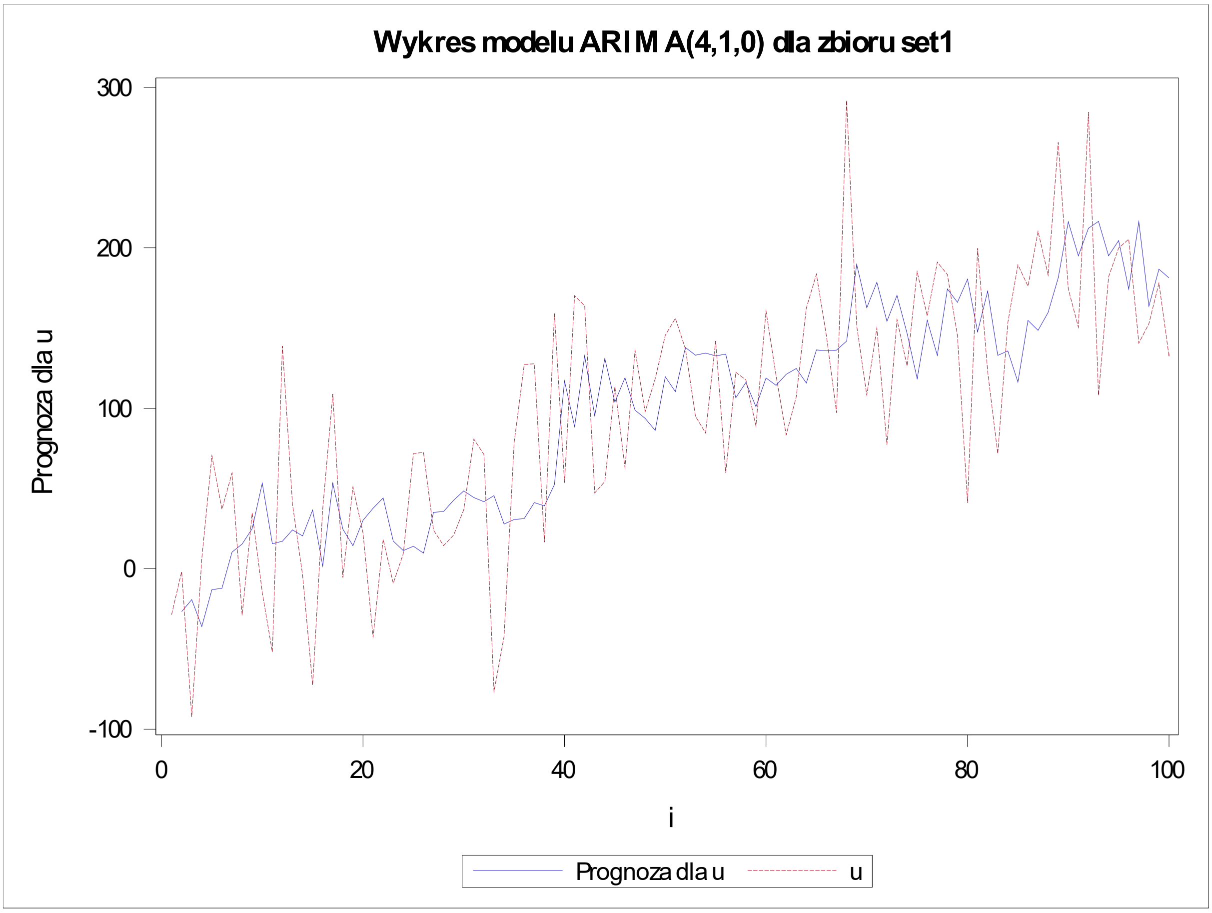
3) <u>Oblicz średni błąd kwadratowy (MSE) dla modelu znalezionych modeli.</u>

set1

Model ARIMA(p=4,d=1,q=0)



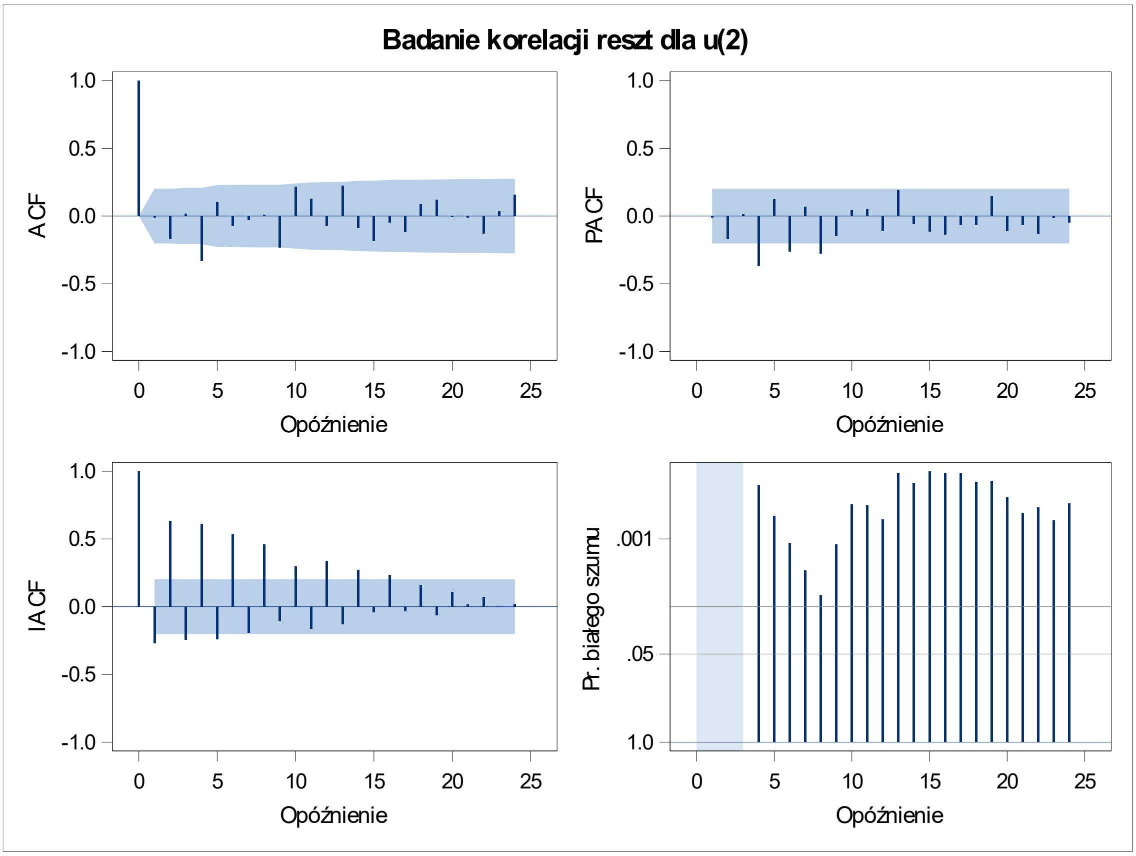


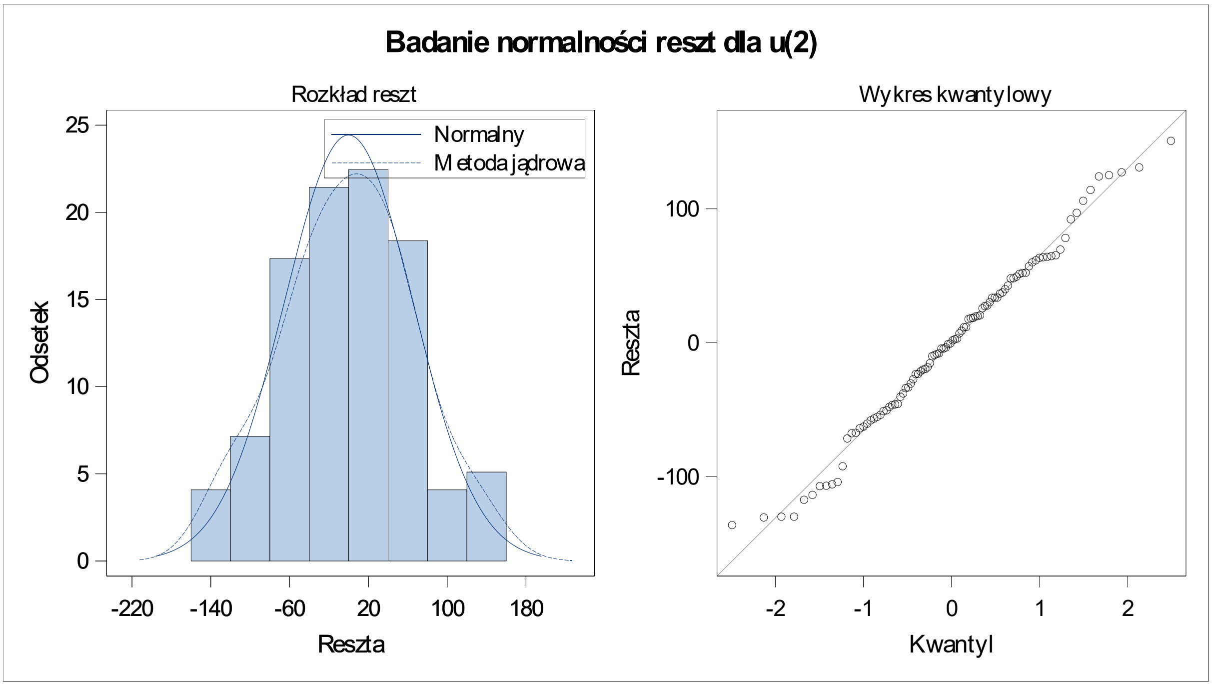


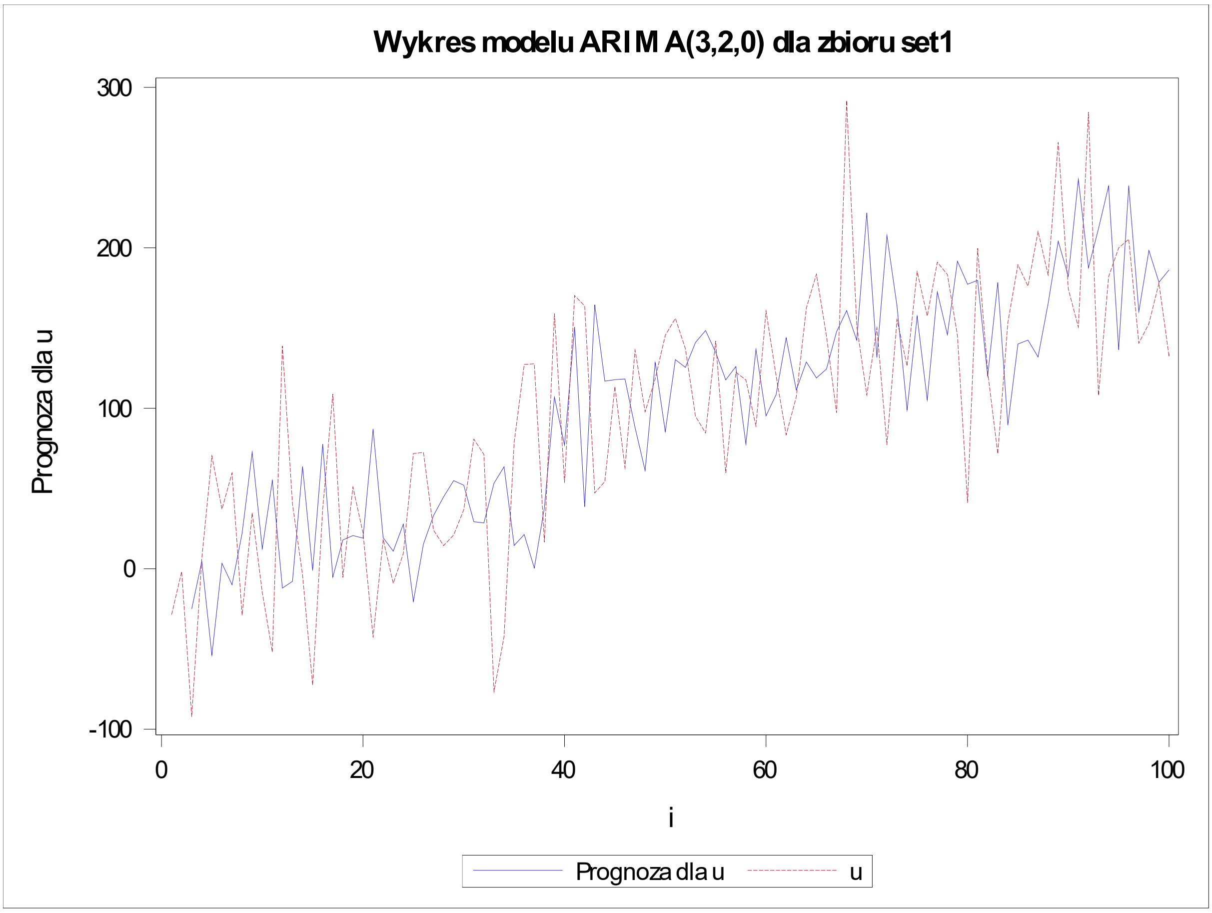
Zmienna analizowana: mean_squared_error				
N	**Średnia**	**Odch. std.**	**Minimum**	**Maksimum**
99	3037.21	4118.73	0.6395354	22474.64

MSE wynosi ok. 3037.21 (kolumna Średnia).

Model ARIMA(p=3, d=2, q=0)



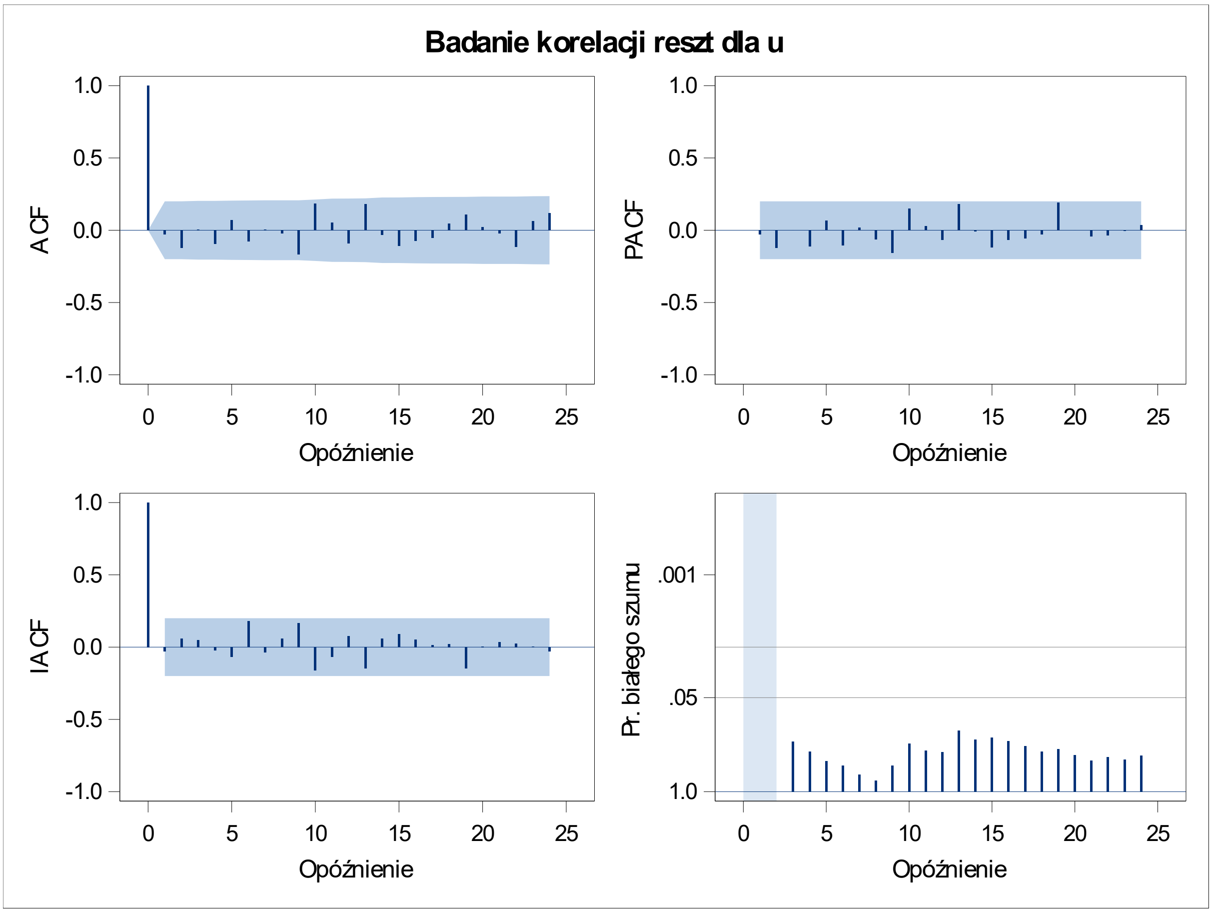


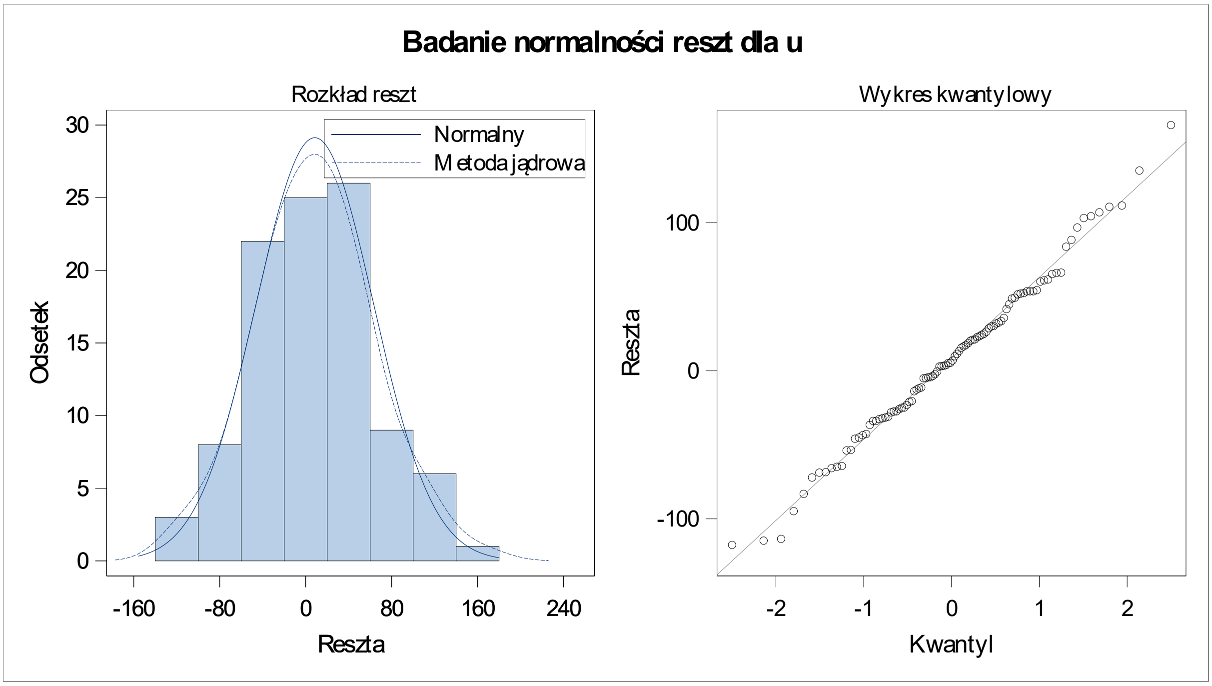


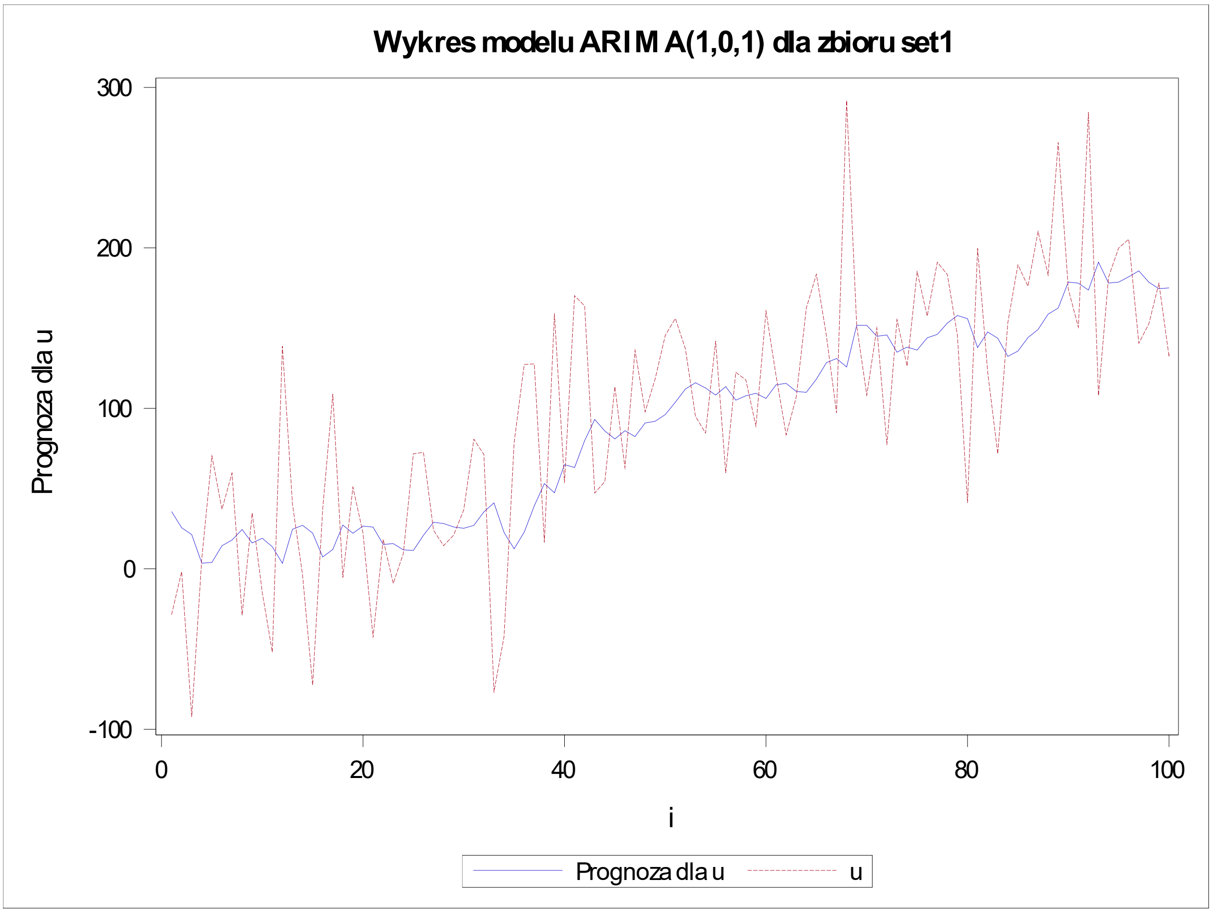
Zmienna analizowana: mean_squared_error				
N	Średnia	Odch. std.	Minimum	Maksimum
98	4221.17	5402.52	0.3268366	22736.51

MSE wynosi ok. 4221.17 (kolumna Średnia).

Model ARIMA(p=1,d=0,q=1)





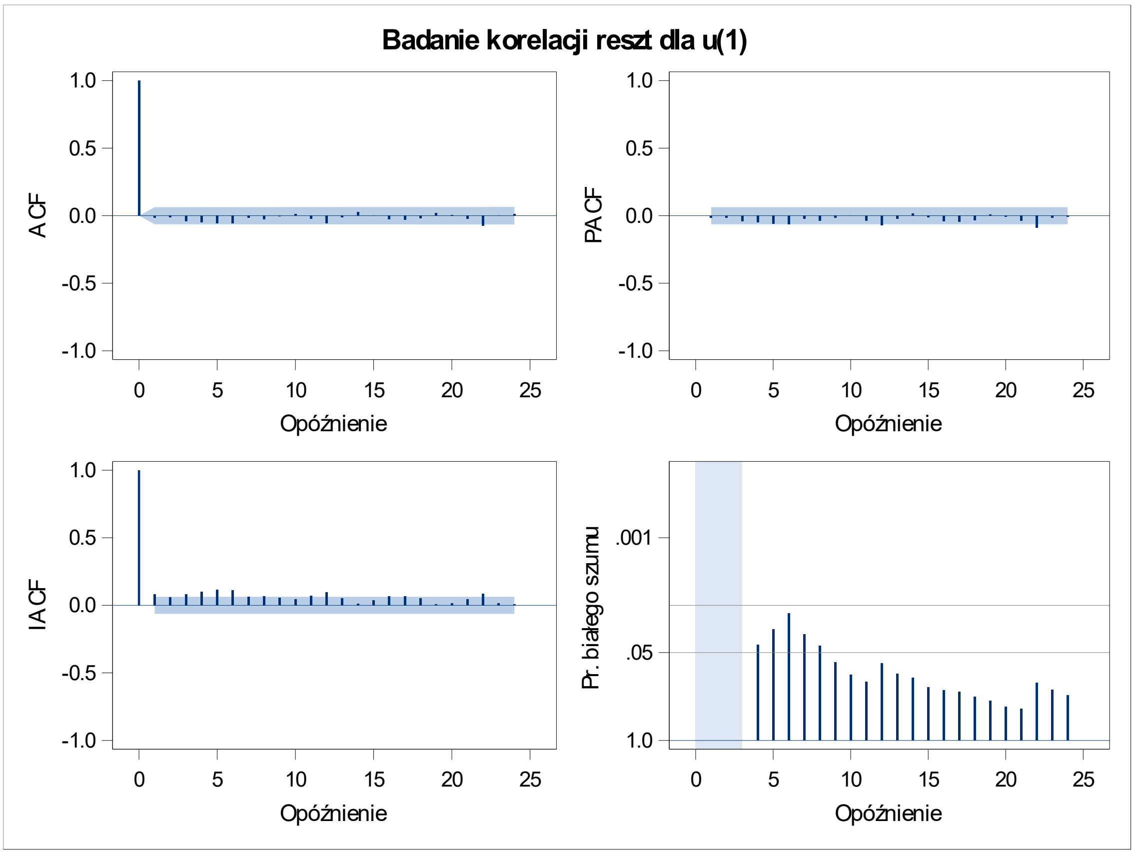


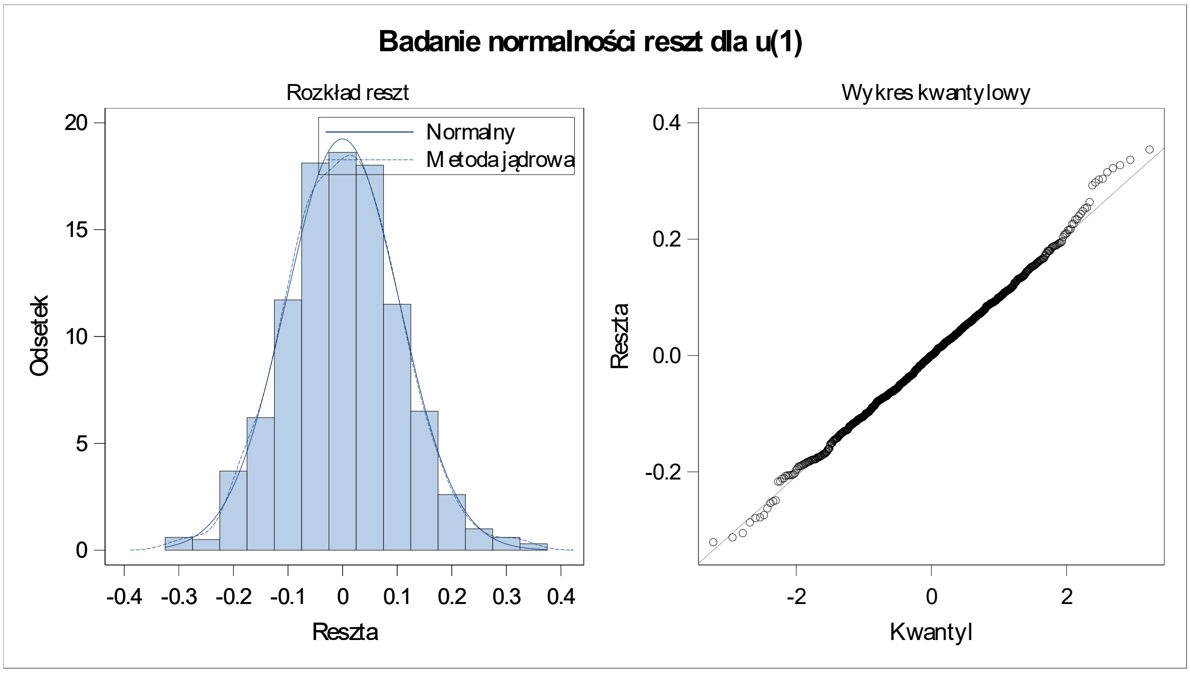
Zmienna analizowana: mean_squared_error				
N	**Średnia**	**Odch. std.**	**Minimum**	**Maksimum**
100	3042.45	4550.39	0.1436704	27568.76

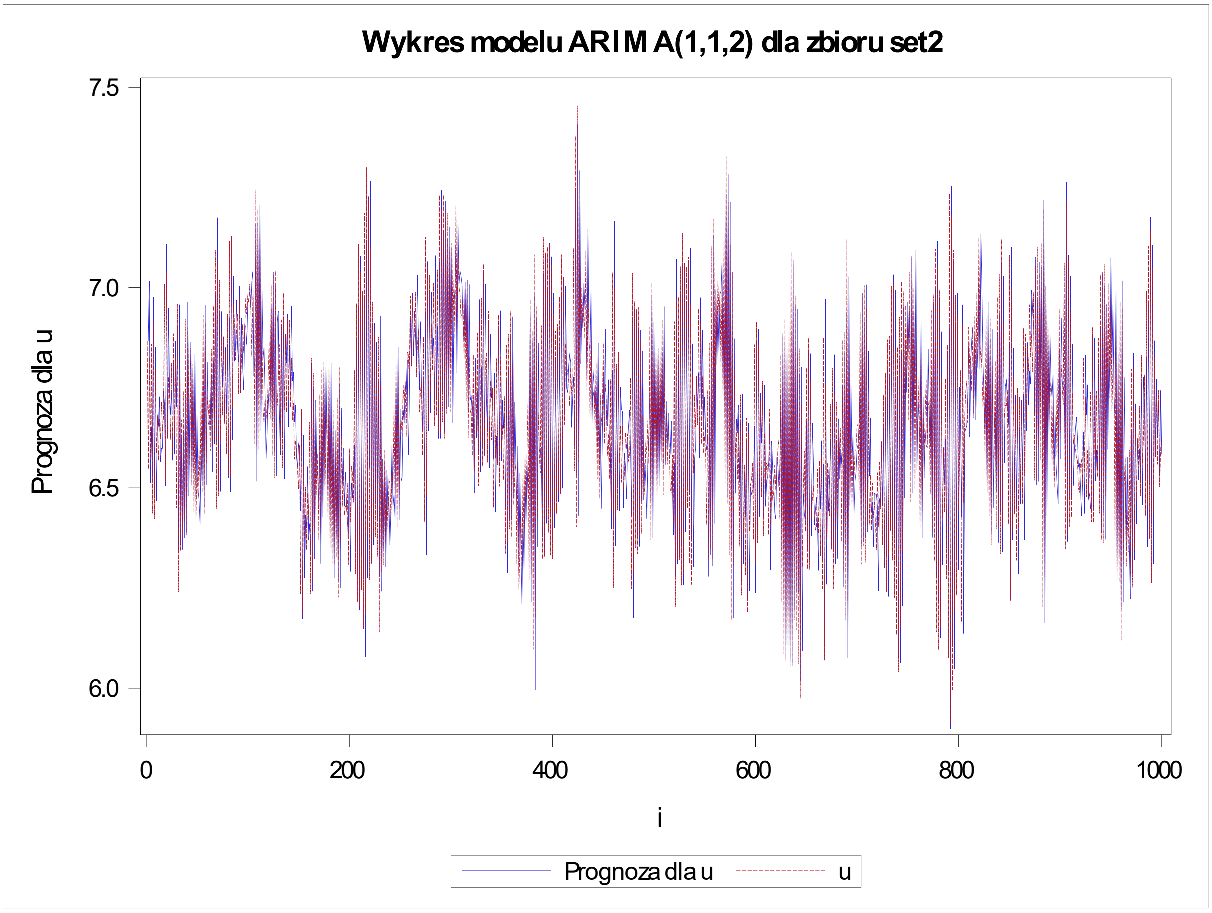
MSE wynosi ok. 3042.45 (kolumna Średnia).

set2

Model ARIMA(p=1,d=1,q=2)



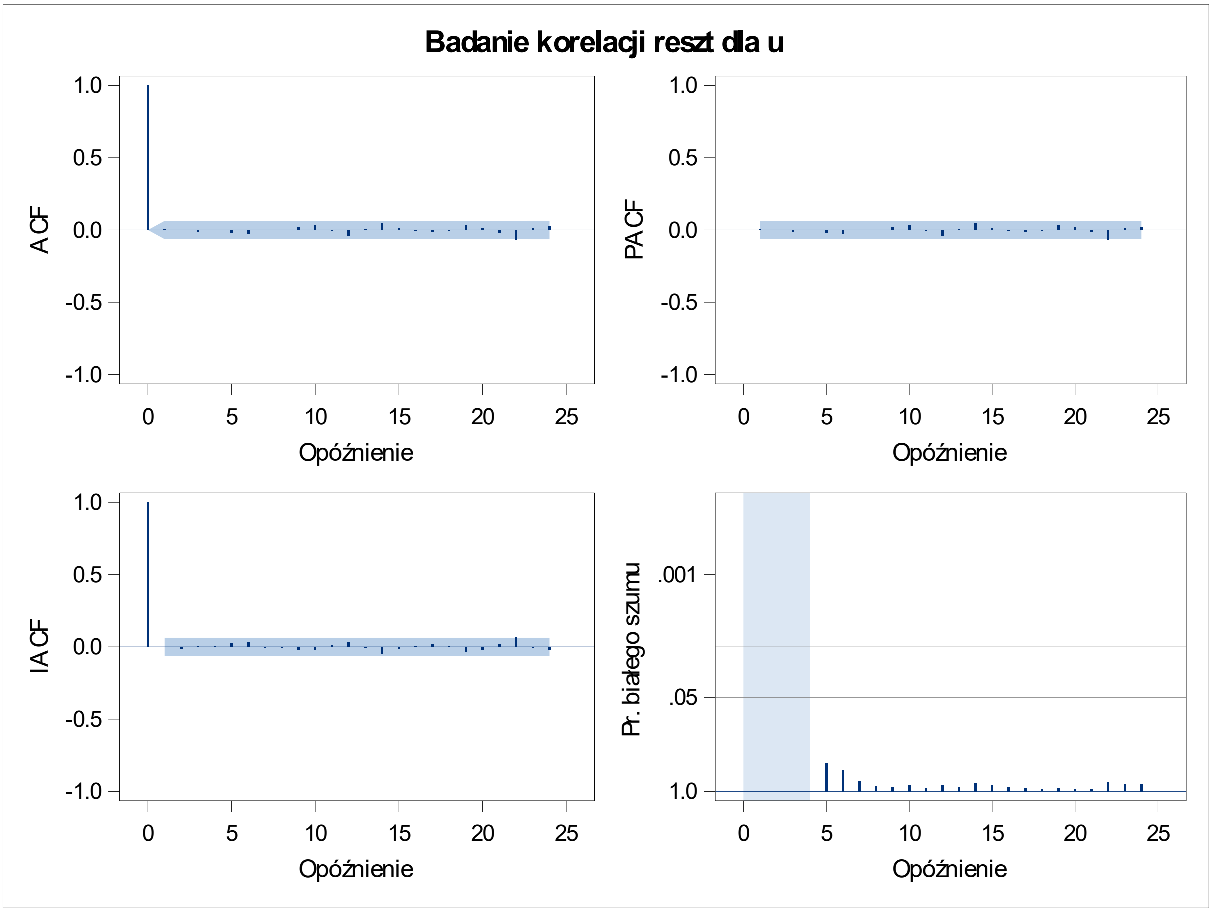


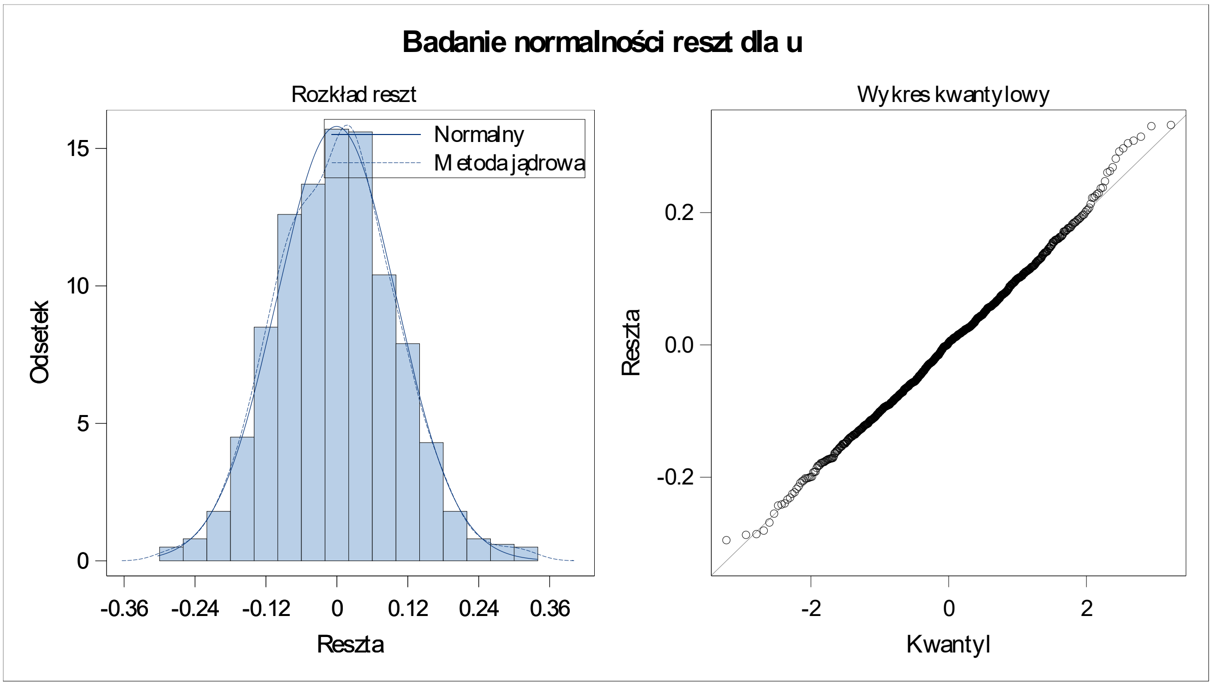


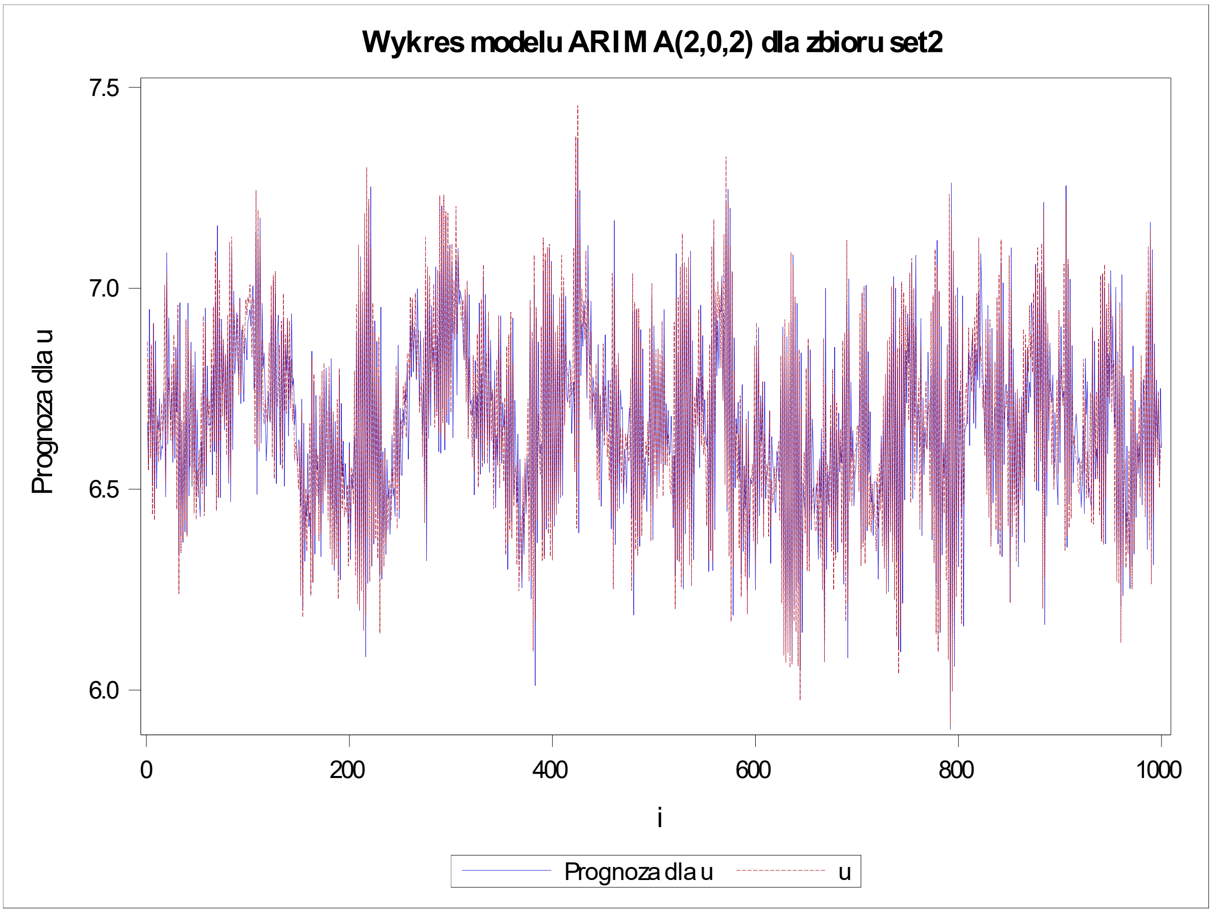
Zmienna analizowana: mean_squared_error				
N	**Średnia**	**Odch. std.**	**Minimum**	**Maksimum**
999	0.0107314	0.0160935	2.8553294E-9	0.1253228

MSE wynosi ok. 0.0107314 (kolumna Średnia).

Model ARIMA(p=2,d=0,q=2)





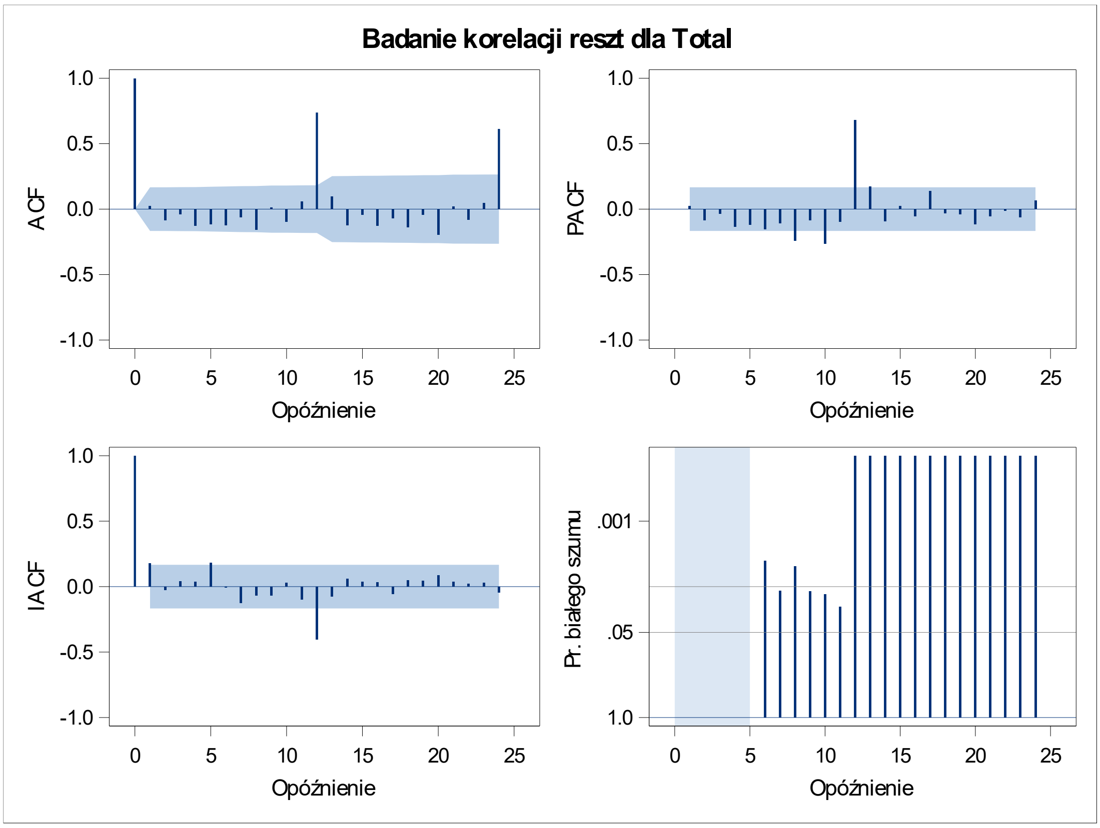


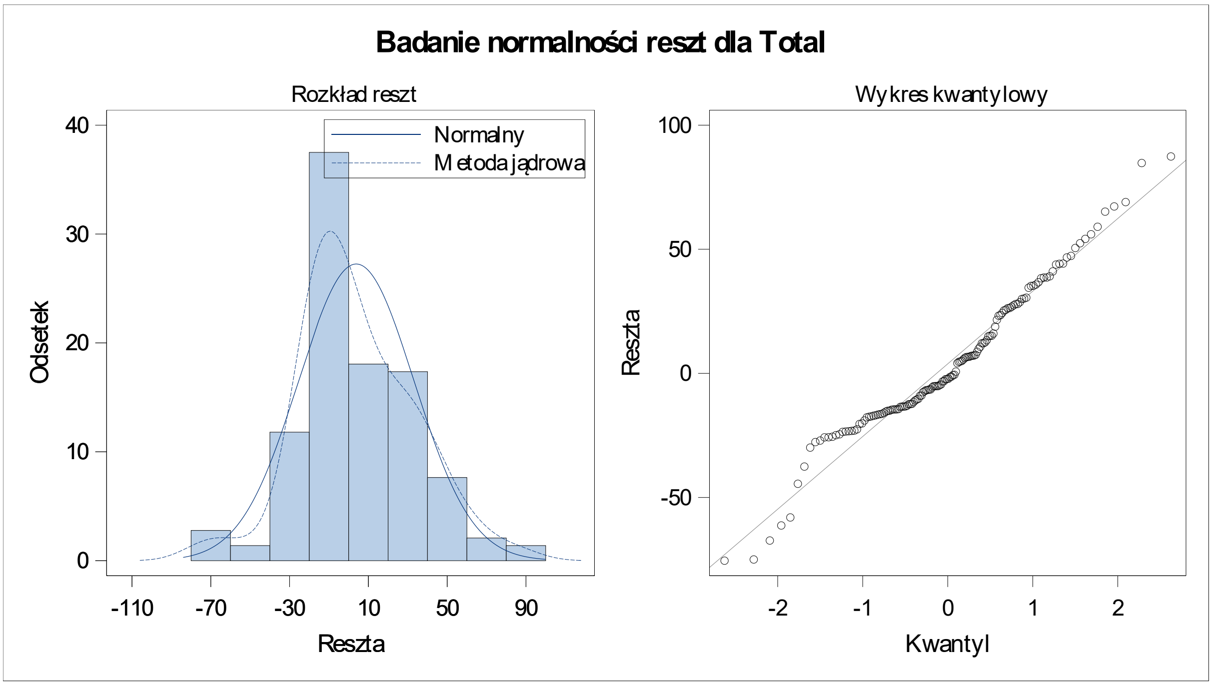
Zmienna analizowana: mean_squared_error				
N	**Średnia**	**Odch. std.**	**Minimum**	**Maksimum**
1000	0.0101902	0.0149297	2.9904771E-9	0.1105449

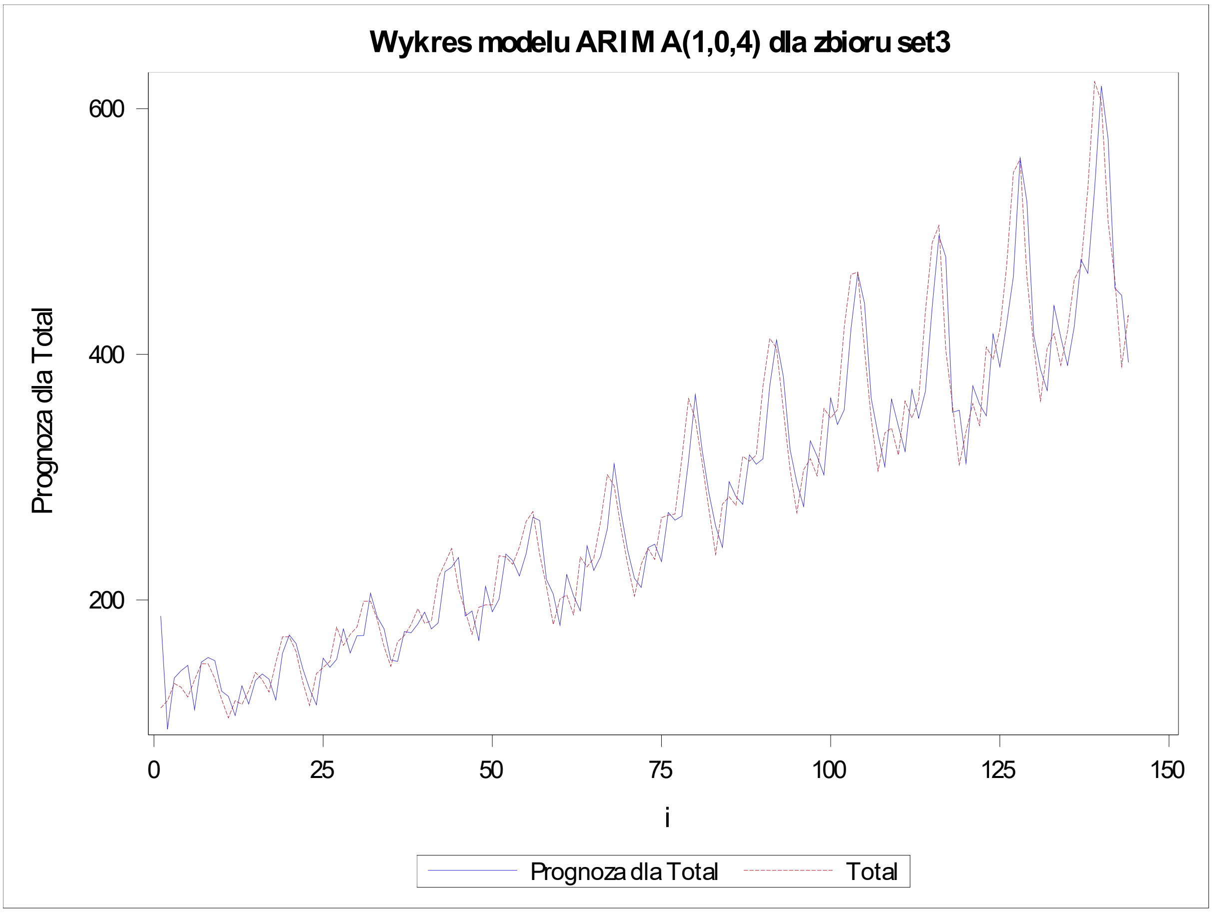
MSE wynosi ok. 0.0101902 (kolumna Średnia).

set3

Model ARIMA(p=1,d=0,q=4)



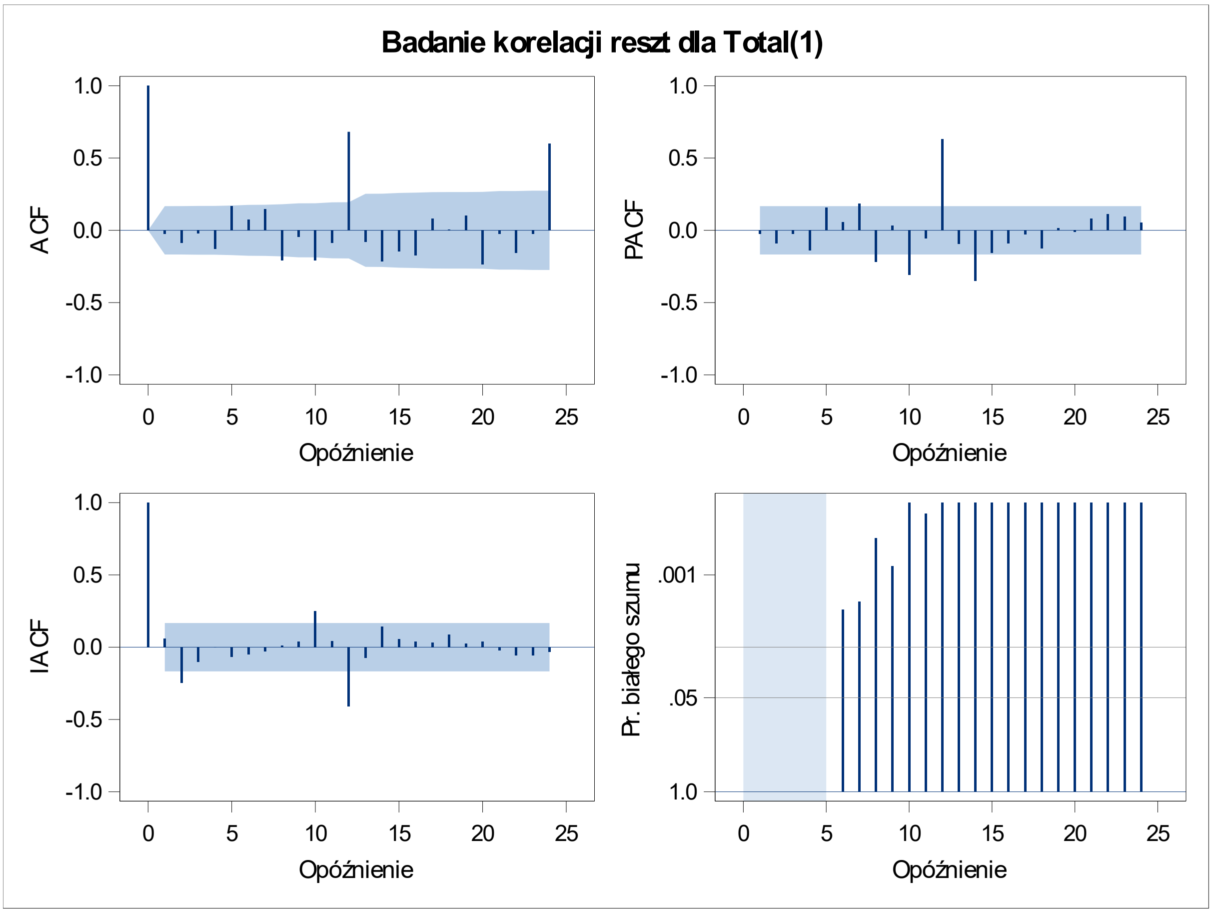


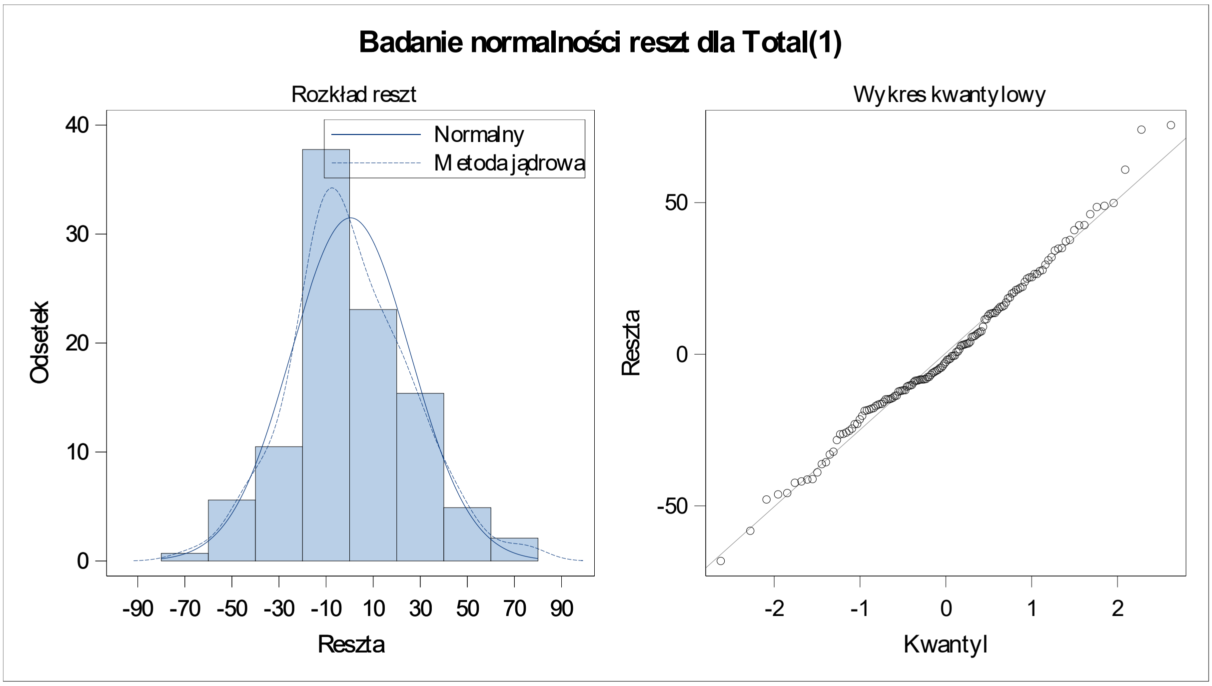


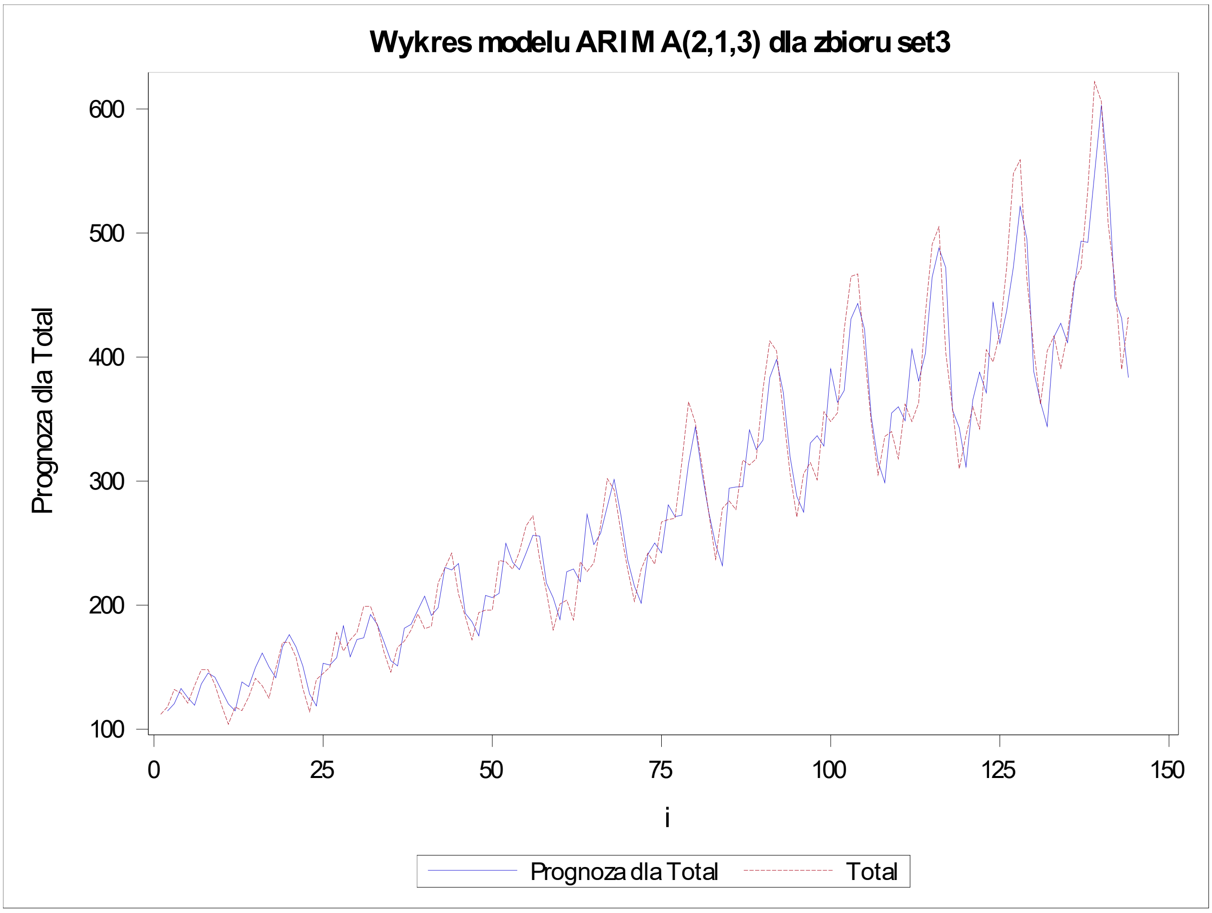
Zmienna analizowana: mean_squared_error				
N	Średnia	Odch. std.	Minimum	Maksimum
144	866.1432509	1402.12	0.3268770	7636.24

MSE wynosi ok. 866.1432509 (kolumna Średnia).

Model ARIMA(p=2,d=1,q=3)







Zmienna analizowana: mean_squared_error				
N	**Średnia**	**Odch. std.**	**Minimum**	**Maksimum**
143	637.5556943	989.0739798	0.1768718	5708.63

MSE wynosi ok. 637.5556943 (kolumna Średnia).

Wnioski z budowy modeli

set1

Model ARIMA(4,1,0) charakteryzuje się rozkładem reszt dość dobrze przypominającym rozkład normalny (wykres normalności reszt). Wykresy korelacji reszt sugerują, że występuje cykliczność błędów (szczęśliwie nie są bardzo duże, choć nie są pomijalne), co oznacza, że model nie oddaje danych najwierniej.

Model ARIMA(3,2,0) charakteryzuje się rozkładem reszt niezbyt dobrze przypominającym rozkład normalny (wykres normalności reszt). Wykresy korelacji reszt sugerują, że występuje cykliczność błędów (są dość spore), co oznacza, że model nie oddaje danych najwierniej.

Model ARIMA(1,0,1) charakteryzuje się rozkładem reszt niezbyt dobrze przypominającym rozkład normalny (wykres normalności reszt). Wykresy korelacji reszt sugerują, że występuje cykliczność błędów (nie są do pominięcia), co oznacza, że model nie oddaje danych najwierniej.

Możemy zauważyć, że najmniejszy błąd średniokwadratowy (wśród analizowanych modeli) jest dla modelu ARIMA(4,1,0) i wynosi ok. 3037.21. Przyjmujemy więc, że ten właśnie model (wśród analizowanych modeli) jest najdokładniejszy.

set2

Model ARIMA(1,1,2) charakteryzuje się rozkładem reszt dobrze przypominającym rozkład normalny (wykres normalności reszt). Wykresy korelacji reszt świadczą o występowaniu niewielkich błędów, co oznacza, że model dobrze oddaje dane.

Model ARIMA(2,0,2) charakteryzuje się rozkładem reszt dobrze przypominającym rozkład normalny (wykres normalności reszt). Wykresy korelacji reszt świadczą o występowaniu niewielkich błędów, co oznacza, że model dobrze oddaje dane.

Możemy zauważyć, że najmniejszy błąd średniokwadratowy (wśród analizowanych modeli) jest dla modelu ARIMA(2,0,2) i wynosi ok. 0.0101902. Przyjmujemy więc, że ten właśnie model (wśród analizowanych modeli) jest najdokładniejszy.

set3

Model ARIMA(1,0,4) charakteryzuje się rozkładem reszt słabo przypominającym rozkład normalny (wykres normalności reszt). Wykresy korelacji reszt sugerują, że występuje cykliczność sporych błędów, co oznacza, że model nienajlepiej oddaje dane.

Model ARIMA(2,1,3) charakteryzuje się rozkładem reszt słabo przypominającym rozkład normalny (wykres normalności reszt). Wykresy korelacji reszt sugerują, że występuje cykliczność niemałych błędów, co oznacza, że model nienajlepiej oddaje dane.

Możemy zauważyć, że najmniejszy błąd średniokwadratowy (wśród analizowanych modeli) jest dla modelu ARIMA(2,1,3) i wynosi ok. 637.5556943. Przyjmujemy więc, że ten właśnie model (wśród analizowanych modeli) jest najdokładniejszy.

Ogólne wnioski z prognoz

Możemy zauważyć, że nasze prognozy zbiegają do pewnej ustalonej wartości
(czasem dość szybko, czasem występują początkowe oscylacje), ewentualnie wykazują określoną tendencję czy też sezonowość (w zależności od parametrów modelu).
Nasze prognozy "przechodzą" przez dane rzeczywiste, niemniej obserwujemy błędy (często dość spore) między wartościami prognozowanymi a rzeczywistymi.