

Projekt ekonometryczny

Badanie czynników wpływających na popyt krajowy

STUDENTKA

*Informatyka i Ekonometria
Wydział Zarządzania AGH
2015*

Spis treści

1.	Dane: opis i źródła-----	3
2.	Podstawowe charakterystyki zmiennych-----	3
3.	Dopasowanie modelu – wykresy zależności-----	4
4.	Wstępna analiza modelu -----	8
5.	Badanie korelacji między zmiennymi-----	9
6.	Metoda Hellwiga -----	10
7.	Budowanie modelu z uwzględnieniem istotnych zmiennych -----	10
8.	Badanie normalności rozkładów reszt -----	11
9.	Testowanie autokorelacji -----	12
10.	Usuwanie autokorelacji z modelu -----	12
11.	Badanie heteroskedastyczności-----	13
12.	Ponowne badanie normalności rozkładu reszt-----	15
13.	Testowanie współliniowości zmiennych modelu -----	16
14.	Test Ramsey'a RESET -----	16
15.	Badanie efektu katalizy-----	17
16.	Badanie koincydencji-----	188
17.	Ostateczna postać modelu oraz jej interpretacja -----	18

1. Dane: opis i źródła

Projekt dotyczy czynników, jakie wpływają na popyt krajowy w Polsce. Dane pochodzą ze strony internetowej Głównego Urzędu Statystycznego (www.stat.gov.pl) z zakładki „Wskaźniki makroekonomiczne”. Projekt zawiera 8 zmiennych objaśniających i opiera się na 44 kwartalnych obserwacjach z lat 2004 -2014.

Y: Popyt krajowy (ceny bieżące)

X1: Kurs oficjalny NBP 100 euro (zł)

X2: Eksport towarem ogółem (od początku roku do końca okresu) w mln zł

X3: Wskaźnik ogólnego klimatu koniunktury w handlu, naprawie pojazdów (w miesiącu kończącym okres)

X4: Podaż pieniądza M3 (stan na koniec okresu, mln zł)

X5: Stopa bezrobocia rejestrowanego ogółem

X6: Przeciętne miesięczne wynagrodzenie nominalne brutto (zł)

X7: Wskaźniki cen towarów i usług konsumpcyjnych

X8: Rachunek bieżący bilansu płatniczego importu towarów (mln Euro)

Link do pobrania pliku, z którego pochodzą dane:

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

2. Podstawowe charakterystyki zmiennych

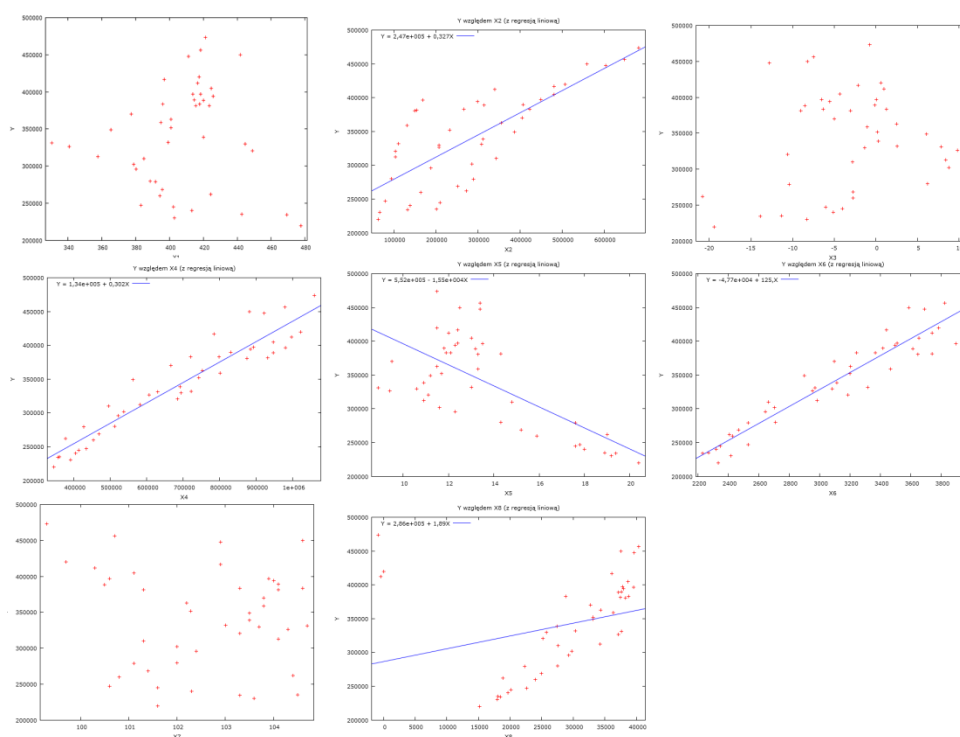
Podstawowe statystyki opisowe

	Y	X1	X2	X3	X4	X5	X6	X7	X8
Średnia	340 106,3	406,96	285852,6	-3,03409	683 698,8	13,6	3091,71	102,6	28374,82
Mediana	343 799,2	412,2	269 414,0	-2,8	692 480,7	12,8	3 149,7	102,9	30 066,5
Minimalna	219 832,4	330,6	62 998,1	-20,8	345 881,0	8,9	2 230,5	99,3	-826,0
Maksymalna	473 506,8	477,6	682 360,0	10,5	1 059 185,6	20,4	3 942,7	104,7	40 361,0
Odch. Stand.	69407,84	29,08	162516,39	7,21	217966,65	2,97	526,27	1,50	10751,53
Wsp. Zmienności	0,20408	0,071448	0,56853	2,3767	0,31881	0,21863	0,17022	0,014667	0,37891
Skośność	-0,06442	-0,17627	0,67507	-0,20004	-0,0011029	0,78855	-0,1293	-0,34699	-1,2144
Kurtoza	-0,99915	0,78227	-0,31578	-0,01826	-1,2988	-0,38643	-1,3006	-1,0514	1,0966

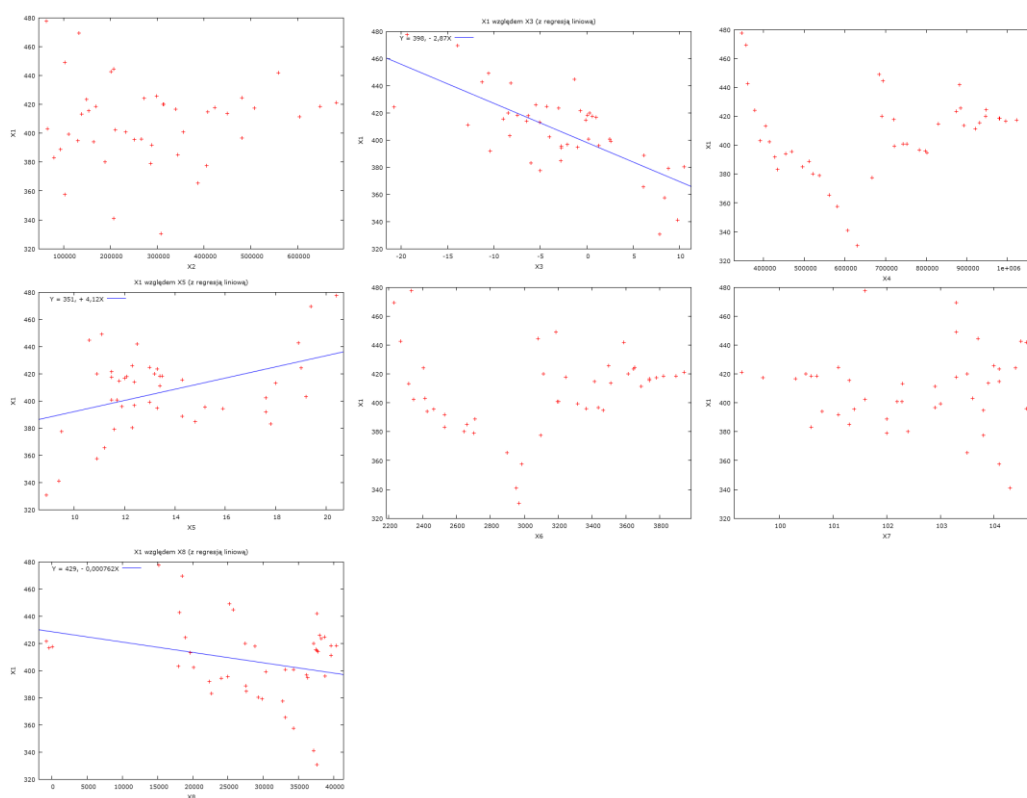
Zmienna objaśniająca waha się pomiędzy wartościami 219832,4 i 473506,8. Średnia i mediana nie są zbliżone ani do wartości minimalnych ani maksymalnych. Podobnie jest ze średnią i medianą pozostałych zmiennych.

3. Dopasowanie modelu – wykresy zależności

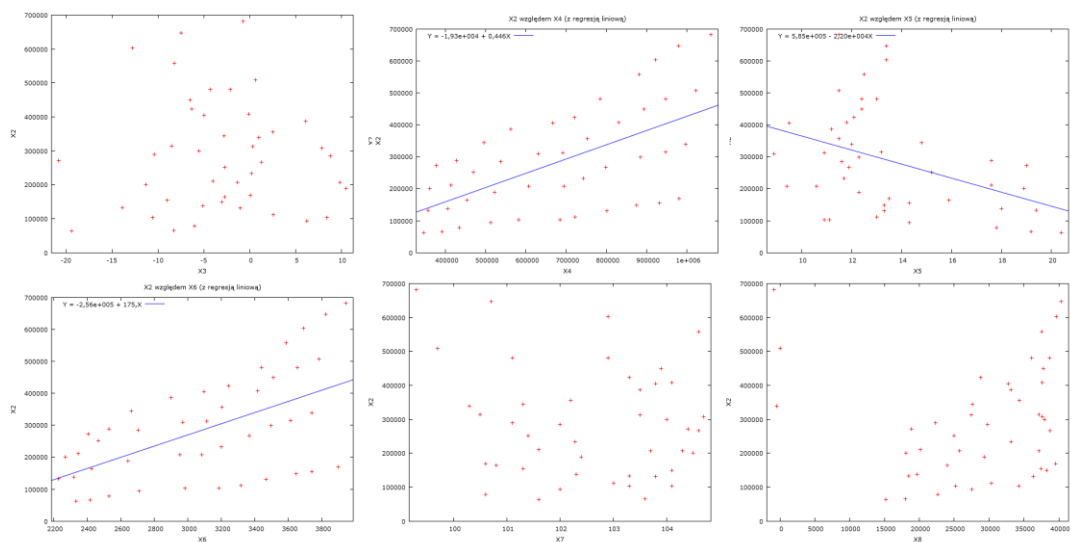
Wykresy zależności zmiennej Y od zmiennych $X1$ - $X8$ (w kolejności od lewej do prawej, zmienne X na osi X , zmienna Y na osi Y)



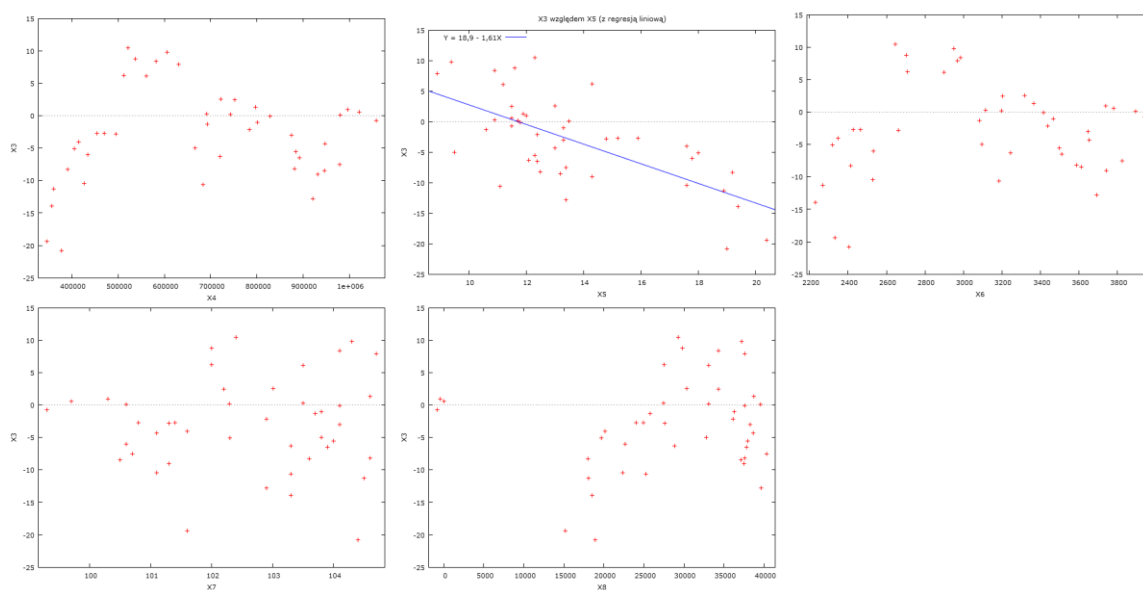
Wykresy zależności zmiennej $X1$ od zmiennych $X2$ - $X8$ (w kolejności od lewej do prawej, zmienne $X2$ - $X8$ na osi X , zmienna $X1$ na osi Y)



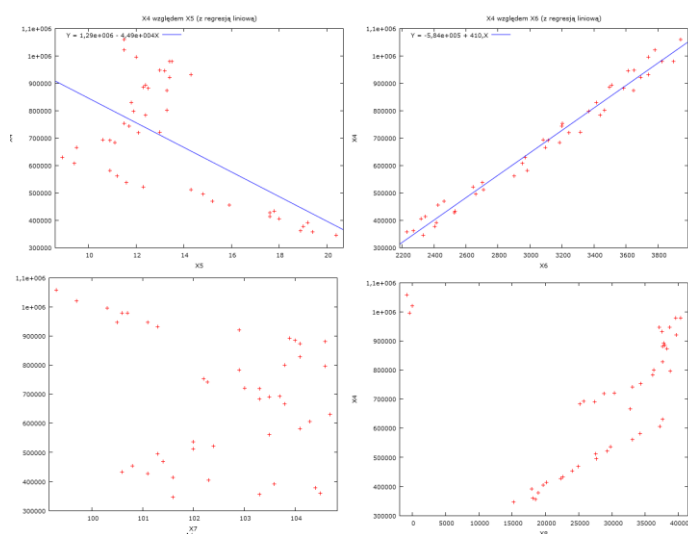
Wykresy zależności zmiennej X_2 od zmiennych X_3 - X_8 (w kolejności od lewej do prawej, zmienne X_3 - X_8 na osi X , zmienna X_2 na osi Y)



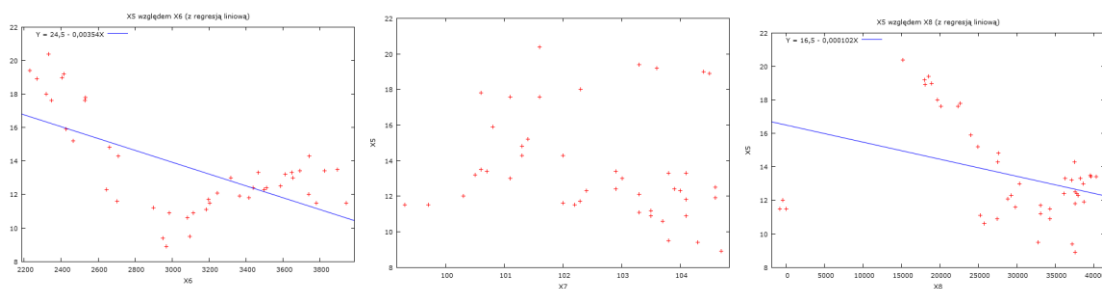
Wykresy zależności zmiennej X_3 od zmiennych X_4 - X_8 (w kolejności od lewej do prawej, zmienne X_4 - X_8 na osi X , zmienna X_3 na osi Y)



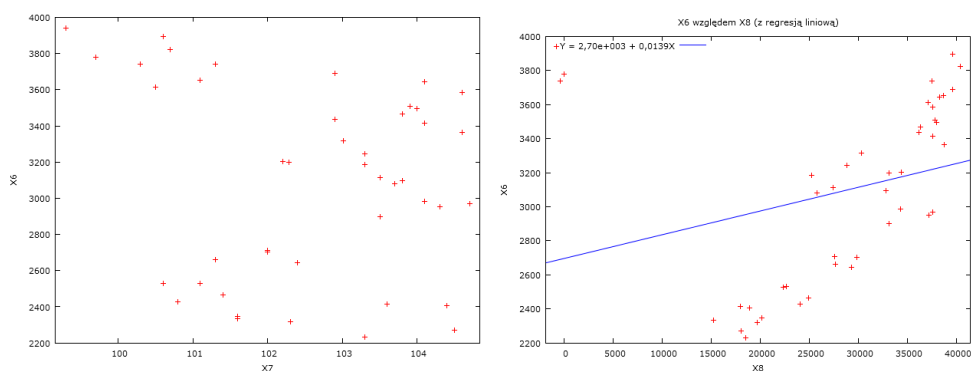
Wykresy zależności zmiennej X_4 od zmiennych X_5 - X_8 (w kolejności od lewej do prawej, zmienne X_5 - X_8 na osi X , zmienna X_4 na osi Y)



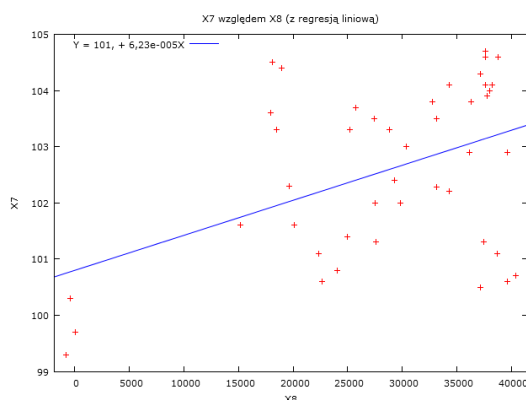
Wykresy zależności zmiennej X_5 od zmiennych X_6 - X_8 (w kolejności od lewej do prawej, zmienne X_6 - X_8 na osi X , zmienna X_5 na osi Y)



Wykresy zależności zmiennej X_6 od zmiennych X_7 - X_8 (w kolejności od lewej do prawej, zmienne X_7 - X_8 na osi X , zmienna X_6 na osi Y)



Wykresy zależności zmiennej X7 od zmiennej X8 (zmienna X8 na osi X, zmienna X7 na osi Y)



Na wykresach można zaobserwować zależności liniowe między poszczególnymi zmiennymi. Zmienna objaśniana Y jest zależna od pięciu zmiennych (X2, X4, X5, X6, X8), co zwiększa szanse, że model z tak dobranymi zmiennymi będzie poprawny. Istnieje jednak również sporo niekorzystnych zależności między poszczególnymi zmiennymi objaśniającymi X.

X1 jest skorelowane z: X3, X5, X8;

X2 z: X4, X5, X6;

X3 z: X1, X5;

X4 z: X2, X5, X6;

X5 z: X1, X2, X3, X4, X6, X8;

X6 z: X2, X4, X5, X8;

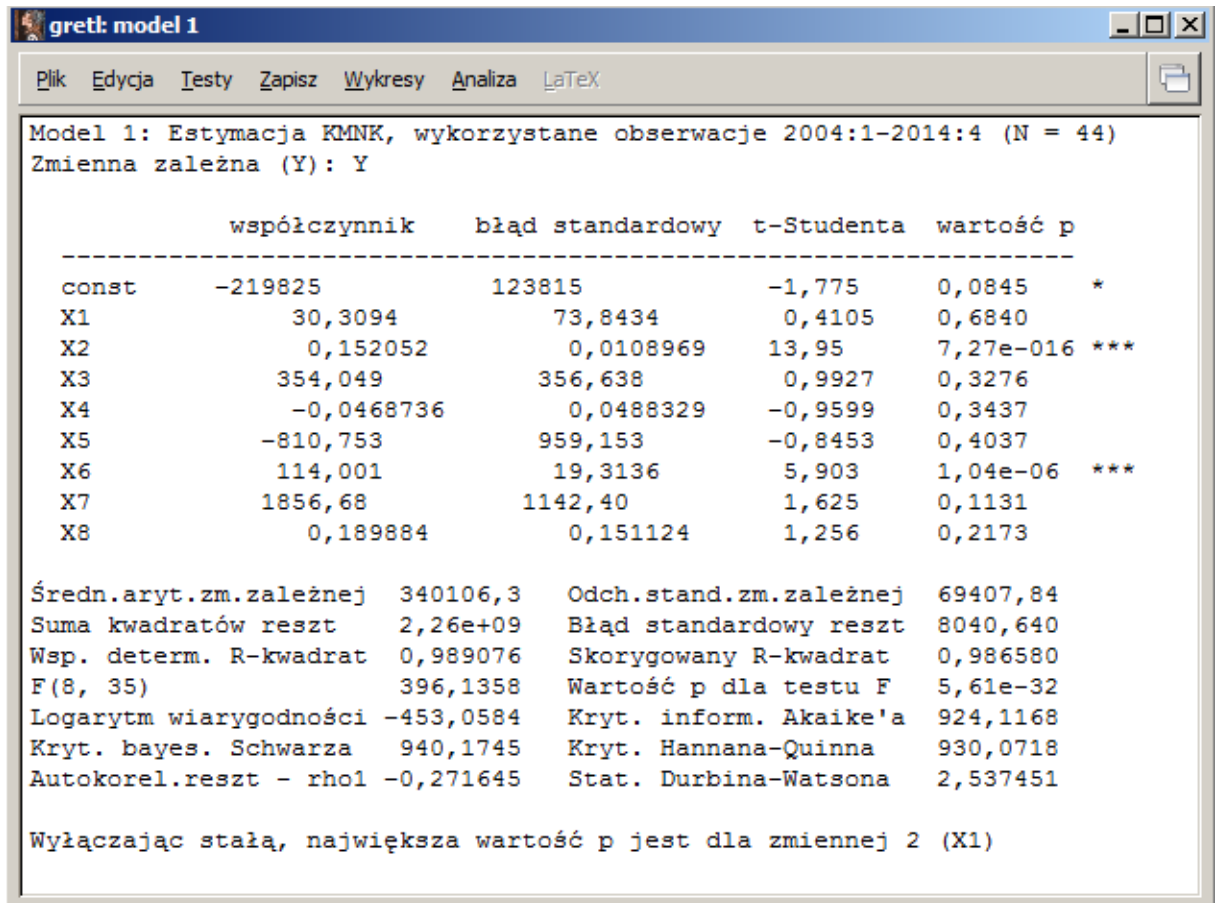
X7 z: X8;

X8 z: X1, X5, X7.

Mimo tak sporej liczby zależności istnieją jednak pary liczb nieskorelowane ze sobą, a jednocześnie skorelowane z Y (np. para X2-X8). Jeżeli jednak nie uda się znaleźć odpowiednio istotnej, nieskorelowanej pary być może będzie trzeba poszukać innych zmiennych objaśniających.

4. Wstępna analiza modelu

Wykonano estymację modelu klasyczną metodą najmniejszych kwadratów:



The screenshot shows the 'gretl: model 1' window. The title bar is blue with the text 'gretl: model 1'. Below the title bar is a menu bar with the following items: 'Plik', 'Edycja', 'Testy', 'Zapisz', 'Wykresy', 'Analiza', and 'LaTeX'. The main window area contains the following text:

Model 1: Estymacja KMNK, wykorzystane obserwacje 2004:1-2014:4 (N = 44)
Zmienna zależna (Y): Y

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	-219825	123815	-1,775	0,0845	*
X1	30,3094	73,8434	0,4105	0,6840	
X2	0,152052	0,0108969	13,95	7,27e-016	***
X3	354,049	356,638	0,9927	0,3276	
X4	-0,0468736	0,0488329	-0,9599	0,3437	
X5	-810,753	959,153	-0,8453	0,4037	
X6	114,001	19,3136	5,903	1,04e-06	***
X7	1856,68	1142,40	1,625	0,1131	
X8	0,189884	0,151124	1,256	0,2173	

Średn. arytm. zm. zależnej 340106,3 Odch. stand. zm. zależnej 69407,84
Suma kwadratów reszt 2,26e+09 Błąd standardowy reszt 8040,640
Wsp. determ. R-kwadrat 0,989076 Skorygowany R-kwadrat 0,986580
F(8, 35) 396,1358 Wartość p dla testu F 5,61e-32
Logarytm wiarygodności -453,0584 Kryt. inform. Akaike'a 924,1168
Kryt. bayes. Schwarza 940,1745 Kryt. Hannana-Quinna 930,0718
Autokorel. reszt - rho1 -0,271645 Stat. Durbina-Watsona 2,537451

Wyłączając stałą, największa wartość p jest dla zmiennej 2 (X1)

Największa wartość p jest dla zmiennej X1, co sugeruje, że powinna ona być wykluczona z modelu.

Wartość p dla testu F (H_0 : Wszystkie współczynniki równe 0) jest mniejsza od 0,05, odrzucono więc hipotezę o zerowości wszystkich współczynników.

Współczynnik determinacji R^2 wynosi 98,9%, co oznacza, że model wyjaśnia 98,9% zmienności badanego zjawiska.

5. Badanie korelacji między zmiennymi

gretl: macierz korelacji

Współczynniki korelacji, wykorzystane obserwacje 2004:1 - 2014:4
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,2973 dla n = 44

Y	X1	X2	X3	X4	
1,0000	0,0309	0,7666	0,1528	0,9483	Y
	1,0000	0,0283	-0,7130	0,1203	X1
		1,0000	-0,0282	0,5986	X2
			1,0000	0,1451	X3
				1,0000	X4
X5	X6	X7	X8		
-0,6661	0,9512	-0,1052	0,2929		Y
0,4217	0,0916	-0,1044	-0,2817		X1
-0,4024	0,5676	-0,1683	0,0513		X2
-0,6639	0,1471	0,0265	0,1764		X3
-0,6130	0,9899	-0,2076	0,2350		X4
1,0000	-0,6262	-0,1763	-0,3688		X5
	1,0000	-0,1562	0,2848		X6
		1,0000	0,4453		X7
			1,0000		X8

Z macierzy korelacji wynika, że zmienne X1, X3 oraz X7 są najmniej skorelowane ze zmienną objaśnianą Y (współczynnik korelacji X1-Y wynosi 0,0309, X3-Y: 0,1528, a X7-Y: 0,1052), dlatego zostaną one wykluczone z modelu.

Zmienna X8 ma trochę silniejszą korelację z Y niż wykluczone zmienne, jednak wciąż jest ona tak słaba, że tę zmienną również postanowiono usunąć.

Pozostałe zmienne są silnie skorelowane z Y, jednak są również zależne od innych zmiennych X. Najsilniejsza jest korelacja pomiędzy X4 a X6 (0,9899), dlatego należy usunąć jedną z tych zmiennych. Zdecydowano, że lepiej wykluczyć zmienną X4, ponieważ ma trochę słabszą korelację z Y i jednocześnie ma silniejszą niż X6 korelację z X2 (która jest jedną z 3 zmiennych o najsilniejszej korelacji z Y).

Po wstępnej analizie macierzy korelacji zdecydowano pozostawić zmienne X2, X5, X6, jednak postanowiono przeprowadzić test istotności zmiennych metodą Hellwiga i wtedy dokonać wyboru.

6. Metoda Hellwiga

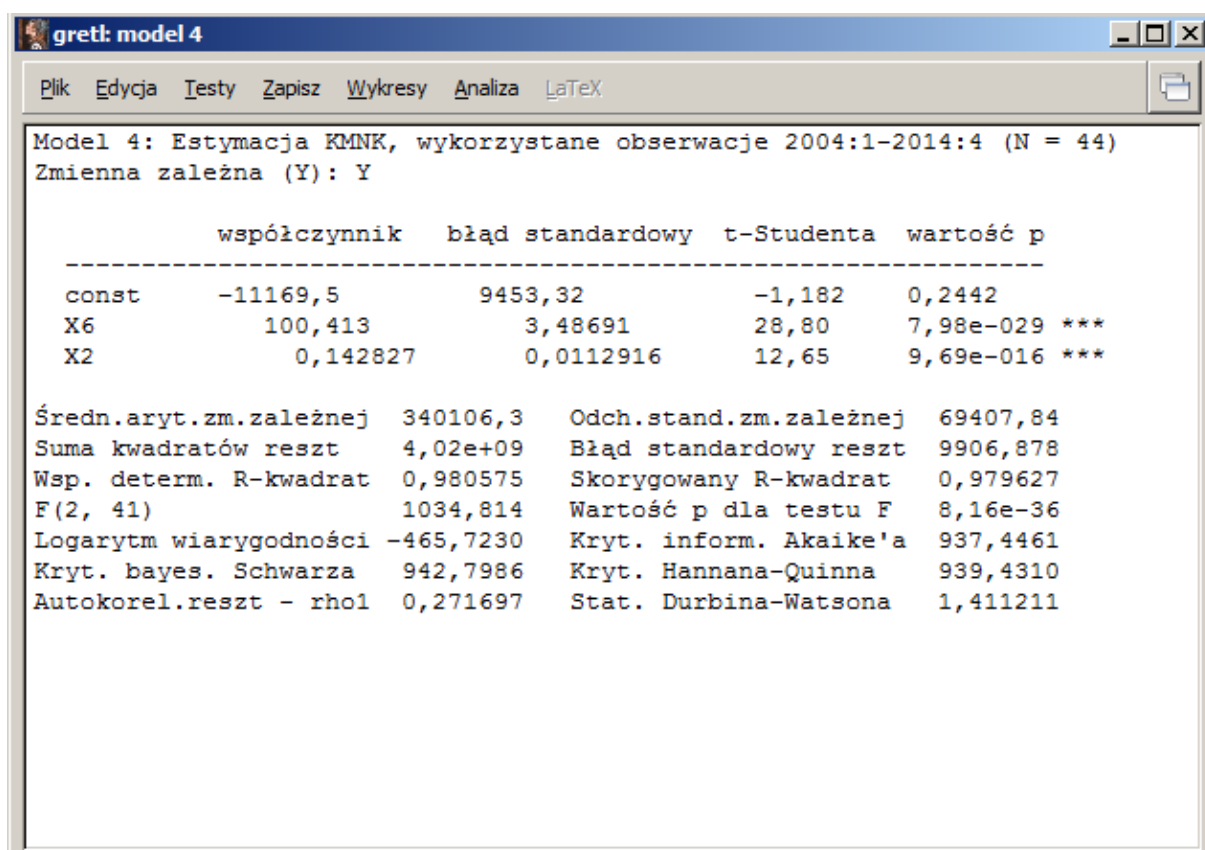
```
? H_max
0,97245413
? najlepszalista
X2 X4 X6
```

Metoda Hellwiga wskazała zmienne X2 X4 i X6 jako najbardziej istotne. Pojemność integralna przy tym zestawie wynosi 0,97245 .

Po wzięciu pod uwagę korelacji między zmiennymi, zwłaszcza wysokiej korelacji między zmiennymi X4 i X6 oraz wyniku metody Hellwiga zdecydowano, by włączyć do modelu zmienne X2 oraz X6.

7. Budowanie modelu z uwzględnieniem istotnych zmiennych

Wyestymowano nowy model KMNK z uwzględnieniem dwóch zmiennych X2 i X6:



Model 4: Estymacja KMNK, wykorzystane obserwacje 2004:1-2014:4 (N = 44)
Zmienna zależna (Y): Y

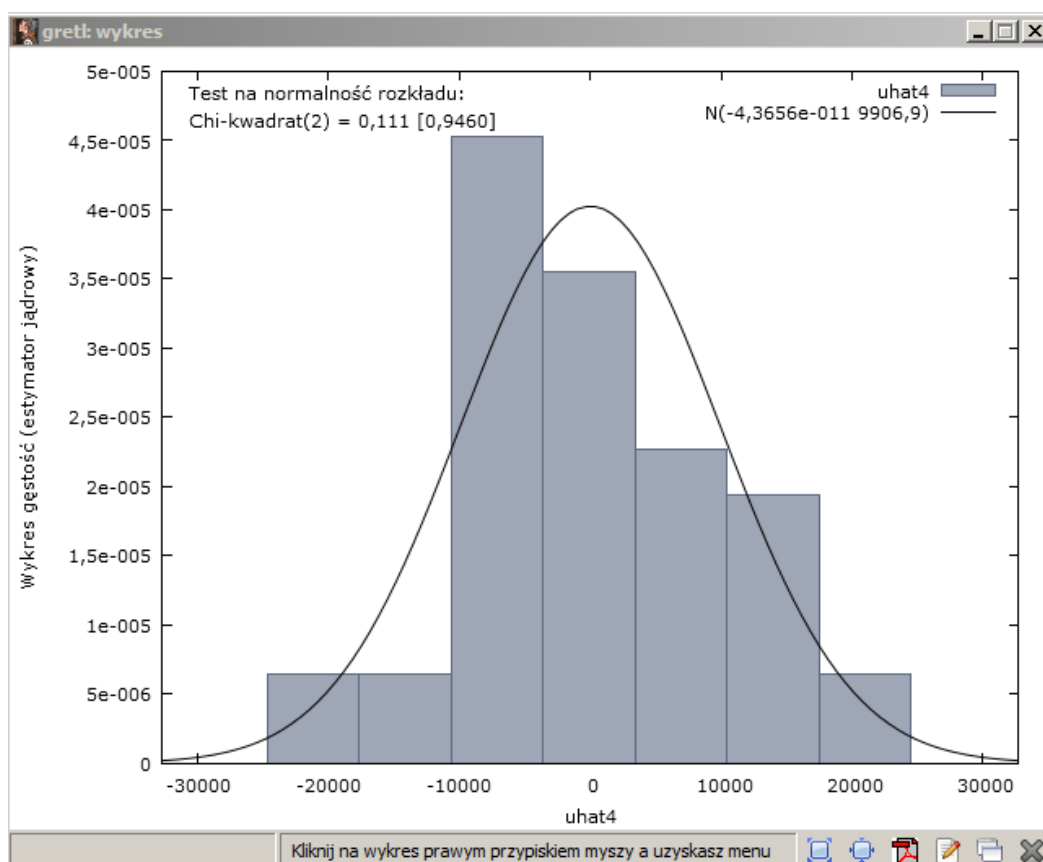
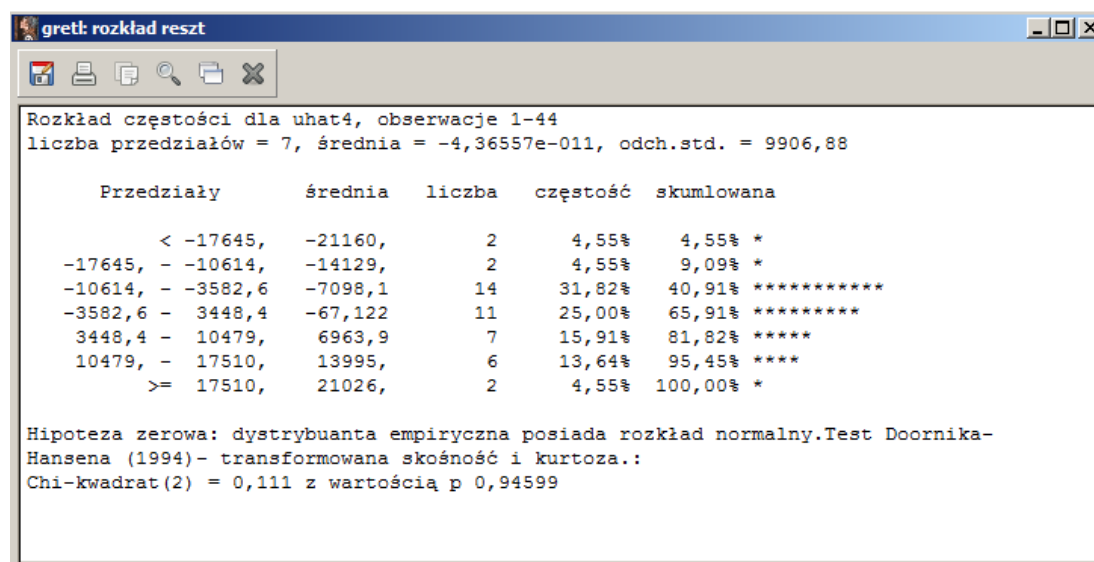
	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-11169,5	9453,32	-1,182	0,2442
X6	100,413	3,48691	28,80	7,98e-029 ***
X2	0,142827	0,0112916	12,65	9,69e-016 ***

Średn. aryt. zm. zależnej	340106,3	Odch. stand. zm. zależnej	69407,84
Suma kwadratów reszt	4,02e+09	Błąd standardowy reszt	9906,878
Wsp. determ. R-kwadrat	0,980575	Skorygowany R-kwadrat	0,979627
F(2, 41)	1034,814	Wartość p dla testu F	8,16e-36
Logarytm wiarygodności	-465,7230	Kryt. inform. Akaike'a	937,4461
Kryt. bayes. Schwarza	942,7986	Kryt. Hannana-Quinna	939,4310
Autokorel. reszt - rho1	0,271697	Stat. Durbina-Watsona	1,411211

Współczynnik determinacji R^2 wynosi 0,980575, co oznacza, że zmienna Y jest objaśniana w 98% przez zmienne objaśniające. Natomiast różnica pomiędzy R^2 a skorygowanym R^2 wynosi 0,00948, więc model nie jest przeparametryzowany. Zmniejszył się logarytm wiarygodności i zwiększyły kryteria informacyjne, co nie jest

korzystne. Nie jest to jednak ostateczny model, prawdopodobnie w dalszych etapach tworzenia modelu wartości ulegną poprawie.

8. Badanie normalności rozkładów reszt



Wartość p wynosi 0,94599, jest większa niż 0,05. Nie ma więc podstaw do odrzucenia hipotezy H_0 : dystrybuanta empiryczna posiada rozkład normalny. Stwierdzono więc, że rozkład reszt jest rozkładu normalnego.

9. Testowanie autokorelacji

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

gret: autokorelacja

Test Breuscha-Godfrey'a na autokorelację do rzędu 6
Estymacja KMNK, wykorzystane obserwacje 2004:1-2014:4 (N = 44)
Zmienna zależna (Y): uhat

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	-5295,63	5678,20	-0,9326	0,3574	
X2	-0,0169498	0,00833005	-2,035	0,0495	**
X6	3,31033	2,24659	1,473	0,1496	
uhat_1	0,209208	0,165234	1,266	0,2138	
uhat_2	0,542098	0,166170	3,262	0,0025	***
uhat_3	0,0813102	0,160897	0,5054	0,6165	
uhat_4	0,566169	0,169978	3,331	0,0021	***
uhat_5	-0,566784	0,172636	-3,283	0,0023	***
uhat_6	-0,290792	0,189947	-1,531	0,1348	

Wsp. determ. R-kwadrat = 0,746856

Statystyka testu: LMF = 17,210249,
z wartością p = $P(F(6,35) > 17,2102) = 3,78e-009$

Statystyka testu: $TR^2 = 32,861686$,
z wartością p = $P(\text{Chi-kwadrat}(6) > 32,8617) = 1,11e-005$

Ljung-Box $Q' = 44,7812$,
z wartością p = $P(\text{Chi-kwadrat}(6) > 44,7812) = 5,17e-008$

Dla wszystkich statystyk testowych wartości p są małe, dużo mniejsze niż 5%, należy więc odrzucić hipotezę zerową i przyjąć, że model zawiera autokorelację.

10. Usuwanie autokorelacji z modelu

W celu usunięcia autokorelacji zastosowano metodę Cochrane'a-Orcutta. Zastosowano transformację przy pomocy kodu w skrypcie:

ols Y const X2 X6

ro = \$rho

series ygw = Y - ro*Y(-5)

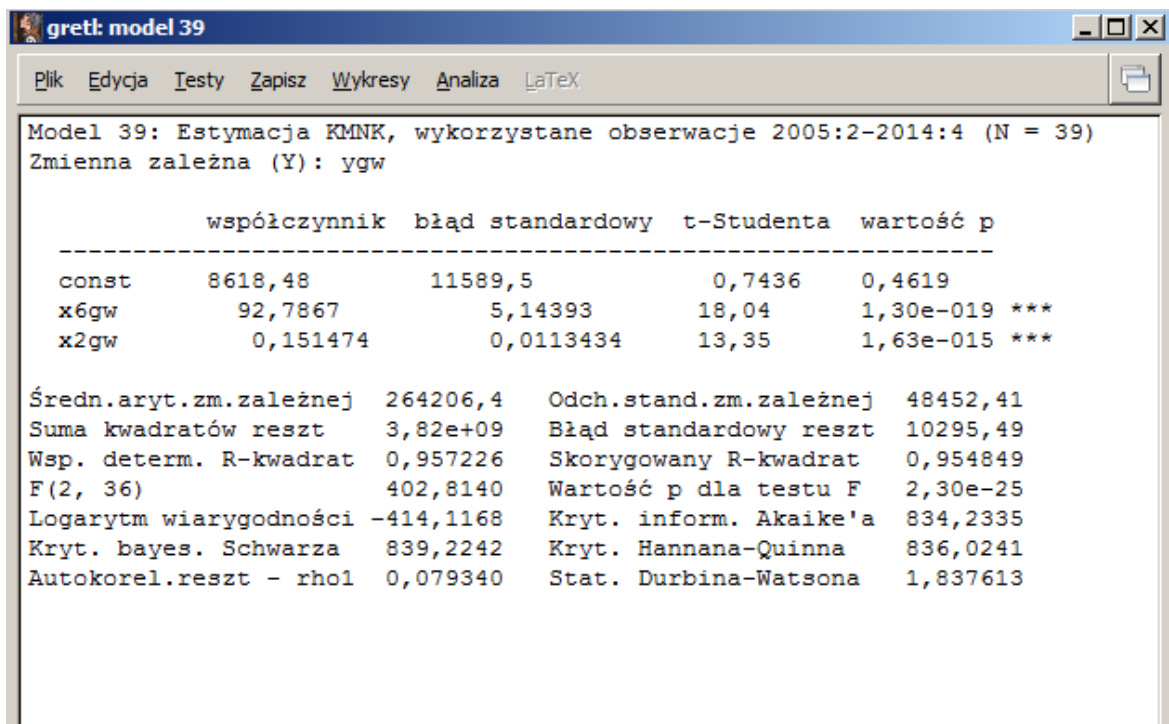
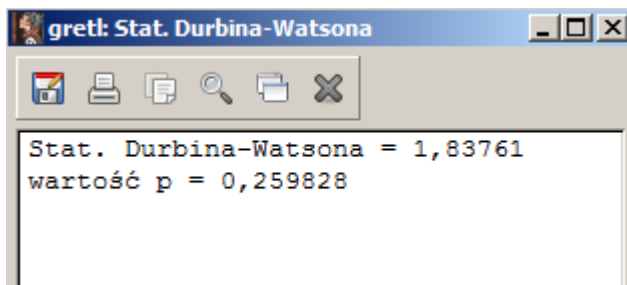


series x6gw = X6- ro*X6(-5)

series x2gw = X2- ro*X2(-5),

a następnie wyestymowano model: ygw const x2gw x6gw.

Wartość autokorelacji reszt – rho1 spadła z 0,271697 do 0,079340. Wartość p testu Durбина-Watsona wynosi 0,2598, czyli znacznie przewyższa wartość przeciętnego poziomu istotności 0,05. Nie ma więc podstaw do odrzucenia hipotezy zerowej o braku autokorelacji.



11. Badanie heteroskedastyczności

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

gretl: test LM - mnożnika Lagrange'a (heteroskedastyczność)

Test White'a na heteroskedastyczność reszt (zmiennosc wariacji resztowej)
 Estymacja KMNK, wykorzystane obserwacje 2005:2-2014:4 (N = 39)
 Zmienna zależna (Y): uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	2,66589e+07	9,98227e+08	0,02671	0,9789
x6gw	8559,37	898892	0,009522	0,9925
x2gw	-578,893	1410,76	-0,4103	0,6842
sq_x6gw	-4,60113	199,408	-0,02307	0,9817
X2_X3	0,613403	0,682696	0,8985	0,3754
sq_x2gw	-0,00177134	0,00149196	-1,187	0,2436

Wsp. determ. R-kwadrat = 0,164997

Statystyka testu: $TR^2 = 6,434872$,
 z wartością $p = P(\text{Chi-kwadrat}(5) > 6,434872) = 0,266172$

gretl: test LM - mnożnika Lagrange'a (heteroskedastyczność)

Test Breuscha-Pagana na heteroskedastyczność
 Estymacja KMNK, wykorzystane obserwacje 2005:2-2014:4 (N = 39)
 Zmienna zależna (Y): standaryzowane uhat^2

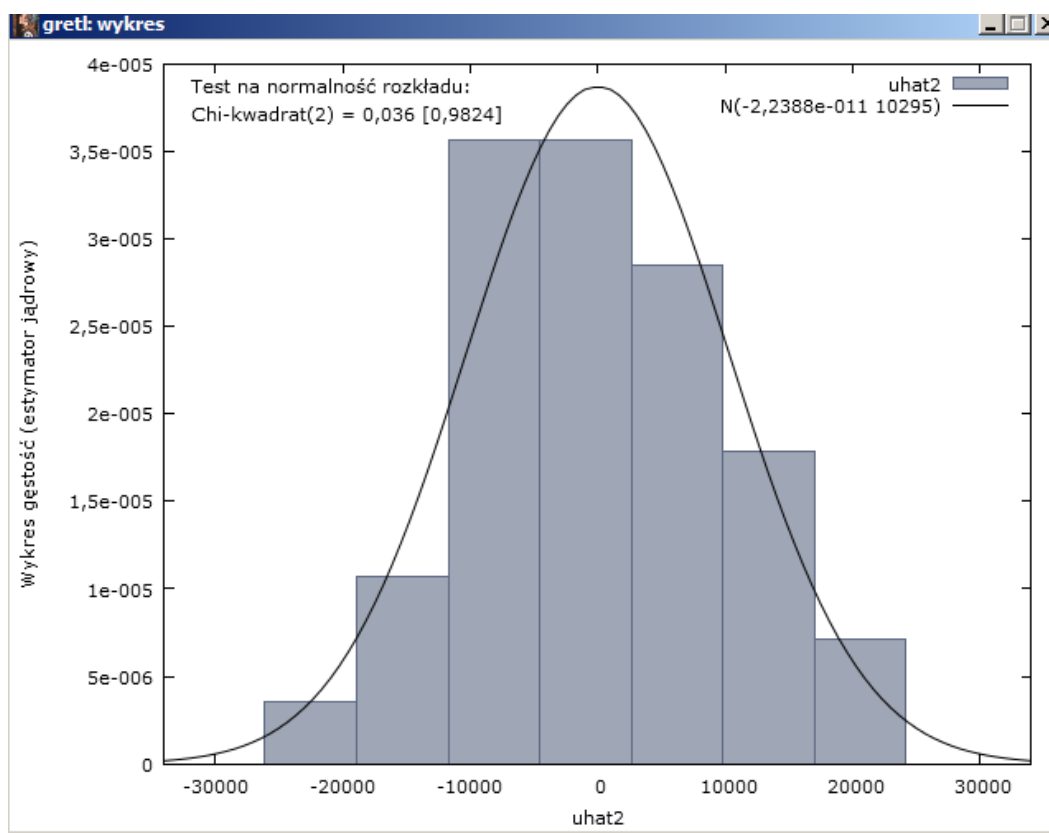
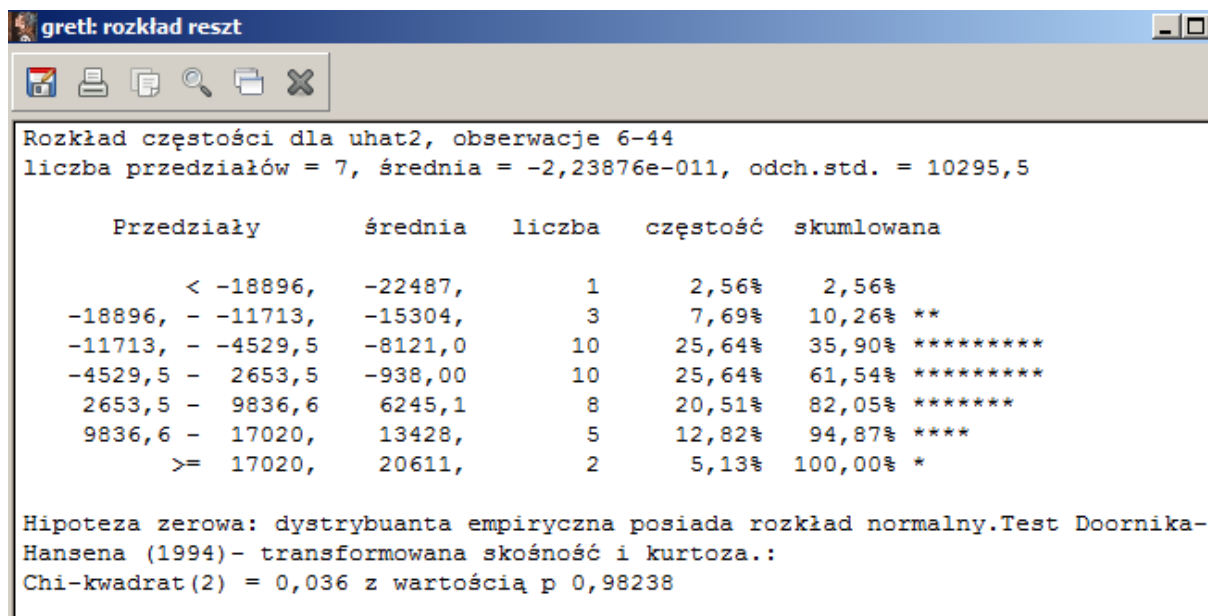
	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-0,0716129	1,40335	-0,05103	0,9596
x6gw	0,000190221	0,000622868	0,3054	0,7618
x2gw	2,66711e-06	1,37355e-06	1,942	0,0600 *

Wyjaśniona suma kwadr. = 7,8582

Statystyka testu: LM = 3,929098,
 z wartością $p = P(\text{Chi-kwadrat}(2) > 3,929098) = 0,140219$

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.

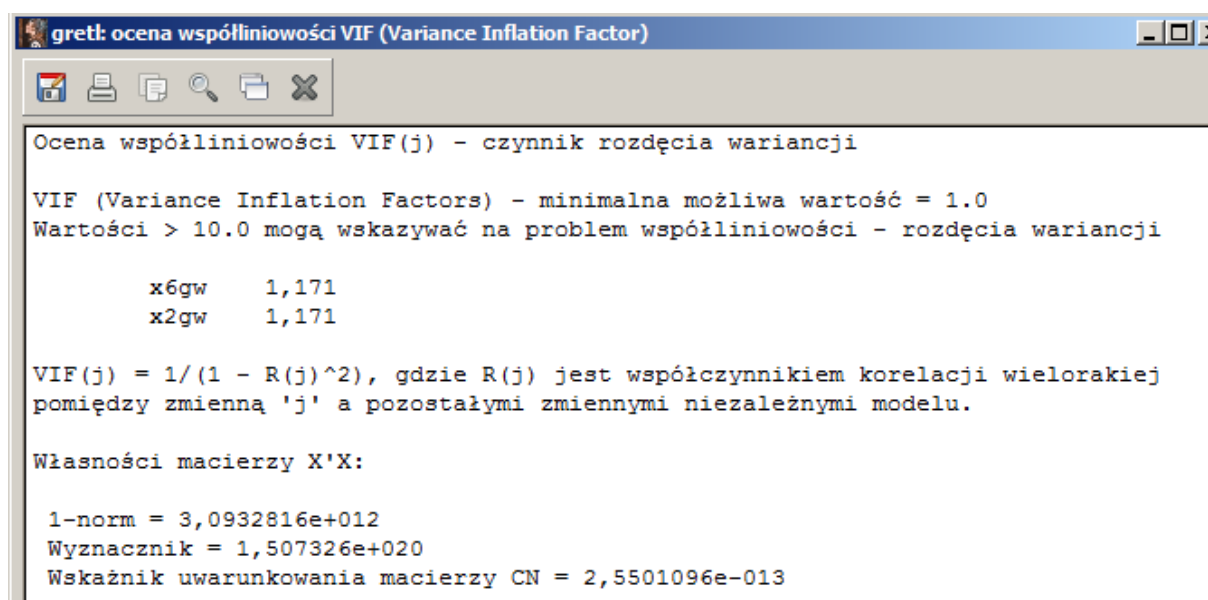
12. Ponowne badanie normalności rozkładu reszt



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 nie występuje współliniowość, ponieważ wartość czynnika jest mniejsza od 10, co więcej znajduje się w pobliżu minimalnej wartości czynnika.

The screenshot shows a window titled "gretl: ocena współliniowości VIF (Variance Inflation Factor)". It contains text explaining the VIF test, its formula, and the results for variables x6gw and x2gw. The results show VIF values of 1.171 for both variables, which are well below the threshold of 10.0.

```
gretl: ocena współliniowości VIF (Variance Inflation Factor)

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

      x6gw      1,171
      x2gw      1,171

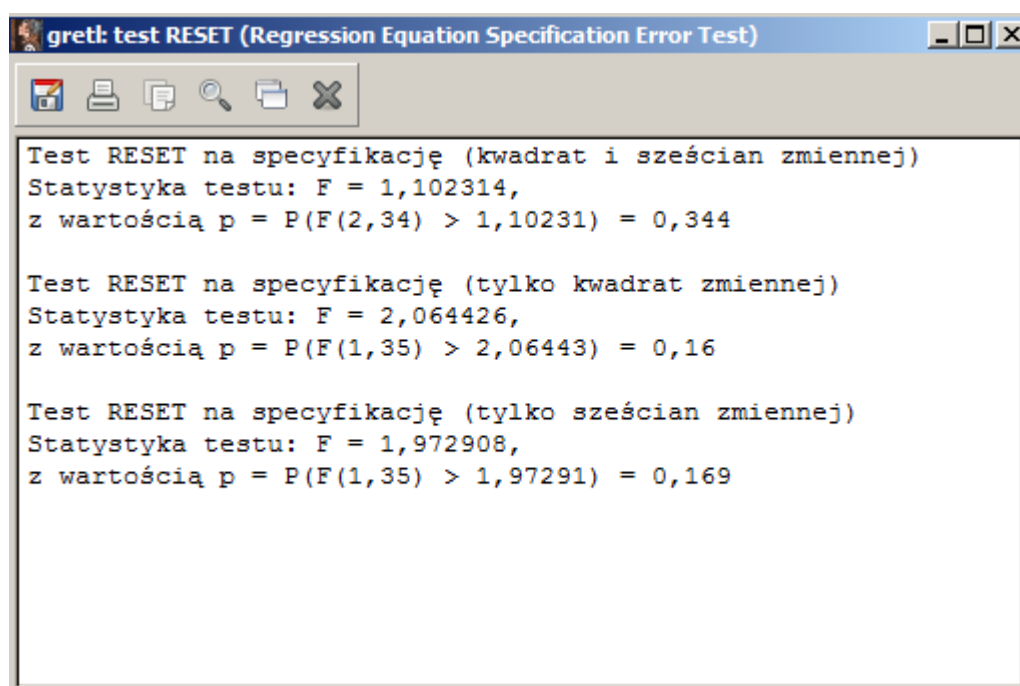
VIF(j) = 1/(1 - R(j)^2), gdzie R(j) jest współczynnikiem korelacji wielorakiej
pomiedzy zmienną 'j' a pozostałymi zmiennymi niezależnymi modelu.

Własności macierzy X'X:

1-norm = 3,0932816e+012
Wyznacznik = 1,507326e+020
Wskaźnik uwarunkowania macierzy CN = 2,5501096e-013
```

14. Test Ramsey'a RESET

Hipotezą zerową tego testu jest liniowość modelu.

The screenshot shows a window titled "gretl: test RESET (Regression Equation Specification Error Test)". It displays the results of three RESET tests: one for the full specification (quadratic and cubic terms), one for the quadratic term only, and one for the cubic term only. All three tests have p-values greater than 0.1, indicating no significant non-linearity.

```
gretl: test RESET (Regression Equation Specification Error Test)

Test RESET na specyfikację (kwadrat i sześćcian zmiennej)
Statystyka testu: F = 1,102314,
z wartością p = P(F(2,34) > 1,10231) = 0,344

Test RESET na specyfikację (tylko kwadrat zmiennej)
Statystyka testu: F = 2,064426,
z wartością p = P(F(1,35) > 2,06443) = 0,16

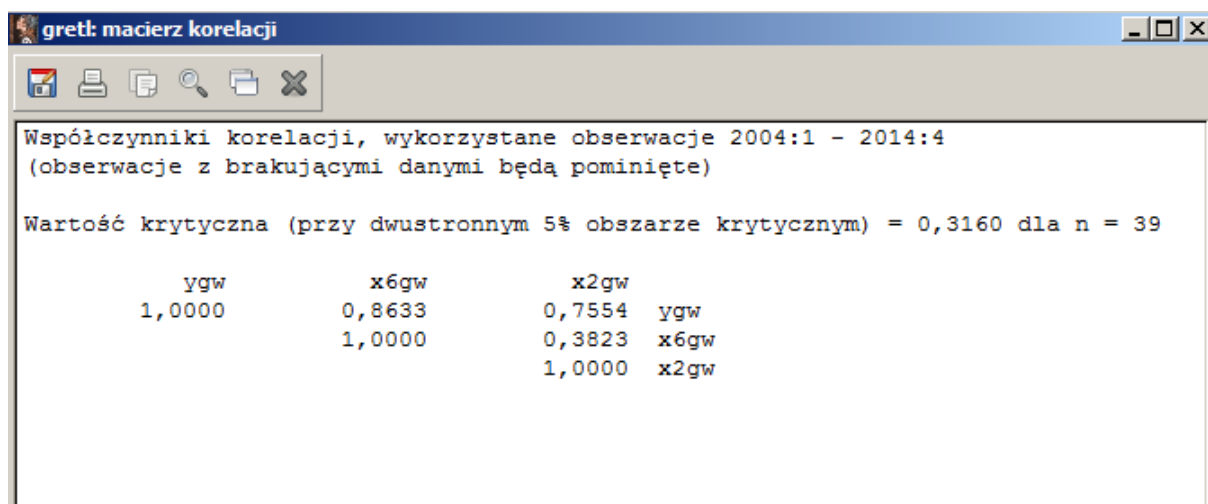
Test RESET na specyfikację (tylko sześćcian zmiennej)
Statystyka testu: F = 1,972908,
z wartością p = P(F(1,35) > 1,97291) = 0,169
```


Ponieważ wszystkie wartości p są większe od 0,05, nie ma podstaw do odrzucenia hipotezy o liniowości modelu.

15. Badanie efektu katalizy

Zmienna X_i z pary zmiennych $\{X_i, X_j\}, i < j$, jest katalizatorem, jeśli:

$$r_{ij} < 0 \text{ lub } r_{ij} > \frac{r_i}{r_j}.$$



R0
0,7554
0,8633

R	
1	0,3823
0,3823	1

$\frac{r_2}{r_6}$	r_{26}
0,875014	0,3823

Z macierzy korelacji wynika, że korelacja zmiennej x2gw z x6gw nie jest mniejsza od 0.

Korelacja x2gw z ygw podzielona przez korelację x6gw z ygw wynosi: 0,875014, nie jest więc mniejsza niż korelacja x2gw z x6gw.

Z tego wynika, że zmienna x2gw nie jest katalizatorem, nie zaburza więc wyniku.

16. Badanie koincydencji

Model ekonometryczny posiada własność koincydencji, jeśli dla każdej zmiennej objaśniającej znak współczynnika stojącego przy zmiennej w modelu jest równy znakowi współczynnika korelacji ze zmienną objaśnianą.

	współczynnik	korelacja
x2gw	0,151474	0,7554
x6gw	92,7867	0,8633

W tym modelu zachodzi zjawisko koincydencji, ponieważ spełniony jest warunek, że dla każdej zmiennej znak współczynnika jest równy znakowi współczynnika korelacji.

17. Ostateczna postać modelu oraz jej interpretacja

Współczynnik determinacji modelu wynosi 0,957, a różnica między R^2 a skorygowanym R^2 wynosi 0,002377. Model opisuje więc 95,7% zjawiska, przy czym nie jest przeparametryzowany. W porównaniu z modelem wejściowym wartość R^2 zmalała o 3%. Zmalały wartości kryteriów informacyjnych przy jednoczesnym wzroście logarytmu wiarygodności.

$$Y = 8618,48 + 0,151474 \cdot x2gw + 92,7867 \cdot x6gw$$

W modelu uwzględniono dwie zmienne objaśniające:

X2: Eksport towarem ogółem

X6: Przeciętne miesięczne wynagrodzenie nominalne

Popyt krajowy:

- wzrasta o 0,151474 jeżeli eksport towarów wzrośnie o jednostkę w stosunku do kwartału poprzedniego
- wzrasta o 92,7867 jeżeli przeciętne miesięczne wynagrodzenie nominalne brutto wzrośnie o jednostkę w stosunku do kwartału poprzedniego