

Badanie czynników wpływających PKB w Polsce

model ekonometryczny

Spis treści

1.Dane: opis i źródła	2
2.Podstawowe charakterystyki zmiennych	2
3.Badanie stacjonarności.....	3
3.1.Zmienna Y.....	3
3.2.Zmienne X.....	5
3.3.Usuwanie niestacjonarności zmiennej Y.....	8
3.4.Usuwanie niestacjonarności zmiennych X.....	9
4.Dopasowanie modelu – wykresy zależności	14
4.1.Zależność zmiennej d_Y i wszystkich zmiennych d_X	14
4.2.Zależności pomiędzy zmiennymi d_X	15
5.Wstępna analiza modelu	18
6.Badanie korelacji między zmiennymi	18
7.Metoda Hellwiga	19
8.Budowanie modelu z uwzględnieniem istotnych zmiennych.....	19
9.Badanie normalności rozkładów reszt	20
10.Testowanie autokorelacji.....	20
10.1. Usuwanie autokorelacji.....	21
11.Badanie heteroskedastyczności.....	22
12.Ponowne badanie normalności rozkładu reszt	24
13.Testowanie współliniowości zmiennych modelu	24
13.1.Próba usunięcia współliniowości.....	25
14.Test Ramsey'a RESET	27
14.1.Doprowadzenie do poprawności funkcyjnej.....	27
15.Badanie efektu katalizy	30
16.Badanie koincydencji	30
17.Ostateczna postać modelu oraz jej interpretacja	30

1. Dane: opis i źródła

Celem tego projektu jest zbadanie wpływu poszczególnych czynników wchodzących w skład ogólnego rachunku narodowego na wartość Produktu Krajowego Brutto w Polsce. Analizowane dane pochodzą ze strony Głównego Urzędu Statystycznego. W projekcie wzięto pod uwagę 9 zmiennych objaśniających pochodzących z 35 kwartałów (2007:1 – 2015:3).

Wszystkie poniższe zmienne wchodzą w skład ogólnego rachunku narodowego a ich wartości wyrażają się w milionach PLN.

Zmienna opisywana:

Y: Produkt krajowy brutto (ceny bieżące)

Zmienne opisujące:

X1: Wartość dodana brutto ogółem (ceny bieżące)

X2: Przemysł

X3: Przetwórstwo przemysłowe

X4: Budownictwo

X5: Transport i gospodarka magazynowa

X6: Zakwaterowanie i gastronomia

X7: Informacja i komunikacja

X8: Działalność usługowa (nie wchodząca w skład powyższych)

X9: Import towarów i usług (ceny bieżące)

Link do pobrania pliku, z którego pochodzą dane:

http://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultstronaopisowa/1772/1/5/kwartalne_ws_kazniki_makroekonomiczne_cz_ii_.xls

2. Podstawowe charakterystyki zmiennych

Statystyki opisowe, dla obserwacji z próby 2007:1 – 2015:3

Zmienna	Średnia	Mediana	Minimalna	Maksymalna
Y	375067,	379557,	272329,	476952,
X1	331263,	337634,	242000,	425175,
X2	83100,2	82527,9	58765,2	113989,
X3	60715,8	58622,5	42292,1	84578,3
X4	26115,1	27158,6	11114,4	37300,3
X5	19344,4	18489,6	11578,7	31121,8
X6	3853,43	3695,70	2187,90	6127,30
X7	12966,2	13121,7	9288,20	16864,6
X8	4583,21	4471,50	2857,90	8077,40
X9	164201,	171196,	117019,	209834,

Zmienna	Odch.stand.	Wsp. zmienności	Skośność	Kurtoza
Y	53569,7	0,142827	-0,0892361	-0,874209
X1	48539,8	0,146529	-0,0856322	-0,881111
X2	15029,4	0,180859	0,453487	-0,615302
X3	10920,9	0,179869	0,509341	-0,521105
X4	7478,47	0,286366	-0,298029	-1,17463
X5	4968,84	0,256862	0,658415	-0,443300
X6	975,317	0,253103	0,389283	-0,266098
X7	1785,24	0,137684	0,116416	-0,544375
X8	1323,34	0,288736	0,873648	0,348515
X9	28830,4	0,175580	-0,0896995	-1,44772

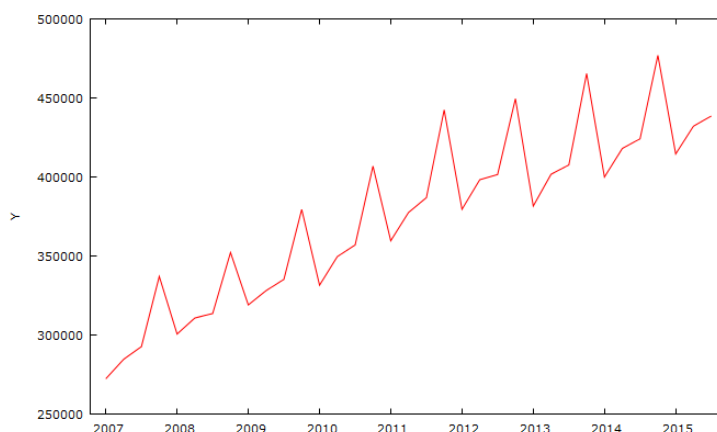
Zmienna	Percentyl 5%	Percentyl 95%	Zakres Q3-Q1	Brakujące obs.
Y	282174,	467735,	83026,1	0
X1	246593,	415273,	76655,8	0
X2	61104,1	112600,	19725,9	0
X3	44448,4	82833,8	16429,2	0
X4	13353,0	36785,4	14374,7	0
X5	12660,9	29196,8	6821,80	0
X6	2214,30	5812,82	1375,10	0
X7	10019,2	16117,1	2936,10	0
X8	2897,66	7721,96	1879,60	0
X9	121187,	208287,	55754,0	0

Podstawowe charakterystyki opisowe wyznaczono powyżej. Na szczególną uwagę zasługuje współczynnik zmienności, który dla wszystkich zmiennych był większy niż 10%. Jest to poziom warunkujący istnienie zmiennej w modelu.

3. Badanie stacjonarności

3.1. Zmienna Y

W celu pozbycia się problemu regresji pozornej badam stacjonarność zmiennej Y.



Z powyższego wykresu szeregu czasowego wynika, że zmienna zmienna Y nie jest stacjonarna. Z charakteru wykresu dopatruje się raczej trendostacjonarności. Aby potwierdzić przypuszczenia wykonuje szereg testów statystycznych:

- **Test ADF – niestacjonarnosci zmiennej Y (testuje istotność)**

Rozszerzony test Dickeya-Fullera dla procesu Y
dla opóźnienia rzędu 7 procesu (1-L)Y
(maksymalne było 8, kryterium AIC)
liczebność próby 28
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,0581616
Statystyka testu: tau_c(1) = -1,84518
asymptotyczna wartość p = 0,3588
Autokorelacja reszt rzędu pierwszego: 0,009
opóźnione różnice: F(7, 19) = 144,365 [0,0000]
```

- **Test ADF – niestacjonarnosci zmiennej Y (nie testuje istotności)**

Rozszerzony test Dickeya-Fullera dla procesu Y
dla opóźnienia rzędu 2 procesu (1-L)Y
liczebność próby 33
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,169252
Statystyka testu: tau_c(1) = -1,52934
asymptotyczna wartość p = 0,5189
Autokorelacja reszt rzędu pierwszego: -0,230
opóźnione różnice: F(2, 29) = 6,325 [0,0053]
```

- **Test ADF-GLS – niestacjonarnosci zmiennej Y (testuje istotność)**

Rozszerzony test Dickeya-Fullera (GLS) dla procesu Y
dla opóźnienia rzędu 4 procesu (1-L)Y
(maksymalne było 5, kryterium zmodyfikowane AIC)
liczebność próby 30
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: 0,00767318
Statystyka testu: tau = 0,333674
asymptotyczna wartość p = 0,7817
Autokorelacja reszt rzędu pierwszego: -0,082
opóźnione różnice: F(4, 25) = 270,196 [0,0000]
```

- **Test KPSS – stacjonarności zmiennej Y**

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej Y
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,971222

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p < .01

3.2. Zmienne X

Wykonuje testy podobne testy dla wszystkich zmiennych X.

- **Test KPSS – stacjonarności**

```
gretl wersja 2016a-git
Bieżąca sesja: 2016-01-31 18:39
? kpss 3 X1
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X1
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,962242
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X2
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X2
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,919977
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X3
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X3
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,945275
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X4
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X4
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,778219
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X5
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X5
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,990735
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X6
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X6
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,982458
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X7
Hipoteza zerowa: proces stacjonarny.
```

```

Test KPSS dla zmiennej X7
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,941098
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X8
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X8
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,902569
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01
? kpss 3 X9
Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej X9
T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,94095
          10%      5%      1%
Krytyczna wart.: 0,353  0,462  0,716
wartość p < .01

```

- **Test ADF – niestacjonarnosci (nie testuje istotności)**

```

gretl wersja 2016a-git
Bieżąca sesja: 2016-01-31 18:51

```

```

? adf 2 X1 --c

```

```

Rozszerzony test Dickeya-Fullera dla procesu X1
dla opóźnienia rzędu 2 procesu (1-L)X1
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

```

```

test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,184777
Statystyka testu: tau_c(1) = -1,63879
asymptotyczna wartość p = 0,4626
Autokorelacja reszt rzędu pierwszego: -0,218
opóźnione różnice: F(2, 28) = 5,955 [0,0070]

```

```

? adf 2 X2 --c

```

```

Rozszerzony test Dickeya-Fullera dla procesu X2
dla opóźnienia rzędu 2 procesu (1-L)X2
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

```

```

test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,36857
Statystyka testu: tau_c(1) = -2,01762
asymptotyczna wartość p = 0,2794
Autokorelacja reszt rzędu pierwszego: -0,407
opóźnione różnice: F(2, 28) = 4,070 [0,0281]

```

```

? adf 2 X3 --c

```

```

Rozszerzony test Dickeya-Fullera dla procesu X3
dla opóźnienia rzędu 2 procesu (1-L)X3
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

```

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,228386
Statystyka testu: tau_c(1) = -1,40563
asymptotyczna wartość p = 0,5812
Autokorelacja reszt rzędu pierwszego: -0,505
opóźnione różnice: F(2, 28) = 9,627 [0,0007]
```

? adf 2 X4 --c

Rozszerzony test Dickeya-Fullera dla procesu X4
dla opóźnienia rzędu 2 procesu (1-L)X4
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -1,47258
Statystyka testu: tau_c(1) = -4,37196
asymptotyczna wartość p = 0,0003275 (test KPSS nie potwierdza stacjonarnosci)
Autokorelacja reszt rzędu pierwszego: 0,060
opóźnione różnice: F(2, 28) = 0,788 [0,4647]
```

? adf 2 X5 --c

Rozszerzony test Dickeya-Fullera dla procesu X5
dla opóźnienia rzędu 2 procesu (1-L)X5
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,0342717
Statystyka testu: tau_c(1) = -0,307916
asymptotyczna wartość p = 0,9215
Autokorelacja reszt rzędu pierwszego: -0,701
opóźnione różnice: F(2, 28) = 35,850 [0,0000]
```

? adf 2 X6 --c

Rozszerzony test Dickeya-Fullera dla procesu X6
dla opóźnienia rzędu 2 procesu (1-L)X6
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,389817
Statystyka testu: tau_c(1) = -1,67288
asymptotyczna wartość p = 0,4452
Autokorelacja reszt rzędu pierwszego: -0,565
opóźnione różnice: F(2, 28) = 14,323 [0,0001]
```

? adf 2 X7 --c

Rozszerzony test Dickeya-Fullera dla procesu X7
dla opóźnienia rzędu 2 procesu (1-L)X7
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,189641
Statystyka testu: tau_c(1) = -1,39632
asymptotyczna wartość p = 0,5858
Autokorelacja reszt rzędu pierwszego: 0,005
opóźnione różnice: F(2, 28) = 32,294 [0,0000]
```

```
? adf 2 X8 --c
```

Rozszerzony test Dickeya-Fullera dla procesu X8

dla opóźnienia rzędu 2 procesu (1-L)X8

liczebność próby 32

Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,368494
Statystyka testu: tau_c(1) = -1,80055
asymptotyczna wartość p = 0,3807
Autokorelacja reszt rzędu pierwszego: 0,019
opóźnione różnice: F(2, 28) = 9,552 [0,0007]
```

```
? adf 2 X9 --c
```

Rozszerzony test Dickeya-Fullera dla procesu X9

dla opóźnienia rzędu 2 procesu (1-L)X9

liczebność próby 32

Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -0,0286817
Statystyka testu: tau_c(1) = -0,57617
asymptotyczna wartość p = 0,8735
Autokorelacja reszt rzędu pierwszego: -0,012
opóźnione różnice: F(2, 28) = 0,327 [0,7239]
```

Wniosek dla zmiennej Y:

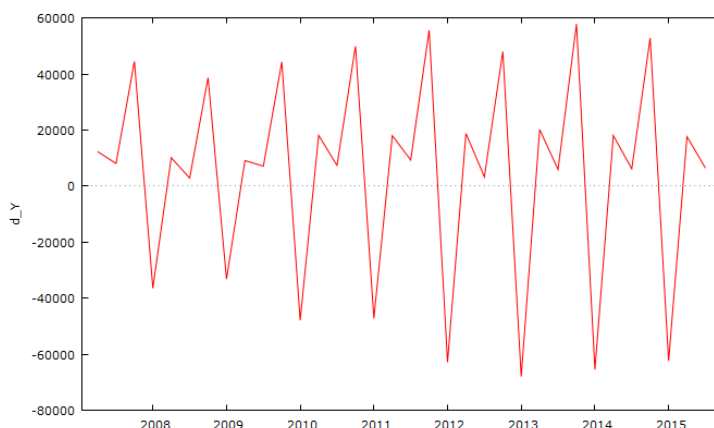
Wszystkie testy ADF dają podobne rezultaty, może zatem powiedzieć, że nie ma podstaw do odrzucenia hipotezy zerowej mówiącej, że szereg jest niestacjonarny. Należy również odrzucić hipotezę zerową testu KPSS mówiącą, że szereg jest stacjonarny. Stwierdzam więc, że mamy do czynienia z szeregiem niestacjonarnym – nie mogę więc wziąć do modelu oryginalnej zmiennej Y.

Wniosek dla zmiennych X:

Wszystkie wykonane testy pokazały podobne rezultaty jak dla zmiennej Y. Należy pozbyć się problemu niestacjonarności przed kontynuowaniem analizy.

3.3. Usuwanie niestacjonarności zmiennej Y

Próbuje pozbyć się stacjonarności poprzez **zróźnicowanie danych**. Poniższy wykres czasowy kwartalnej zmiany wartości zmiennej Y wygląda dużo lepiej i wykazuje cechy zbliżone do szeregu stacjonarnego. Testuje więc stacjonarność nowej zmiennej d_Y .



- **Test ADF – niestacjonarności zmiennej d_Y (nie testuje istotności)**

Rozszerzony test Dickeya-Fullera dla procesu d_Y
 dla opóźnienia rzędu 2 procesu (1-L)d_Y
 liczebność próby 32
 Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)
 test z wyrazem wolnym (const)
 model: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
 estymowana wartość (a-1) wynosi: -3,95829
 Statystyka testu: $\tau_c(1) = -28,6855$
asymptotyczna wartość p = 2,898e-050
 Autokorelacja reszt rzędu pierwszego: 0,693
 opóźnione różnice: $F(2, 28) = 187,374 [0,0000]$

- **Test KPSS – stacjonarności zmiennej DY**

Hipoteza zerowa: proces stacjonarny.
 Test KPSS dla zmiennej d_Y
 T = 35
 Parametr rzędu opóźnienia (lag truncation) = 3
 Statystyka testu = 0,313926

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p > .10

Wniosek:

Wyniki wszystkich testów pozwalają przyjąć założenie, że nowy szereg DY jest stacjonarny. Do modelu biorę zatem d_Y (kwartalna zmiana wartości produktu krajowego brutto) zamiast Y.

3.4. *Usuwanie niestacjonarności zmiennych X*

Aby pozbyć się niestacjonarności postępuje podobnie jak dla zmiennej Y – zamiast zmiennych X będę posługiwał się ich przyrostami. Ponieważ program Gretl nie umożliwia usunięcia poszczególnych wierszy wygenerowanej nowej zmiennej przyrostowej – brakujące dane (dla okresu 1) wprowadzam ręcznie (aby nie doprowadzić do interpolacji w czasie testów). Każdą zmianę testuje.

- **Test KPSS – stacjonarności zmiennych d_X**

gretl wersja 2016a-git
 Bieżąca sesja: 2016-01-31 19:18
 ? kpss 3 d_X1
 Hipoteza zerowa: proces stacjonarny.
 Test KPSS dla zmiennej d_X1
 T = 35
 Parametr rzędu opóźnienia (lag truncation) = 3
 Statystyka testu = 0,313483

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

 wartość p > .10
 ? kpss 3 d_X2
 Hipoteza zerowa: proces stacjonarny.
 Test KPSS dla zmiennej d_X2

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,265684

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p > .10

? kpss 3 d_X3

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej d_X3

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,237177

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p > .10

? kpss 3 d_X4

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej d_X4

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,321024

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p > .10

? kpss 3 d_X5

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej d_X5

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,453989

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

Interpolowana wartość p 0,054

? kpss 3 d_X6

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej d_X6

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,435986

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

Interpolowana wartość p 0,062

? kpss 3 d_X7

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej d_X7

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,243517

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p > .10

? kpss 3 d_X8

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej d_X8

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,25198

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p > .10

? kpss 3 d_X9

Hipoteza zerowa: proces stacjonarny.
Test KPSS dla zmiennej d_X9

T = 35
Parametr rzędu opóźnienia (lag truncation) = 3
Statystyka testu = 0,0655011

	10%	5%	1%
Krytyczna wart.:	0,353	0,462	0,716

wartość p > .10

- **Test ADF – niestacjonarnosci zmiennych d_X (nie testuje istotności)**

gretl wersja 2016a-git

Bieżąca sesja: 2016-01-31 19:22

? adf 2 d_X1 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X1
dla opóźnienia rzędu 2 procesu (1-L)d_X1
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

test z wyrazem wolnym (const)
model: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
estymowana wartość (a-1) wynosi: -3,99771
Statystyka testu: $\tau_c(1) = -27,5564$
asymptotyczna wartość p = 2,243e-051
Autokorelacja reszt rzędu pierwszego: 0,500
opóźnione różnice: $F(2, 28) = 172,647 [0,0000]$

? adf 2 d_X2 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X2
dla opóźnienia rzędu 2 procesu (1-L)d_X2
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy a = 1; proces I(1)

test z wyrazem wolnym (const)
model: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
estymowana wartość (a-1) wynosi: -3,87536
Statystyka testu: $\tau_c(1) = -24,0164$
asymptotyczna wartość p = 5,667e-052
Autokorelacja reszt rzędu pierwszego: 0,220
opóźnione różnice: $F(2, 28) = 159,992 [0,0000]$

? adf 2 d_X3 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X3
dla opóźnienia rzędu 2 procesu (1-L)d_X3
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -3,65636
Statystyka testu: tau_c(1) = -16,4273
asymptotyczna wartość p = 3,089e-039
Autokorelacja reszt rzędu pierwszego: 0,111
opóźnione różnice: F(2, 28) = 93,485 [0,0000]
```

? adf 2 d_X4 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X4
dla opóźnienia rzędu 2 procesu (1-L)d_X4
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -3,95307
Statystyka testu: tau_c(1) = -32,9628
asymptotyczna wartość p = 2,109e-042
Autokorelacja reszt rzędu pierwszego: 0,565
opóźnione różnice: F(2, 28) = 251,549 [0,0000]
```

? adf 2 d_X5 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X5
dla opóźnienia rzędu 2 procesu (1-L)d_X5
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -3,80783
Statystyka testu: tau_c(1) = -17,4552
asymptotyczna wartość p = 6,497e-042
Autokorelacja reszt rzędu pierwszego: 0,334
opóźnione różnice: F(2, 28) = 195,123 [0,0000]
```

? adf 2 d_X6 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X6
dla opóźnienia rzędu 2 procesu (1-L)d_X6
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -3,94341
Statystyka testu: tau_c(1) = -28,7608
asymptotyczna wartość p = 3,56e-050
Autokorelacja reszt rzędu pierwszego: 0,403
opóźnione różnice: F(2, 28) = 344,431 [0,0000]
```

? adf 2 d_X7 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X7
dla opóźnienia rzędu 2 procesu (1-L)d_X7
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces I(1)

```
test z wyrazem wolnym (const)
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
estymowana wartość (a-1) wynosi: -3,45459
```

Statystyka testu: $\tau_c(1) = -12,0973$
asymptotyczna wartość $p = 5,029e-026$
Autokorelacja reszt rzędu pierwszego: 0,176
opóźnione różnice: $F(2, 28) = 34,736 [0,0000]$

? adf 2 d_X8 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X8
dla opóźnienia rzędu 2 procesu $(1-L)d_X8$
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces $I(1)$

test z wyrazem wolnym (const)
model: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
estymowana wartość $(a-1)$ wynosi: -3,90195
Statystyka testu: $\tau_c(1) = -18,3239$
asymptotyczna wartość $p = 5,575e-044$
Autokorelacja reszt rzędu pierwszego: 0,364
opóźnione różnice: $F(2, 28) = 75,732 [0,0000]$

? adf 2 d_X9 --c

Rozszerzony test Dickeya-Fullera dla procesu d_X9
dla opóźnienia rzędu 2 procesu $(1-L)d_X9$
liczebność próby 32
Hipoteza zerowa: występuje pierwiastek jednostkowy $a = 1$; proces $I(1)$

test z wyrazem wolnym (const)
model: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
estymowana wartość $(a-1)$ wynosi: -1,50875
Statystyka testu: $\tau_c(1) = -4,18179$
asymptotyczna wartość $p = 0,0007036$
Autokorelacja reszt rzędu pierwszego: 0,087
opóźnione różnice: $F(2, 28) = 0,887 [0,4233]$

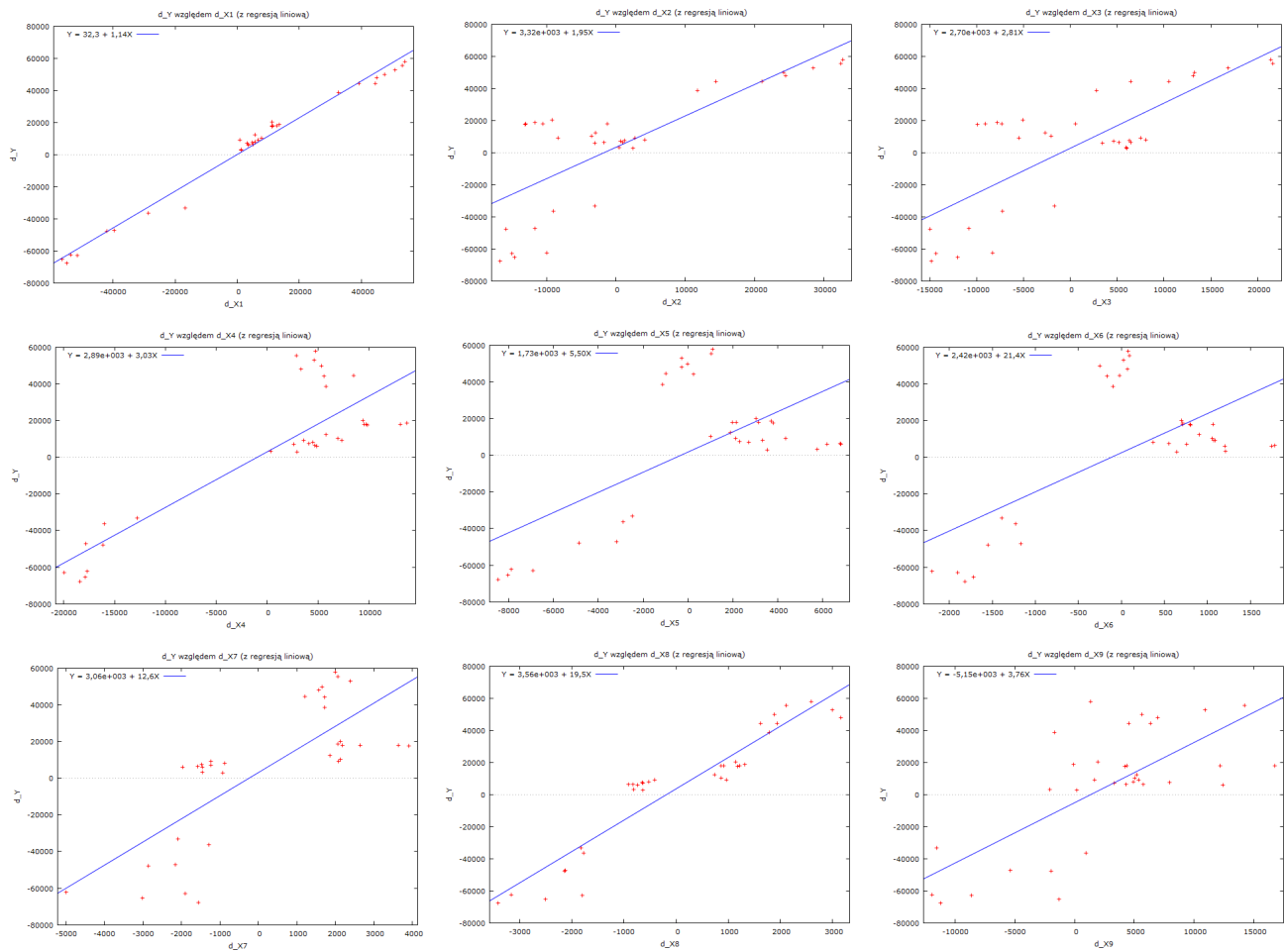
Wnioski:

Z powodów analogicznych jak dla zmiennej Y do dalszej analizy jako zmienne opisujące będziemy brać ich zmiany.

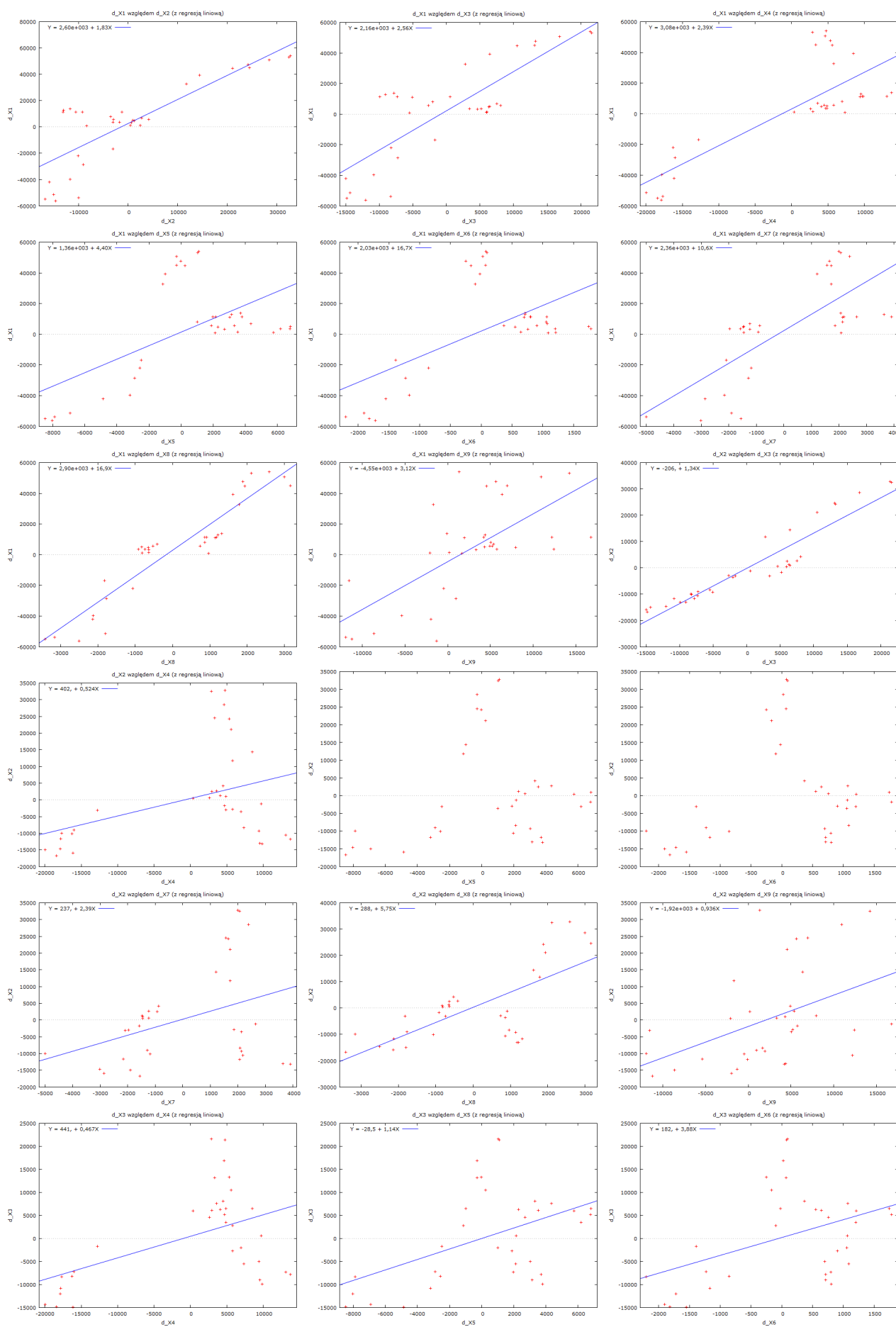


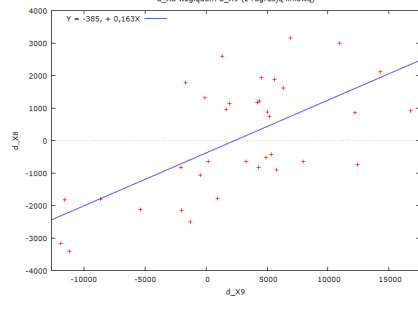
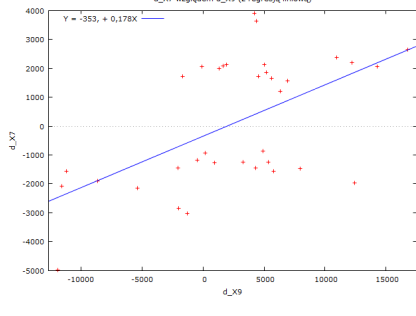
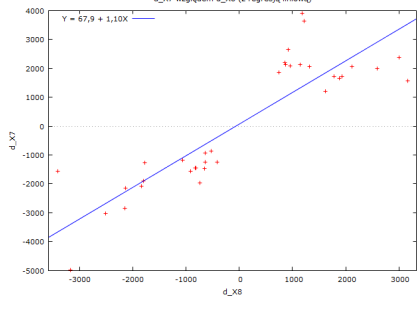
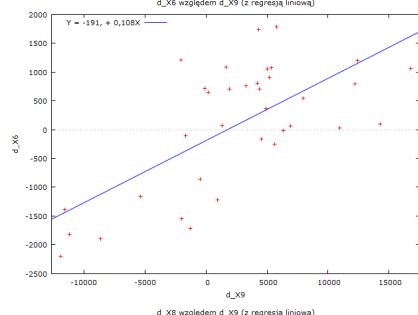
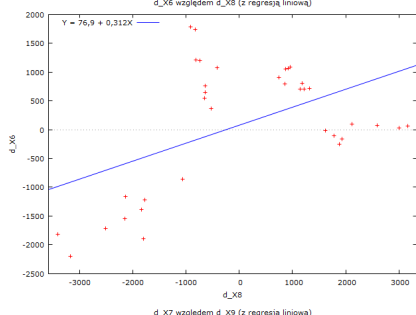
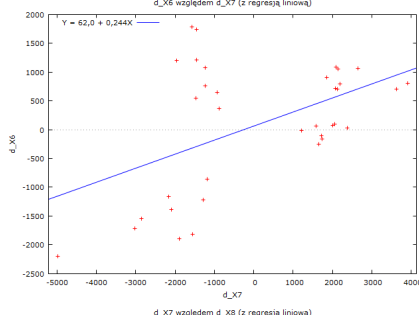
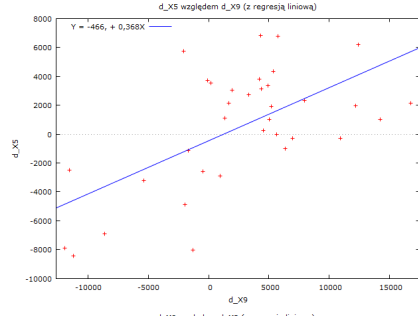
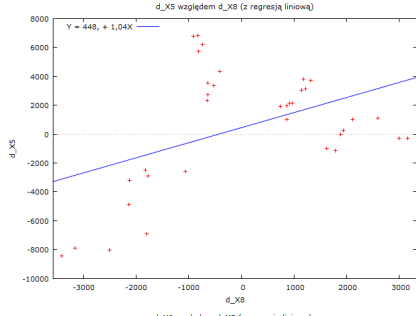
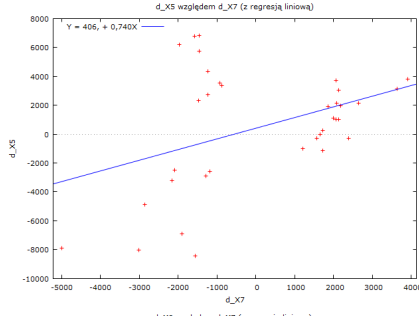
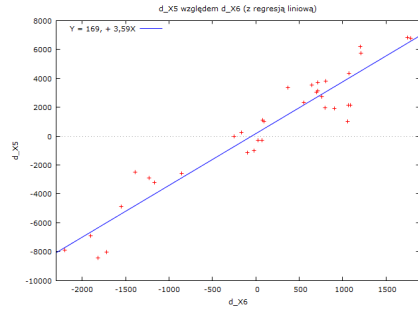
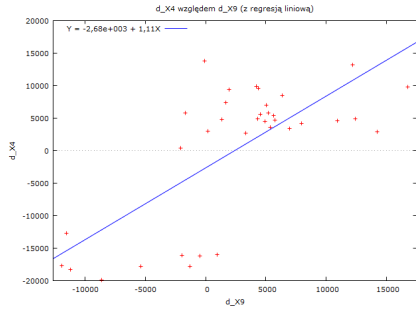
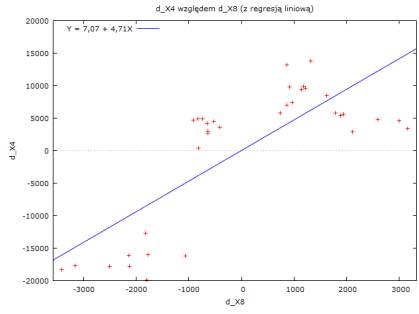
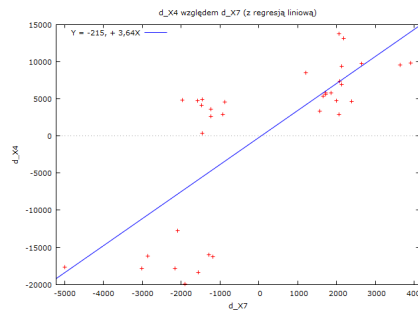
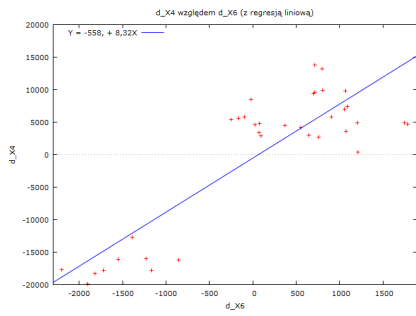
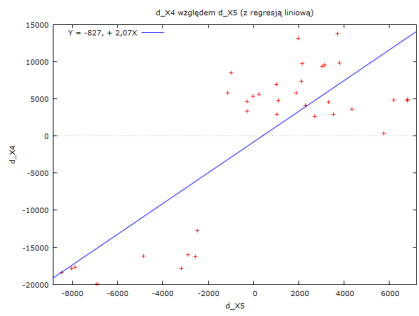
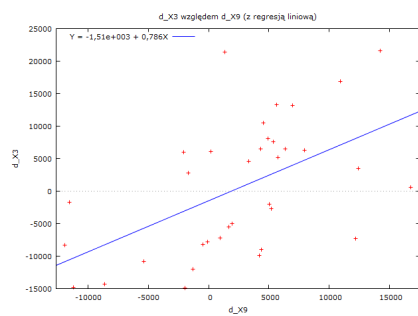
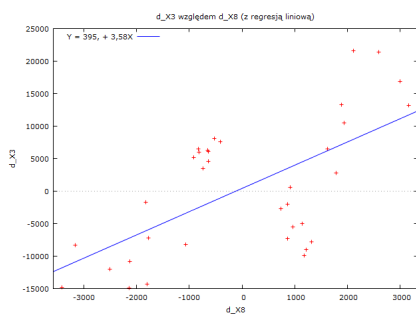
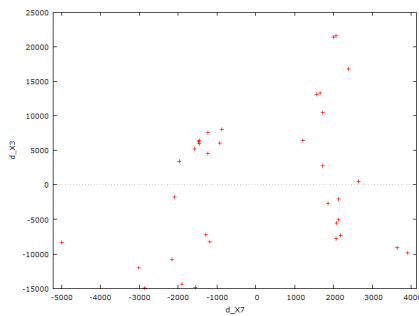
4. Dopasowanie modelu – wykresy zależności

4.1. Zależność zmiennej d_Y i wszystkich zmiennych d_X



4.2. Zależności pomiędzy zmiennymi d_X





Jak można obserwować zmienna Y wykazuje zależności liniowe z wszystkimi zmiennymi X. Jest to na chwilę obecną bardzo dobry prognostyk. Niestety zależności pomiędzy zmiennymi X nawzajem nie wyglądają już tak dobrze - możemy dopatrzeć się wielu zależności liniowych, które nie są dobre dla modelu. Szczęśliwie udaje się również zauważyć zmiennie niezależne liniowo. Siłę poszczególnych zależności przebadam w kolejnych podpunktach. Na chwilę obecną poniżej załączam krótkie podsumowanie.

- pary wykazujące liniową zależność:

d_Y: d_X1,d_X2,d_X3,d_X4,d_X5,d_X6,d_X7,d_X8,d_X9

d_X1: d_X2,d_X3,d_X4,d_X5,d_X6,d_X7,d_X8,d_X9

d_X2: d_X3,d_X4,d_X7,d_X8,d_X9

d_X3: d_X4,d_X5,d_X6,d_X8,d_X9

d_X4: d_X5,d_X6,d_X7,d_X8,d_X9

d_X5: d_X6,d_X7,d_X8,d_X9

d_X6: d_X7,d_X8,d_X9

d_X7: d_X8,d_X9

d_X8: d_X9

- pary nie wykazujące liniowej zależności:

d_X2: d_X5,d_X6

d_X3: d_X7

5. Wstępna analiza modelu

Wykonano estymację modelu klasyczną metodą najmniejszych kwadratów:

Model 8: Estymacja KMNK, wykorzystane obserwacje 2007:1-2015:3 (N = 35)
Zmienna zależna (Y): d_Y

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	1736,05	717,969	2,4180	0,0229	**
d_X2	1,94172	0,421833	4,6031	<0,0001	***
d_X3	-1,35247	0,559751	-2,4162	0,0230	**
d_X4	1,36899	0,196253	6,9756	<0,0001	***
d_X5	2,80229	0,617801	4,5359	0,0001	***
d_X6	-4,16053	2,45975	-1,6914	0,1027	
d_X7	0,596066	0,868997	0,6859	0,4988	
d_X8	4,20717	1,76139	2,3886	0,0245	**
d_X9	0,163941	0,142893	1,1473	0,2617	

Średn.aryt.zm.zależnej	3858,009	Odch.stand.zm.zależnej	36920,29
Suma kwadratów reszt	3,35e+08	Błąd standardowy reszt	3589,937
Wsp. determ. R-kwadrat	0,992770	Skorygowany R-kwadrat	0,990545
F(8, 26)	446,2664	Wartość p dla testu F	8,09e-26
Logarytm wiarygodności	-330,9671	Kryt. inform. Akaike'a	679,9342
Kryt. bayes. Schwarza	693,9323	Kryt. Hannana-Quinna	684,7663
Autokorel.reszt - rho1	-0,232680	Stat. Durbina-Watsona	2,327687

Największe wartość p jest dla zmiennych d_X6, d_X7 i d_X9 co sugeruje, że powinny być one wykluczone z modelu. Wartość p dla testu F (wszystkie współczynniki równe 0) jest mniejsza od 0,05, odrzucono więc hipotezę o zerowości wszystkich współczynników. Współczynnik determinacji R-kwadrat wynosi 99,2%, co oznacza, że model wyjaśnia 99,2% zmienności badanego zjawiska.

6. Badanie korelacji między zmiennymi

Badam korelacje pomiędzy zmiennymi:

Współczynniki korelacji liniowej dla obserwacji z próby 2007:1-2015:3
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,3338 dla n = 35

d_Y	d_X1	d_X2	d_X3	d_X4	
1,0000	0,9921	0,7753	0,7812	0,8571	d_Y
	1,0000	0,8298	0,8177	0,7988	d_X1
		1,0000	0,9396	0,3859	d_X2
			1,0000	0,4902	d_X3
				1,0000	d_X4
d_X5	d_X6	d_X7	d_X8	d_X9	
0,6279	0,6472	0,7631	0,9384	0,7085	d_Y
0,5717	0,5767	0,7363	0,9331	0,6723	d_X1
0,2171	0,1964	0,3642	0,6992	0,4443	d_X2
0,4630	0,4197	0,2782	0,6197	0,5321	d_X3
0,8025	0,8592	0,7528	0,7781	0,7129	d_X4
1,0000	0,9543	0,3939	0,4439	0,6117	d_X5
	1,0000	0,4886	0,4998	0,6755	d_X6
		1,0000	0,8757	0,5560	d_X7

1,0000	0,6354	d_X8
	1,0000	d_X9

Macierz korelacji potwierdziła, że zmienna d_Y jest nieco słabiej skorelowana ze zmiennymi d_X6, d_X7 i d_X9 - podtrzymuje zatem opinie o ich wykluczeniu z modelu. Zmienne d_X2 i d_X3 są mocno ze sobą skorelowane oraz prawie identycznie opisują zmienną opisywaną, dlatego warto byłoby pozbyć się jednej z nich. Zmienna d_X2 jest mniej skorelowana z ważną zmienną d_X4 oraz jest nieznacznie bardziej (w stosunku do d_X3) skorelowana ze zmienną d_X8. Zmienna d_X2 będzie chyba nieco bardziej wartościowa w modelu niż d_X3.

7. Metoda Hellwiga

```
Liczba iteracji: 255
? H_max
0,98319183
? najlepsza_lista
d_X2 d_X4 d_X8
```

Metoda Hellwiga wskazała zmienne d_X2, d_X4, d_X8 jako najbardziej istotne. Pojemność integralna przy tym zestawie wynosi 0,98319183. Metoda ta potwierdziła, iż zmienne wyselekcjonowane w poprzednich punktach są bardzo wartościowe dla modelu. Mamy jeszcze zmienną d_X5 którą decyduje włączyć do ostatecznego modelu, ponieważ jest ona słabiej skorelowana z pozostałymi zmiennymi.

8. Budowanie modelu z uwzględnieniem istotnych zmiennych

Wyestymowano nowy model z uwzględnieniem czterech zmiennych d_X2, d_X4, d_X5 i d_X8:

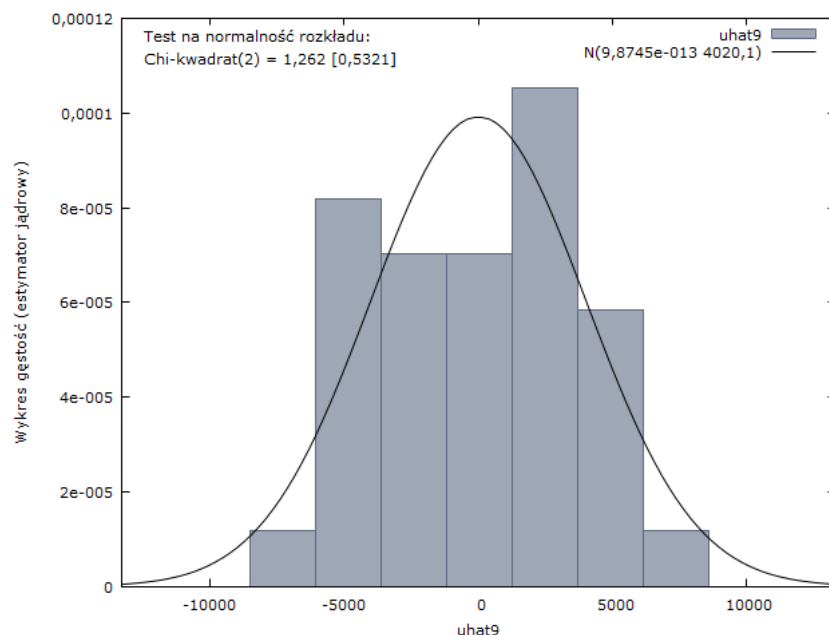
Model 9: Estymacja KMNK, wykorzystane obserwacje 2007:1-2015:3 (N = 35)
Zmienna zależna (Y): d_Y

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	2415,13	694,436	3,4778	0,0016	***
d_X2	0,907826	0,0736152	12,3320	<0,0001	***
d_X4	1,13646	0,192915	5,8910	<0,0001	***
d_X5	1,05028	0,32954	3,1871	0,0033	***
d_X8	7,99854	0,999885	7,9995	<0,0001	***

Średn. arytm. zm. zależnej	3858,009	Odch. stand. zm. zależnej	36920,29
Suma kwadratów reszt	4,85e+08	Błąd standardowy reszt	4020,065
Wsp. determ. R-kwadrat	0,989539	Skorygowany R-kwadrat	0,988144
F(4, 30)	709,4403	Wartość p dla testu F	3,12e-29
Logarytm wiarygodności	-337,4321	Kryt. inform. Akaike'a	684,8642
Kryt. bayes. Schwarza	692,6409	Kryt. Hannana-Quinna	687,5487
Autokorel. reszt - rho1	-0,094400	Stat. Durbina-Watsona	2,131014

Współczynnik determinacji R-kwadrat praktycznie nie uległ zmianie (zmałał o 0,3%) i wynosi 0,9895, co oznacza, że zmienna Y jest objaśniana prawie w 99% przez zmienne objaśniające. Natomiast różnica pomiędzy R-kwadrat a skorygowanym R-kwadrat wynosi 0,001395, model więc nie jest przeparametryzowany. Zmniejszył się nieznacznie logarytm wiarygodności i zwiększyły kryteria informacyjne.


9. Badanie normalności rozkładów reszt



Rozkład częstości dla uhat9, obserwacje 1-35
 liczba przedziałów = 7, średnia = 9,87451e-013, odch.std. = 4020,07

Przedziały	średnia	liczba	częstość	skumulowana
< -6069,2	-7289,8	1	2,86%	2,86% *
-6069,2 - -3628,0	-4848,6	7	20,00%	22,86% *****
-3628,0 - -1186,9	-2407,4	6	17,14%	40,00% *****
-1186,9 - 1254,3	33,710	6	17,14%	57,14% *****
1254,3 - 3695,4	2474,9	9	25,71%	82,86% *****
3695,4 - 6136,6	4916,0	5	14,29%	97,14% *****
>= 6136,6	7357,2	1	2,86%	100,00% *

Hipoteza zerowa: dystrybuanta empiryczna posiada rozkład normalny. Test Doornika-Hansena (1994) - transformowana skośność i kurtoza.:
 Chi-kwadrat(2) = 1,262 z wartością p 0,53208

Wartość p wynosi 0,53208, jest większa niż 0,05. Nie ma więc podstaw do odrzucenia hipotezy H0: dystrybuanta empiryczna posiada rozkład normalny. Stwierdzam więc, że rozkład reszt jest rozkładu normalnego. 

10. Testowanie autokorelacji

W celu sprawdzenia autokorelacji wykonano test Breuscha-Godfrey'a oparty o mnożniki Lagrange'a. Hipotezą zerową tego testu jest brak autokorelacji.

Test Breuscha-Godfrey'a na autokorelację do rzędu 6
 Estymacja KMNK, wykorzystane obserwacje 2007:1-2015:3 (N = 35)
 Zmienna zależna (Y): uhat

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-12,8144	637,287	-0,02011	0,9841
d_X2	-0,0785980	0,0760556	-1,033	0,3117
d_X4	-0,116428	0,220385	-0,5283	0,6021
d_X5	0,172259	0,352167	0,4891	0,6292

d_X8	0,742106	1,10895	0,6692	0,5098	
uhat_1	-0,107882	0,187134	-0,5765	0,5696	
uhat_2	-0,157599	0,188149	-0,8376	0,4105	
uhat_3	0,333181	0,213278	1,562	0,1313	
uhat_4	0,0708569	0,204458	0,3466	0,7319	
uhat_5	-0,0278292	0,197650	-0,1408	0,8892	
uhat_6	-0,509670	0,206689	-2,466	0,0212	**

Wsp. determ. R-kwadrat = 0,335924

Statystyka testu: LMF = 2,023406,
z wartością $p = P(F(6,24) > 2,02341) = 0,102$

Statystyka testu: $TR^2 = 11,757336$,
z wartością $p = P(\text{Chi-kwadrat}(6) > 11,7573) = 0,0676$

Ljung-Box $Q' = 14,7715$,
z wartością $p = P(\text{Chi-kwadrat}(6) > 14,7715) = 0,0221$

Dla jednego testu wartość p było mniejsze od 5% - Postaram się zatem usunąć autokorelację rzędu 6.

10.1. Usuwanie autokorelacji

W celu usunięcia autokorelacji zastosowano metodę Cochrane'a-Orcutta.
Skrypt:

```
gretl wersja 2016a-git
Bieżąca sesja: 2016-02-01 00:42
? eps=reszty
Wygenerowano serie eps (ID 22)
? eps6=reszty(-6)
Wygenerowano serie eps6 (ID 23)
? rho6=corr(eps,eps6)
Wygenerowano skalar rho6 = -0,480267
? d_Ynowy=d_Y-rho6*d_Y(-6)
Wygenerowano serie d_Ynowy (ID 24)
? d_X2nowy=d_X2-rho6*d_X2(-6)
Wygenerowano serie d_X2nowy (ID 25)
? d_X4nowy=d_X4-rho6*d_X4(-6)
Wygenerowano serie d_X4nowy (ID 26)
? d_X5nowy=d_X5-rho6*d_X5(-6)
Wygenerowano serie d_X5nowy (ID 27)
? d_X8nowy=d_X8-rho6*d_X8(-6)
Wygenerowano serie d_X8nowy (ID 28)
```

Poprawiony model:

Model 11: Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)
Zmienna zależna (Y): d_Ynowy

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	3275,93	626,232	5,2312	<0,0001	***
d_X2nowy	0,89353	0,0869521	10,2761	<0,0001	***
d_X4nowy	1,25568	0,19662	6,3863	<0,0001	***
d_X5nowy	1,18074	0,313124	3,7708	0,0009	***
d_X8nowy	7,5253	0,983734	7,6497	<0,0001	***

Średn. arytm. zm. zależnej	6003,939	Odch. stand. zm. zależnej	44915,40
Suma kwadratów reszt	2,46e+08	Błąd standardowy reszt	3199,119
Wsp. determ. R-kwadrat	0,995652	Skorygowany R-kwadrat	0,994927
F(4, 24)	1373,837	Wartość p dla testu F	5,92e-28

Logarytm wiarygodności	-272,4535	Kryt. inform. Akaike'a	554,9070
Kryt. bayes. Schwarza	561,7435	Kryt. Hannana-Quinna	557,0481
Autokorel.reszt - rho1	-0,081175	Stat. Durbina-Watsona	2,110607

Powtarzam test Breuscha-Godfrey:

Test Breuscha-Godfrey na autokorelację do rzędu 6
Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)
Zmienna zależna (Y): uhat

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-64,2229	706,397	-0,09092	0,9286
d_X2nowy	-0,0324180	0,108569	-0,2986	0,7687
d_X4nowy	-0,0736922	0,255671	-0,2882	0,7765
d_X5nowy	0,103036	0,371098	0,2777	0,7844
d_X8nowy	0,407607	1,30925	0,3113	0,7591
uhat_1	-0,0709332	0,248079	-0,2859	0,7782
uhat_2	0,0940998	0,244778	0,3844	0,7052
uhat_3	0,257038	0,260886	0,9853	0,3376
uhat_4	-0,0421056	0,266909	-0,1578	0,8764
uhat_5	0,0134094	0,252417	0,05312	0,9582
uhat_6	0,0448915	0,254734	0,1762	0,8621

Wsp. determ. R-kwadrat = 0,073146

Statystyka testu: LMF = 0,236755,
z wartością $p = P(F(6,18) > 0,236755) = 0,959$

Statystyka testu: $TR^2 = 2,121229$,
z wartością $p = P(\text{Chi-kwadrat}(6) > 2,12123) = 0,908$

Ljung-Box $Q' = 2,35233$,
z wartością $p = P(\text{Chi-kwadrat}(6) > 2,35233) = 0,885$

I dodatkowo wykonuje test Durbina-Watsona:

Stat. Durbina-Watsona = 2,11061
wartość $p = 0,724155$

Pozbyliśmy się zatem niepożądaney autokorelacji.

11.Badanie heteroskedastyczności

W celu sprawdzenia heteroskedastyczności wykonano test White'a oraz test Breuscha-Pagana.
Hipotezą zerową obu tych testów jest brak heteroskedastyczności.

Test White'a na heteroskedastyczność reszt (zmiennosc wariacji resztowej)
Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)
Zmienna zależna (Y): uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	1,19925e+07	8,04886e+06	1,490	0,1584
d_X2nowy	-532,249	621,200	-0,8568	0,4060
d_X4nowy	-457,694	797,992	-0,5736	0,5754
d_X5nowy	-188,218	1667,77	-0,1129	0,9117
d_X8nowy	4197,97	3784,91	1,109	0,2861
sq_d_X2nowy	-0,0335938	0,0347271	-0,9674	0,3498
X2_X3	-0,0220421	0,160032	-0,1377	0,8924
X2_X4	0,163222	0,345415	0,4725	0,6438
X2_X5	0,489840	0,728274	0,6726	0,5121
sq_d_X4nowy	-0,281318	0,252512	-1,114	0,2840
X3_X4	0,671292	0,567626	1,183	0,2567
X3_X5	2,42520	2,54804	0,9518	0,3573

sq_d_X5nowy	-0,552826	0,462812	-1,194	0,2521
X4_X5	-3,61791	3,31078	-1,093	0,2929
sq_d_X8nowy	-5,58396	5,34347	-1,045	0,3137

Wsp. determ. R-kwadrat = 0,454900

Statystyka testu: $TR^2 = 13,192086$,
z wartością $p = P(\text{Chi-kwadrat}(14) > 13,192086) = 0,511458$

Test Breuscha-Pagana na heteroskedastyczność
Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)
Zmienna zależna (Y): standaryzowane uhat^2

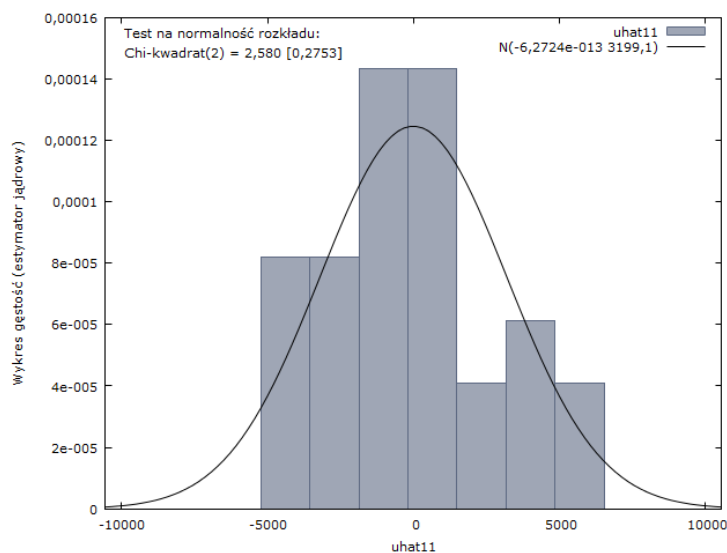
	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1,02764	0,203266	5,056	3,61e-05	***
d_X2nowy	-2,34985e-05	2,82234e-05	-0,8326	0,4133	
d_X4nowy	-9,11945e-05	6,38200e-05	-1,429	0,1659	
d_X5nowy	2,28250e-06	0,000101635	0,02246	0,9823	
d_X8nowy	0,000421015	0,000319306	1,319	0,1998	

Wyjaśniona suma kwadr. = 4,99058

Statystyka testu: LM = 2,495291,
z wartością $p = P(\text{Chi-kwadrat}(4) > 2,495291) = 0,645479$

W obu testach wartość p jest większa od poziomu istotności 0,05. Nie ma więc podstaw do odrzucenia hipotezy zerowej. W modelu nie występuje heteroskedastyczność.

12. Ponowne badanie normalności rozkładu reszt



Rozkład częstości dla uhat11, obserwacje 7-35
liczba przedziałów = 7, średnia = -6,27238e-013, odch.std. = 3199,12

Przedziały	średnia	liczba	częstość	skumulowana	
< -3550,7	-4392,4	4	13,79%	13,79%	***
-3550,7 - -1867,3	-2709,0	4	13,79%	27,59%	***
-1867,3 - -183,92	-1025,6	7	24,14%	51,72%	*****
-183,92 - 1499,5	657,77	7	24,14%	75,86%	*****
1499,5 - 3182,8	2341,1	2	6,90%	82,76%	**
3182,8 - 4866,2	4024,5	3	10,34%	93,10%	***
>= 4866,2	5707,9	2	6,90%	100,00%	**

Hipoteza zerowa: dystrybucja empiryczna posiada rozkład normalny. Test Doornika-Hansena (1994) - transformowana skośność i kurtosis.
Chi-kwadrat(2) = 2,580 z wartością p 0,27529

Wartość p testu przekracza 0,05, rozkład reszt jest normalny.

13. Testowanie współliniowości zmiennych modelu

Przeprowadzono test VIF w celu sprawdzenia współliniowości zmiennych. Z testu wynika, że w modelu może występować współliniowość.

Ocena współliniowości VIF(j) - czynnik rozdęcia wariancji
VIF (Variance Inflation Factors) - minimalna możliwa wartość = 1.0
Wartości > 10.0 mogą wskazywać na problem współliniowości - rozdęcia wariancji

```
d_X2nowy    3,739
d_X4nowy    14,096
d_X5nowy    2,508
d_X8nowy    16,027
```

$VIF(j) = 1/(1 - R(j)^2)$, gdzie $R(j)$ jest współczynnikiem korelacji wielorakiej pomiędzy zmienną 'j' a pozostałymi zmiennymi niezależnymi modelu.

Belsley-Kuh-Welsch collinearity diagnostics:

```
--- variance proportions ---
lambda    cond    const d_X2nowy d_X4nowy d_X5nowy d_X8nowy
3,106     1,000    0,002    0,020    0,007    0,022    0,006
1,043     1,725    0,770    0,002    0,001    0,024    0,001
0,588     2,298    0,115    0,125    0,002    0,375    0,005
0,230     3,675    0,103    0,505    0,079    0,242    0,036
0,033     9,705    0,011    0,348    0,912    0,338    0,951
```

lambda = eigenvalues of $X'X$, largest to smallest
cond = condition index
note: variance proportions columns sum to 1.0

13.1. Próba usunięcia współliniowości

Spróbuję zbudować model bez zmiennej d_X4nowy, gdyż podejrzewam, że zmienna ta może nam psuć właściwości modelu z powodu silnej korelacji z dobrze 'opisującą' zmienną d_X8nowy. Nowy model:

Model 13: Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)
Zmienna zależna (Y): d_Ynowy

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	2662,6	996,173	2,6728	0,0131	**
d_X2nowy	0,693367	0,130564	5,3105	<0,0001	***
d_X5nowy	2,49279	0,380393	6,5532	<0,0001	***
d_X8nowy	13,0153	0,769907	16,9050	<0,0001	***

Średn. arytm. zm. zależnej	6003,939	Odch. stand. zm. zależnej	44915,40
Suma kwadratów reszt	6,63e+08	Błąd standardowy reszt	5149,886
Wsp. determ. R-kwadrat	0,988262	Skorygowany R-kwadrat	0,986854
F(3, 25)	701,6236	Wartość p dla testu F	3,03e-24
Logarytm wiarygodności	-286,8523	Kryt. inform. Akaike'a	581,7046
Kryt. bayes. Schwarz	587,1738	Kryt. Hannana-Quinna	583,4175
Autokorel. reszt - rho1	-0,032180	Stat. Durbina-Watsona	2,038237

Ocena współliniowości VIF(j) - czynnik rozdęcia wariancji

VIF (Variance Inflation Factors) - minimalna możliwa wartość = 1.0

Wartości > 10.0 mogą wskazywać na problem współliniowości - rozdęcia wariancji

```
d_X2nowy    3,253
d_X5nowy    1,428
d_X8nowy    3,788
```

$VIF(j) = 1/(1 - R(j)^2)$, gdzie $R(j)$ jest współczynnikiem korelacji wielorakiej pomiędzy zmienną 'j' a pozostałymi zmiennymi niezależnymi modelu.

Belsley-Kuh-Welsch collinearity diagnostics:

```
--- variance proportions ---
lambda    cond    const d_X2nowy d_X5nowy d_X8nowy
2,245     1,000    0,009  0,048  0,074  0,044
1,026     1,479    0,784  0,009  0,030  0,013
0,575     1,977    0,157  0,101  0,791  0,014
0,154     3,820    0,050  0,842  0,105  0,929
```

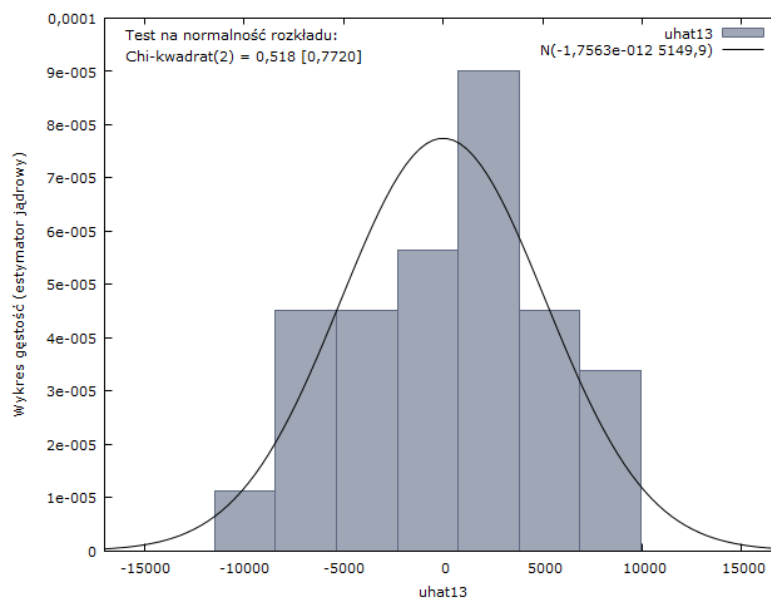
lambda = eigenvalues of $X'X$, largest to smallest

cond = condition index

note: variance proportions columns sum to 1.0

Współliniowość została usunięta. Nowy model wydaje się nie utracić dobrych właściwości, ale żeby się o tym przekonać powtarzam testy z poprzednich podpunktów:

- **Test normalności reszt**



Rozkład częstości dla uhat13, obserwacje 7-35

liczba przedziałów = 7, średnia = -1,75627e-012, odch.std. = 5149,89

Przedziały	średnia	liczba	częstość	skumulowana
< -8427,9	-9958,0	1	3,45%	3,45% *
-8427,9 - -5367,8	-6897,8	4	13,79%	17,24% ****
-5367,8 - -2307,6	-3837,7	4	13,79%	31,03% ****
-2307,6 - 752,49	-777,57	5	17,24%	48,28% *****
752,49 - 3812,6	2282,6	8	27,59%	75,86% *****
3812,6 - 6872,8	5342,7	4	13,79%	89,66% ****
>= 6872,8	8402,8	3	10,34%	100,00% ***

Hipoteza zerowa: dystrybuanta empiryczna posiada rozkład normalny. Test Doornika-Hansena (1994) - transformowana skośność i kurtোza.:

Chi-kwadrat(2) = 0,518 z wartością p 0,77196

- **Test na istnienie autokorelacji**

Test Breuscha-Godfrey na autokorelację do rzędu 6

Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)

Zmienna zależna (Y): uhat

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-101,219	1108,58	-0,09131	0,9282
d_X2nowy	0,0228715	0,158350	0,1444	0,8867
d_X5nowy	0,151335	0,460236	0,3288	0,7459
d_X8nowy	-0,146698	0,938172	-0,1564	0,8774
uhat_1	-0,0903953	0,248439	-0,3639	0,7200
uhat_2	0,110437	0,237397	0,4652	0,6471
uhat_3	0,277673	0,249217	1,114	0,2791
uhat_4	0,0420283	0,249743	0,1683	0,8681
uhat_5	-0,0625244	0,247539	-0,2526	0,8033
uhat_6	-0,0557644	0,271161	-0,2057	0,8393

Wsp. determ. R-kwadrat = 0,075231

Statystyka testu: LMF = 0,257612,
z wartością p = $P(F(6,19) > 0,257612) = 0,95$

Statystyka testu: $TR^2 = 2,181700$,
z wartością p = $P(\text{Chi-kwadrat}(6) > 2,1817) = 0,902$

Ljung-Box $Q' = 2,06198$,
z wartością p = $P(\text{Chi-kwadrat}(6) > 2,06198) = 0,914$

- **Test na istnienie heteroskedastyczność**

Test Breuscha-Pagana na heteroskedastyczność

Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)

Zmienna zależna (Y): standaryzowane uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1,15244	0,194543	5,924	3,50e-06	***
d_X2nowy	-2,45036e-05	2,54980e-05	-0,9610	0,3458	
d_X5nowy	-0,000202581	7,42873e-05	-2,727	0,0115	**
d_X8nowy	0,000230467	0,000150356	1,533	0,1379	

Wyjaśniona suma kwadr. = 8,23142

Statystyka testu: LM = 4,115708,
z wartością p = $P(\text{Chi-kwadrat}(3) > 4,115708) = 0,249238$

Heteroskedastyczność nadal nie istnieje aczkolwiek wartość p znacznie zmalała.

14. Test Ramsey'a RESET

Za pomocą poniższego testu chcę sprawdzić poprawność formy funkcyjnej modelu:

Hipotezą zerową tego testu jest założenie, że postać linowa modelu jest poprawna..

Test RESET na specyfikację (kwadrat i sześcián zmiennej)

Statystyka testu: F = 13,405284,
z wartością p = $P(F(2,23) > 13,4053) = 0,000138$

Test RESET na specyfikację (tylko kwadrat zmiennej)

Statystyka testu: F = 1,633549,

z wartością $p = P(F(1,24) > 1,63355) = 0,213$

Test RESET na specyfikację (tylko sześcián zmiennej)

Statystyka testu: $F = 24,408092$,

z wartością $p = P(F(1,24) > 24,4081) = 4,83e-005$

Niestety tylko jeden test potwierdził hipotezę H_0 . Wskazywać to może na istnienie innego modelu, który lepiej byłby dopasowany do danych.

14.1. Doprowadzenie do poprawności funkcyjnej

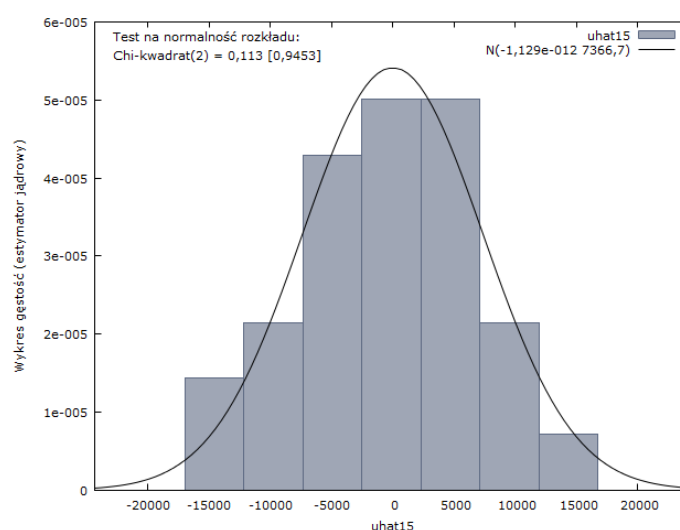
Spróbuję zatem raz jeszcze zmodyfikować mój model oraz sprawdzić jego parametry. Z podobnych powodów jak poprzednio (silna korelacja z dobrze 'opisującą' zmienną d_X8nowy) usunę z modelu zmienną d_X2nowy . Nowy model:

Model 15: Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)
Zmienna zależna (Y): d_Ynowy

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	3423,59	1410,17	2,4278	0,0224	**
d_X8nowy	16,2457	0,675093	24,0643	<0,0001	***
d_X5nowy	2,37528	0,543217	4,3726	0,0002	***

Średn. aryt. zm. zależnej	6003,939	Odch. stand. zm. zależnej	44915,40
Suma kwadratów reszt	1,41e+09	Błąd standardowy reszt	7366,719
Wsp. determ. R-kwadrat	0,975021	Skorygowany R-kwadrat	0,973100
F(2, 26)	507,4396	Wartość p dla testu F	1,47e-21
Logarytm wiarygodności	-297,8029	Kryt. inform. Akaike'a	601,6059
Kryt. bayes. Schwarz	605,7078	Kryt. Hannana-Quinna	602,8905
Autokorel. reszt - rho1	-0,086355	Stat. Durbina-Watsona	2,167942

• Test normalności reszt



Rozkład częstości dla $uhat15$, obserwacje 7-35

liczba przedziałów = 7, średnia = $-1,12903e-012$, odch.std. = 7366,72

Przedziały średnia liczba częstość skumulowana

< -12157,	-14565,	2	6,90%	6,90%	**
-12157, -	-7340,3	3	10,34%	17,24%	***
-7340,3 -	-2523,7	6	20,69%	37,93%	*****
-2523,7 -	2292,9	7	24,14%	62,07%	*****
2292,9 -	7109,5	7	24,14%	86,21%	*****
7109,5 -	11926,	3	10,34%	96,55%	***
>= 11926,	14334,	1	3,45%	100,00%	*

Hipoteza zerowa: dystrybucja empiryczna posiada rozkład normalny. Test Doornika-Hansena (1994) - transformowana skośność i kurtoza.:
Chi-kwadrat(2) = 0,113 z wartością p 0,94527

• Test na istnienie współliniowości

Ocena współliniowości VIF(j) - czynnik rozdęcia wariancji

VIF (Variance Inflation Factors) - minimalna możliwa wartość = 1.0
Wartości > 10.0 mogą wskazywać na problem współliniowości - rozdęcia wariancji

```
d_X8nowy    1,423
d_X5nowy    1,423
```

$VIF(j) = 1/(1 - R(j)^2)$, gdzie $R(j)$ jest współczynnikiem korelacji wielorakiej pomiędzy zmienną 'j' a pozostałymi zmiennymi niezależnymi modelu.

Belsley-Kuh-Welsch collinearity diagnostics:

```
--- variance proportions ---
lambda      cond      const d_X8nowy d_X5nowy
1,589      1,000      0,049   0,188   0,207
0,982      1,272      0,823   0,100   0,001
0,429      1,926      0,128   0,712   0,792
```

lambda = eigenvalues of $X'X$, largest to smallest
cond = condition index
note: variance proportions columns sum to 1.0

• Test na istnienie autokorelacji

Test Breuscha-Godfrey na autokorelację do rzędu 6

Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)
Zmienna zależna (Y): uhat

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	209,583	1592,79	0,1316	0,8966
d_X8nowy	0,277403	0,869639	0,3190	0,7530
d_X5nowy	-0,256316	0,862584	-0,2971	0,7694
uhat_1	-0,128564	0,244062	-0,5268	0,6041
uhat_2	-0,130712	0,226049	-0,5782	0,5696
uhat_3	0,240180	0,273524	0,8781	0,3903
uhat_4	0,0940894	0,238858	0,3939	0,6978
uhat_5	0,165672	0,268411	0,6172	0,5440
uhat_6	-0,0999601	0,274983	-0,3635	0,7200

Wsp. determ. R-kwadrat = 0,094002

Statystyka testu: LMF = 0,345849,
z wartością p = $P(F(6,20) > 0,345849) = 0,904$

Statystyka testu: $TR^2 = 2,726049$,
z wartością p = $P(\text{Chi-kwadrat}(6) > 2,72605) = 0,842$

Ljung-Box $Q' = 2,3122$,
z wartością p = $P(\text{Chi-kwadrat}(6) > 2,3122) = 0,889$

- **Test na istnienie heteroskedastyczność**

Test Breuscha-Pagana na heteroskedastyczność

Estymacja KMNK, wykorzystane obserwacje 2008:3-2015:3 (N = 29)

Zmienna zależna (Y): standaryzowane $uhat^2$

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1,08220	0,237824	4,550	0,0001	***
d_X8nowy	0,000151585	0,000113854	1,331	0,1946	
d_X5nowy	-0,000136811	9,16132e-05	-1,493	0,1474	

Wyjaśniona suma kwadr. = 4,02932

Statystyka testu: LM = 2,014661,

z wartością $p = P(\text{Chi-kwadrat}(2) > 2,014661) = 0,365193$

- **Test poprawności funkcyjnej modelu Reset**

Test RESET na specyfikację (kwadrat i sześćcian zmiennej)

Statystyka testu: F = 1,223031,

z wartością $p = P(F(2,24) > 1,22303) = 0,312$

Test RESET na specyfikację (tylko kwadrat zmiennej)

Statystyka testu: F = 0,112887,

z wartością $p = P(F(1,25) > 0,112887) = 0,74$

Test RESET na specyfikację (tylko sześćcian zmiennej)

Statystyka testu: F = 2,293970,

z wartością $p = P(F(1,25) > 2,29397) = 0,142$

Wszystkie wersje testu RESET potwierdziły poprawność funkcyjną modelu. Parametry modelu uległy nieznacznemu pogorszeniu ale nowy model przeszedł pozytywnie wszystkie wykonane testy. Dlatego też przyjmuję ten model jako ostateczny.

15.Badanie efektu katalizy

Zmienna X_i z pary zmiennych $\{X_i, X_j\}, i < j$ jest katalizatorem, jeśli: $r_{ij} < 0$ lub $r_{ij} > (r_i / r_j)$

Współczynniki korelacji liniowej dla obserwacji z próby 2008:3-2015:3

Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,3673 dla $n = 29$

d_Ynowy	d_X5nowy	d_X8nowy	
1,0000	0,6470	0,9781	d_Ynowy
	1,0000	0,5454	d_X5nowy
		1,0000	d_X8nowy

R0	R	
0,6470	1,0000	0,5454
0,9781	0,5454	1,0000

Zmienna d_X5nowy nie jest katalizatorem ponieważ:

$r_{ij} = 0,5454$ jest > 0 oraz $r_{ij} = 0,5454$ jest $< (r_i / r_j) = 0,6615$

16. Badanie koincydencji

	współczynnik	korelacja
d_X5nowy	2,37528	0,6470
d_X8nowy	16,2457	0,9781

W modelu zachodzi zjawisko koincydencji, ponieważ dla każdej zmiennej objaśniającej znak współczynnika stojącego przy zmiennej w modelu jest taki sam jak znak współczynnika korelacji ze zmienną objaśnianą.

17. Ostateczna postać modelu oraz jej interpretacja

Współczynnik determinacji R-kwadrat wynosi 0,975021, zmalał on od wstępnie wyestymowanego o 1,45%. Model opisuje 97,5% zjawiska i nie jest przeparametryzowany. Wartości kryteriów informacyjnych zmalały przy jednoczesnym wzroście logarytmu wiarygotności.

Równanie wyjściowego modelu:

$$Y = 3423,59 + (16,2457 \times d_X8) + (2,37528 \times d_X5)$$

gdzie:

d_X5: zmiana (w stosunku do poprzedniego kwartału) wartości rachunku narodowego wyprowadzanego przez branżę transport i gospodarka magazynowa

d_X8: zmiana (w stosunku do poprzedniego kwartału) wartości rachunku narodowego wyprowadzanego przez branżę działalność usługowa

A więc PKB w Polsce:

- wzrasta o 16,2457 jeżeli zmiana kwartalna rachunku narodowego w branży transport i gospodarka wzrosnie o jednostkę w stosunku do poprzedniego okresu (kwartału)
- - wzrasta o 2,37528 jeżeli zmiana kwartalna rachunku narodowego w branży działalność usługowa wzrosnie o jednostkę w stosunku do poprzedniego okresu (kwartału)

