Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział Zarządzania

Model ekonometryczny

Autorka

Kraków 2017

Budownictwo jest powszechnie postrzegane jako dział gospodarki najbardziej ożywiający koniunkturę i istotnie wpływający na wielkość PKB. Znaczenie sektora budowlanego można określić bezpośrednio w wartości dodanej, czyli bezpośrednim wkładzie budownictwa do PKB. Jednakże działalność w sektorze budownictwa skutkuje dodatkową wartością dodaną w innych sektorach i branżach gospodarki, co jest już trudniej mierzalne. Szacuje się, że co najmniej 30% społeczeństwa jest zaangażowanych w branżę budowlaną, jako beneficjenci, wykonawcy czy producenci.

Wiele czynników może mieć wpływ na wyniki w budownictwie. Mogą się one odnosić nie tylko do kosztów w budownictwie, ale i do poziomu inwestycji w gospodarce, być związane z m.in. z możliwościami ekonomicznymi przemysłu, koniunkturą w budownictwie mieszkaniowym. Równie istotnymi czynnikami predykcyjnymi są psychologiczne, związane ze wzrostem pesymizmu lub optymizmu w gospodarce ogółem, czy przedsiębiorców branży budowlanej. Na ich podstawie sporządzane są również prognozy co do wzrostu lub spadku wyników w budownictwie.

Na podstawie literatury przedmiotu można wytypować kilka potencjalnych czynników, które mogą wpływać na aspekt charakteryzujący wkład budownictwa w polską gospodarkę – wartość dodaną budownictwa ogółem. Warto zatem podjąć próbę wyestymowania modelu, który charakteryzowałby zależność wyników budownictwa w gospodarce od wybranych czynników. **Ogólnie postać analityczną** takiego **modelu** można zapisać jako:

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ \ldots + \alpha_n x_n + \varepsilon$$

Model ten ma przedstawiać zależność między zmienną objaśnianą Y a zmiennymi objaśniającymi X_n. Przyjmuje się w pracy, że wartość dodana brutto ogółem w budownictwie stanowić będzie zmienną objaśnianą (zależną) – Y, a potencjalne zmienne objaśniające (niezależne), mogące charakteryzować czynniki wpływające na wartość dodaną w budownictwie zostaną oznaczone jako X_n. Wprowadza się zatem następujące oznaczenia dla wytypowanych zmiennych:

- Y: Wartość dodana brutto ogółem (ceny bieżące) - budownictwo (mln zł)
- X1: Mieszkania w budowie ogółem (w tys.)
- X2: Mieszkania oddane do użytkowania ogółem (w tys.)
- X3: Nakłady inwestycyjne w przemyśle (od początku roku do końca okresu) (w mln zł)
- X4: Podaż pieniądza ogółem M3 (stan w końcu okresu) (mln zł)
- X5: Popyt krajowy (ceny bieżące) (w mln zł)
- X6: Przeciętne miesięczne wynagrodzenie nominalne brutto w budownictwie (w zł)
- X7: Przeciętne zatrudnienie w budownictwie (w tys.)
- X8: Wskaźnik ogólnego klimatu koniunktury w budownictwie (w miesiącu kończącym okres) (od początku roku do końca okresu (analogiczny okres roku poprzedniego=100), niewyrównany sezonowo)
- X9: Wskaźnik prognozy ogólnej sytuacji gospodarczej w budownictwie (w miesiącu kończącym okres) (od początku roku do końca okresu (analogiczny okres roku poprzedniego=100), niewyrównany sezonowo)
- X10: Wskaźnik prognozy ogólnej sytuacji gospodarczej w przetwórstwie przemysłowym (w miesiącu kończącym okres) (od początku roku do końca okresu (analogiczny okres roku poprzedniego=100), niewyrównany sezonowo).

Dane charakteryzujące powyższe składowe zostały zaczerpnięte z zasobów Głównego Urzędu Statystycznego (link do źródła internetowego GUS, z którego pobrano dane: http://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultstronaopisowa/1772/1/5/kwartal

ne_wskazniki_makroekonomiczne_cz_ii_.xls – dostęp 24 lutego 2016). Dane obejmują 11 lat, czyli 44 kwartały (od I kw. 2005 do IV kw. 2015 roku), jednakże dla zmiennej X6 i X7 dane zostały ujęte w statystykach GUS dopiero od I kw. 2007, dla zmiennej X3 nie pojawiają się dane za wybrane okresy pierwszych kwartałów (oznaczone jako niemożliwe lub niecelowe wypełnienie), oraz dla niektórych zmiennych (Y, X3-X7 dane nie są jeszcze dostępne za IV kw. 2015 r. Dane dotyczące produkcji sprzedanej przemysłu ogółem i przeciętnego miesięcznego wynagrodzenia nominalnego brutto i zatrudnienia w budownictwie odnoszą się do podmiotów gospodarczych o liczbie pracujących powyżej 9 osób, czyli nie obejmują mikroprzedsiębiorstw.

Budowa modelu, którego wyniki zostaną poniżej zaprezentowane, ***przeprowadzona*** jest ***przy wykorzystaniu programu Gretl 2016a-git***.

Wstępnie dokonano wyboru 9 potencjalnych zmiennych objaśniających i należy zbadać ich istotność, tzn. czy zmienne te mają istotny wpływ na zmienną objaśnianą (wartość dodaną brutto w budownictwie. Dodatkowo sprawdzamy istotność wyrazu wolnego, gdyby okazał się statystycznie nieistotny, powinniśmy oszacować model bez wyrazu wolnego (usuwając zmienną *const*). Obok założeń dotyczących zmiennych objaśnianej i objaśniających, należy również sprawdzić założenia dotyczące reszt estymowany modelu, które stanowią realizację składnika losowego.

W toku budowy modelu możemy traktować szereg czasowy całościowo i próbować od razu modelować wszystkie jego składowe; inne podejście przewiduje najpierw analizę trendu, ewentualnych interwencji, potem wahań okresowych i na końcu reszt modelu całościowego, które są realizacją składnika losowego. W celach porównawczych **w pracy** podjęto się próby **uwzględnienia budowy modelu w dwóch wariantach - dla zmiennych niestacjonarnych i stacjonarnych**.

* * * * * * * * * * * * * *

Przechodzi się zatem od razu do wstępnej analizy modelu **dla zmiennych niestacjonarnych**.

Wyniki estymacji parametrów dla takiego wstępnego **modelu**, wykonanego **klasyczną metodą najmniejszych kwadratów (KMNK)**, zaprezentowano poniżej w Tabeli 1.

Tabela 1. Pierwszy wstępny model z estymacja parametrów metodą KMNK (ujęcie kwartalne od 2005 do 2015 roku włącznie; obserwacje z brakującymi danymi (13) zostały pominięte, zmienna zależna (Y): Y)

Model 1: Estymacja KMNK, wykorzystane obserwacje 2005:1-2016:4 (N = 31)
Liczba pominiętych niekompletnych obserwacji: 13
Zmienna zależna (Y): Y

	Współczynnik	*Błąd stand.*	*t-Studenta*	*wartość p*	
const	−51476,2	19756,4	−2,6055	0,0169	**
X1	88,313	63,3417	1,3942	0,1785	
X2	−463,345	135,932	−3,4086	0,0028	***
X3	0,377732	0,0771152	4,8983	<0,0001	***
X4	−0,0383233	0,0226113	−1,6949	0,1056	
X5	0,0523369	0,0492719	1,0622	0,3008	
X6	10,8179	5,2512	2,0601	0,0526	*
X7	−21,9416	62,8841	−0,3489	0,7308	

X8	−378,867	250,903	−1,5100	0,1467
X9	186,574	165,467	1,1276	0,2729
X10	49,039	111,63	0,4393	0,6652

Średn.aryt.zm.zależnej	27610,47	Odch.stand.zm.zależnej	6518,407
Suma kwadratów reszt	85368303	Błąd standardowy reszt	2066,014
Wsp. Determ. R-kwadrat	0,933028	Skorygowany R-kwadrat	0,899542
F(10, 20)	27,86328	Wartość p dla testu F	0,00000000141
Logarytm wiarygodności	−273,8288	Kryt. inform. Akaike'a	569,6576
Kryt. bayes. Schwarza	585,4315	Kryt. Hannana-Quinna	574,7995

Wyłączając stałą, największa wartość p jest dla zmiennej 8 (X7)

Współczynniki determinacji $R^2 = 0{,}933$ i skorygowany $R^2 = 0{,}89954$. Ponieważ liczba szacowanych zmiennych w modelu jest mniejsza od liczby obserwacji, zatem wzięto pod uwagę współczynnik determinacji $R^2 = 0{,}933$, który informuje, że 93,3% zmienności zostaje wyjaśnionych przez powyższy model.

Istotność zmiennych została sprawdzona testem t-Studenta. Na podstawie wyników należy stwierdzić, że największa wartość p jest dla zmiennej X7, co pozwala przypuszczać, że powinna zostać wykluczona z modelu. Zmienną X10 także charakteryzuje wysoka wartość p, co również sugeruje, że powinno się ją wykluczyć z modelu. Za statystycznie istotne zostały wskazane zarówno wyraz wolny jak i zmienne objaśniające X2, X3 i X6. Wyraz wolny wskazany został za **istotny statystycznie** na poziomie istotności poniżej 2% ($p < \alpha = 0{,}02$), zmienne objaśniające X2 i X3 na poziomie istotności zdecydowanie poniżej 0,01. Zmienna objaśniająca X6 może wstępnie w rozważaniach również zostać uznana za statystycznie istotną; dla niej wartość p znajduje się nieco powyżej poziomu istotności równego 0,05 (zdecydowanie powyżej poziomu istotności 10%). Jednakże warto zauważyć, że zmienna X4 charakteryzuje się również stosunkowo niską wartością p (nieco powyżej 0,01), a w dalszej kolejności zmienne X8 i X1, co może być ewentualnie wzięte pod uwagę przy kolejnych etapach rozważań w budowie modelu.

W Tabeli 1. podane są również **wyniki dla testu F**, który bada łączną istotność wszystkich parametrów. Testowane hipotezy to: H_0: wszystkie zmienne (parametry) są nieistotne, H_1: przynajmniej jedna zmienna (parametr) jest istotna. Wartość p dla testu F jest bardzo mała, mniejsza od poziomu istotności, który można przyjąć na poziomie 1% ($\alpha = 0{,}01$), co upoważnia do odrzucenia hipotezy zerowej o nieistotności wszystkich zmiennych i przyjęciu hipotezy alternatywnej, że co najmniej jedna zmienna w modelu jest istotna.

Postać **analityczną** takiego **wstępnego modelu** można zapisać jako:

$$Y = -51476{,}2 + 88{,}313\, x_1 - 463{,}345\, x_2 + 0{,}377732\, x_3 - 0{,}0383233\, x_4 + 0{,}0523369\, x_5 + 10{,}8179\, x_6 - 21{,}9416\, x_7 - 378{,}867\, x_8 + 186{,}574\, x_9 + 49{,}039\, x_{10} + \varepsilon$$

W celu budowy ostatecznego modelu ekonometrycznego zostanie przeprowadzony dalszy proces doboru zmiennych objaśniających. Należy dokonać doboru zmiennych do budowanego modelu, które mogłyby najlepiej opisywać wyniki zmiennej objaśnianej. Wstępnie pomocne mogą się okazać dane charakteryzujące statystyki będące realizacjami zmiennej objaśnianej i zmiennych objaśniających, na podstawie których można podjąć decyzję co do ewentualnej możliwości redukcji niektórych zmiennych. Podstawowe

charakterystyki opisowe zmiennej objaśnianej i wytypowanych 10 zmiennych objaśniających przedstawia poniższa Tabela 2.

Tabela 2. Statystyki opisowe, dla obserwacji z próby (ujęcie kwartalne od 2005 do 2015 roku włącznie; obserwacje z brakującymi danymi zostały pominięte)

Statystyki opisowe, dla obserwacji z próby 2005:1 - 2015:4
(obserwacje z brakującymi danymi będą pominięte)

Zmienna	Średnia	Mediana	Minimalna	Maksymalna
Y	24189,3	25867,9	7513,40	37300,3
X1	682,318	693,800	592,400	736,600
X2	35,0932	34,1000	22,1000	57,7000
X3	34752,8	33353,9	9077,00	74086,9
X4	741768,	742764,	391190,	1107490
X5	355747,	366478,	230319,	473141,
X6	3584,03	3621,67	2649,97	4111,21
X7	436,945	437,500	379,800	504,100
X8	-1,90227	-5,30000	-34,9000	31,6000
X9	3,63182	2,70000	-42,8000	38,1000
X10	7,27226	8,95000	-19,3000	23,7000
Zmienna	**Odch.stand.**	**Wsp. zmienności**	**Skośność**	**Kurtoza**
Y	8367,34	0,345911	-0,224126	-1,08930
X1	39,7971	0,0583263	-0,826018	-0,375491
X2	7,32639	0,208770	0,670917	0,627567
X3	17582,2	0,505923	0,390210	-0,833875
X4	216101,	0,291332	-0,0484183	-1,21164
X5	64532,7	0,181401	-0,258630	-0,830170
X6	324,843	0,0906363	-0,825823	0,745240
X7	33,7871	0,0773256	0,190322	-0,825419
X8	17,6868	9,29770	0,293611	-0,865955
X9	21,0674	5,80077	-0,151789	-0,784959
X10	10,9559	1,50653	-0,697163	-0,131179
Zmienna	**Percentyl 5%**	**Percentyl 95%**	**Zakres Q3-Q1**	**Brakujące obs.**
Y	8603,28	36424,5	14236,0	1
X1	601,650	735,325	46,9000	0
X2	23,6250	47,7500	8,60000	0
X3	10270,1	63990,2	27586,3	7
X4	406721,	1076390	394678,	1
X5	240933,	456888,	89350,9	1
X6	2830,57	4089,83	408,270	9
X7	382,280	495,860	50,4000	9
X8	-29,2750	29,3500	30,9250	0
X9	-31,7250	36,2000	30,9250	0
X10	-17,4000	22,4500	15,6500	0

Ponieważ literatura przedmiotu wskazuje, że zmienne modelu powinny charakteryzować się zmiennością powyżej 10%, należałoby wyeliminować zmienne objaśniające o zbyt niskim poziomie zmienności. Na podstawie powyższej tabeli, można uznać, że zmienne X1 (0,0583263), X6 (0,0906363) i X7 (0,0773256) nie spełniają tego

wymagania, wskazane byłoby zatem wyeliminowanie ich z modelu, aczkolwiek zauważyć, że zmienność X6 jest bliska 10%.

W wielu publikacjach naukowych można znaleźć szeroki zestaw metod doboru zmiennych do modelu ekonometrycznego. Większość metod doboru opiera się na dwóch podstawowych założeniach, tj. aby zmienne niezależne wybrane do modelu były nieskorelowane lub słabo skorelowane między sobą, oraz aby wybrane zmienne niezależne były wysoko skorelowane ze zmienną zależną.

W związku z tym kolejnym krokiem jest **zbadanie korelacji między zmiennymi**, przeprowadzając test istotności dla współczynnika korelacji liniowej Pearsona. Przy założeniu, że hipoteza zerowa ma postać H_0: $R_{ij} = 0$ (przy H_1: $\sim H_0$), wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) wyznaczona ze statystyki r* wynosi 0,3388 (dla n = 34). Wyniki współczynników korelacji dla zmiennych z próby przedstawia poniższa Tabela 2.

Tabela 3. Współczynniki korelacji liniowej dla obserwacji z próby (ujęcie kwartalne od 2005 do 2015 roku włącznie; obserwacje z brakującymi danymi zostały pominięte)

Współczynniki korelacji liniowej dla obserwacji z próby 2007:2-2015:3
(obserwacje z brakującymi danymi będą pominięte)
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,3388 dla n = 34

Y	X1	X2	X3	X4	
1,0000	0,4836	0,3731	0,8328	0,3698	Y
	1,0000	-0,1073	0,0147	0,5683	X1
		1,0000	0,7522	-0,0704	X2
			1,0000	0,0620	X3
				1,0000	X4

X5	X6	X7	X8	X9	
0,7169	0,6376	0,3772	-0,3757	-0,5998	Y
0,5854	0,6425	0,4068	-0,2649	-0,2738	X1
0,2603	0,1301	0,1892	-0,3595	-0,5655	X2
0,4898	0,3447	0,1205	-0,2861	-0,5799	X3
0,8443	0,9147	-0,0081	-0,6150	-0,5170	X4
1,0000	0,9170	0,2898	-0,6774	-0,7330	X5
	1,0000	0,1140	-0,6064	-0,6130	X6
		1,0000	-0,5269	-0,5518	X7
			1,0000	0,9316	X8
				1,0000	X9

X10	
-0,5452	Y
-0,1772	X1
-0,7339	X2
-0,7136	X3
-0,2286	X4
-0,4912	X5
-0,3980	X6
-0,4697	X7
0,7222	X8
0,8528	X9
1,0000	X10

Jak wspomniano na początku pracy, charakteryzując dane, dla niektórych lat dane nie są dostępne. Ma to znaczenie dla obliczeń i wyników macierzy korelacyjnej. W powyższej Tabeli 2. dostępne są współczynniki korelacji liniowej dla obserwacji z próby począwszy od II kwartału 2007 do III kwartału 2015 roku.

Na podstawie powyższych wyników, należy stwierdzić, że współczynniki korelacji $|r_{ij}| > r^*$ można uznać, za istotnie różniące się od zera. Ponadto, z macierzy korelacji wyników dostępnych za określone wyżej kwartały wynika, że większość zmiennych objaśniających jest dość istotnie związana ze zmienną objaśnianą :

- zmienna X4 jest **najmniej skorelowana** (ze wszystkich zmiennych) **ze zmienną objaśnianą Y** (współczynnik korelacji X4-Y wynosi 0,3698), dlatego wskazane jest wykluczenie jej z modelu;
- w dalszej kolejności zmienne X2, X7 i X8 są najmniej skorelowane ze zmienną objaśnianą Y, na nieco wyższym poziomie jak zmienna X4 (współczynnik korelacji X2-Y wynosi 0,3731, X7-Y wynosi 0,3772, X8-Y wynosi -0,3757), dlatego można rozważyć wykluczenie ich z modelu – w przypadku zmiennej X7 warto przypomnieć, że jej współczynnik zmienności wynosi mniej niż 8%, co może stanowić dodatkowy argument za odrzuceniem tej zmiennej w dalszych rozważaniach nad równaniem modelu,
- zmienna X3 jest **najbardziej skorelowana** (ze wszystkich zmiennych) **ze zmienną objaśnianą Y** (współczynnik korelacji X3-Y wynosi 0,8328), dlatego wskazane jest pozostawienie jej w modelu; ponadto zmienna ta została uznana za statystycznie istotną na podstawie analizy wyników estymacji parametrów wstępnego modelu wykonanego metodą KMNK,
- w dalszej kolejności zmienna X5 (0,7169) prezentuje stosunkowo wysoki stopień korelacji ze zmienna Y, a następnie zmienna X6 (0,6376) i X9 (-0,5998); aczkolwiek zmienne X5 i X9 charakteryzowały się gorszymi wynikami statystycznej istotności w metodzie KMNK niż większość zmiennych we wstępnym modelu, dlatego rozważa się eliminację ich z ostatecznego modelu,
- X2 została wskazana jako statystycznie istotna w KMNK, jednakże jest w miarę silnie powiązana ze zmienną X3 (0,7522) i X10 (-0,7339), należy zatem ostrożnie podejść do rozważań nad jej obecnością w ostatecznym modelu – prawdopodobnie wykluczając w takim podejściu zmienną X3 i/lub X10,
- X4 jest silnie skorelowane z X6 (0,9147) i relatywnie wysoko z X5 (0,8443); nie została wskazana jako jednak ze zmiennych znaczących statystycznie w KMNK, jednakże charakteryzuje się stosunkowo niską wartością p (nieco powyżej 0,01), można zatem rozważyć jej pozostawienie w ostatecznym modelu; warto jednak zwrócić uwagę na wyniki p dla zmiennej X6, które charakteryzują ją jako jedną ze statystycznie najistotniejszych w modelu, dlatego warto rozważyć jej pozostawienie kosztem X4; w przeciwieństwie do X5,
- X5 jest silnie skorelowane z X6 (0,9170) i na zdecydowanie niższym poziomie z X9 (-0,7330),
- X8 jest silnie skorelowane z X9 (0,9316) i na niższym poziomie z X10 (-0,7222),
- X9 jest w miarę silnie skorelowane z X10 (0,8528); ponadto obie zmienne nie są wskazane jako statystycznie istotne na podstawie wyników KMNK, w szczególności zmienna X10 (aczkolwiek jest istotnie skorelowana ze zmienną objaśnianą).

Na podstawie powyższej analizy macierzy korelacji, wspartej wnioskami z wyników estymacji parametrów dla wstępnego modelu, wykonanego KMNK, jak i znajomości tematyki budownictwa podjęto **decyzję o pozostawieniu** w dalszych rozważaniach nad modelem **zmiennych: X2, X3 i X4**.

Ponieważ jednak wystąpiły wątpliwości co do możliwości uwzględnienia zmiennych X2 i X3 jednocześnie w modelu, wypada jednak dokonać weryfikacji wyboru zmiennych, wszystkich zmiennych. Wyznaczona macierz korelacji umożliwia w dalszej kolejności zastosowanie jednej z wielu metod redukcji potencjalnych zmiennych objaśniających. Jedną z nich jest **metoda pojemności integralnej Hellwiga**.

Metoda Hellwiga (w Gretlu wykorzystano obliczenia w oparciu o skrypt z zajęć Analiza współzależności zjawisk) przeprowadzona dla 10 zmiennych wskazała za najbardziej istotne zmienne **X3, X4, X7** (przy **pojemności integralnej na poziomie 0,96087135**).

Wziąwszy pod uwagę wyniki uzyskane metodą Hellwiga, jak i wnioski płynące z powyżej przeprowadzonej analizy i specyfiki zagadnienia ekonomicznego, uzasadnionym wydaje się **uwzględnienie w dalszych etapach** budowy modelu **jedynie dwóch zmiennych – X3 i X4**. X2 ostatecznie nie zostanie uwzględniona, ponieważ pomimo wysokiej istotności statystycznej, jest jednocześnie stosunkowo wysoko skorelowana z X3, a X3 ma wyższy poziom istotności statystycznej. X3 jest najbardziej statystycznie istotna i najbardziej skorelowana z Y, dlatego pożądane jest pozostawienie jej w modelu. X4 jest o wiele mniej statystycznie istotna (jednak poziom p nieco ponad 0,01), oraz równocześnie słabo skorelowana z X3, dlatego również zostaje uwzględniona w dalszym modelu. Zmienna X7 zdecydowanie nie zostanie ujęta w dalszych rozważaniach, ze względu na to, że jej wartość p w metodzie KMNK jest najwyższa, ponadto jest słabo skorelowana z Y i charakteryzuje się niskim współczynnikiem zmienności.

Wyestymowany zostanie zatem **drugi model** uwzględniający oprócz zmiennej objaśnianej już dwie zmienne objaśniające: X3 i X4. Również i w tym przypadku należy zbadać ich istotność oraz sprawdzić istotność wyrazu wolnego. **Wyniki estymacji parametrów** dla **modelu uwzględniającego wybrane zmienne**, wykonanego **klasyczną metodą najmniejszych kwadratów (KMNK)**, zaprezentowano poniżej w Tabeli 4.

Tabela 4. Drugi model z estymacja parametrów metodą KMNK (ujęcie kwartalne od 2005 do 2015 roku włącznie; obserwacje z brakującymi danymi (13) zostały pominięte, zmienna zależna (Y): Y)

Model 2: Estymacja KMNK, wykorzystane obserwacje 2005:1-2015:4 (N = 37)
Liczba pominiętych niekompletnych obserwacji: 7
Zmienna zależna (Y): Y

	Współczynnik	*Błąd stand.*	*t-Studenta*	*wartość p*	
const	6343,62	2202,74	2,8799	0,0068	***
X3	0,310203	0,0322397	9,6218	<0,0001	***
X4	0,0116532	0,00266629	4,3706	0,0001	***

Średn.aryt.zm.zależnej	26122,89	Odch.stand.zm.zależnej	7202,241
Suma kwadratów reszt	3,77e+08	Błąd standardowy reszt	3329,394
Wsp. determ. R-kwadrat	0,798177	Skorygowany R-kwadrat	0,786305
F(2, 34)	67,23204	Wartość p dla testu F	0,00000000 000153
Logarytm wiarygodności	−351,0266	Kryt. inform. Akaike'a	708,0532
Kryt. bayes. Schwarza	712,8860	Kryt. Hannana-Quinna	709,7570

Dla nowego modelu współczynniki determinacji R^2 = 0,798177, skorygowany R^2 = 0,786305, czyli różnica między współczynnikami wynosi 0,011872, można zatem uznać, że model nie jest przeparametryzowany i wyjaśnia prawie 80% zmienności. Współczynnik ten jest na niższym poziomie niż w pierwszym modelu, gdyż konieczne okazało się wyeliminowanie z modelu zmiennych o wysokim stopniu korelacji ze zmienną objaśnianą lub o wysokim poziomie istotności w modelu. Mimo to w drugim oszacowanym modelu według testu F oceniającego łączną istotność modelu, stopień wyjaśniania zmienności zmiennej zależnej jest istotny. Z kolei na podstawie testu t-Studenta uznać należy, że istotność zmiennych wraz z wyrazem wolnym w modelu jest na bardzo wysokim poziomie ($p < \alpha$ = 0,01). Zmniejszyły się niestety kryteria informacyjne w porównaniu do pierwszego modelu.

Postać **analityczną** takiego **drugiego modelu** można zapisać jako:

$$Y = 6343{,}62 + 0{,}310203\, x_3 + 0{,}0116532\, x_4 + \varepsilon$$

Nie można jeszcze jednak uznać tego modelu za ostateczny, gdyż w kolejnych etapach jego budowania należy sprawdzić pewne własności (szacowanego modelu) dotyczące reszt.

Jednym z kolejnych kroków budowy modelu jest ocena **normalności reszt składnika losowego**. W tym wypadku w programie Gretl testowanie normalności rozkładu reszt wykonywane jest przy wykorzystaniu testu Doornika-Hansena. Test weryfikuje hipotezę zerową o istnieniu normalnego rozkładu reszt, wobec hipotezy alternatywnej, że składnik modelu nie ma rozkładu normalnego.

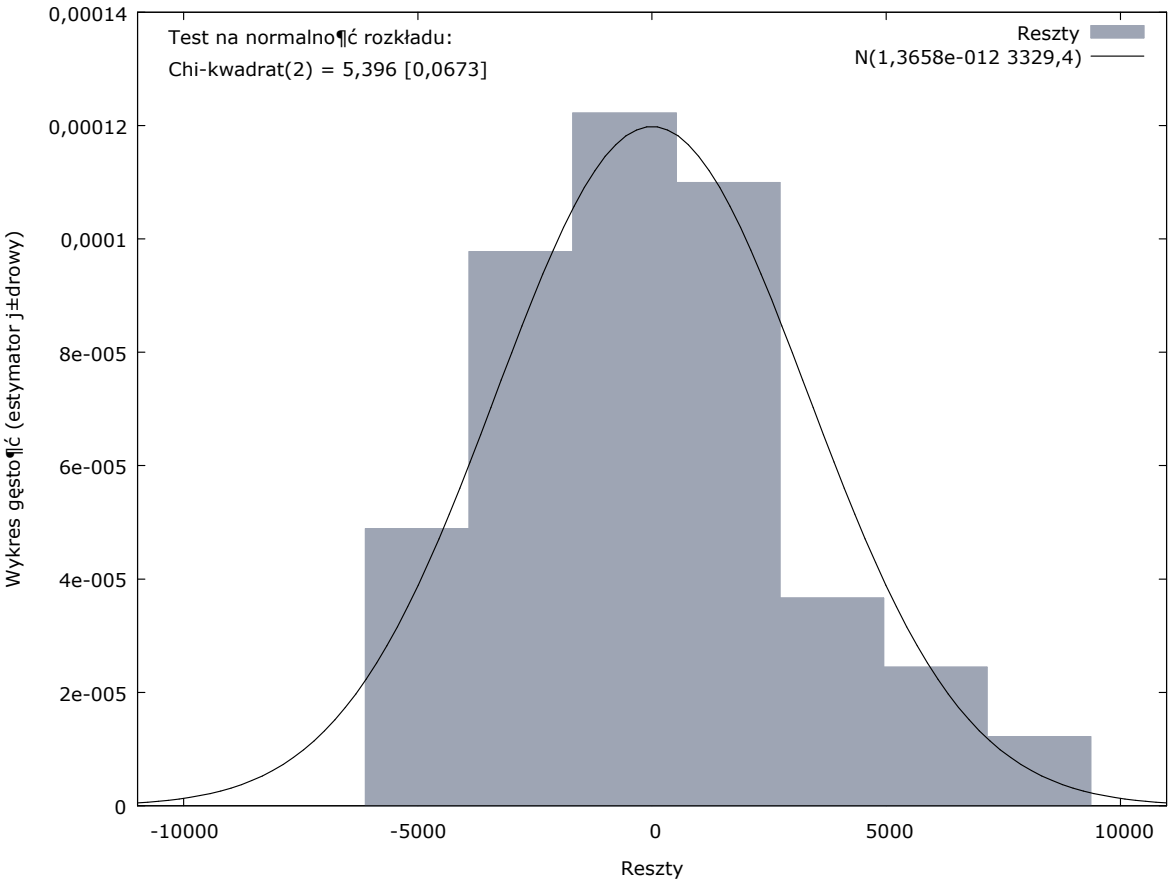
W rozważanym drugim modelu, jedynie z dwoma zmiennymi objaśniającymi X3 i X4, wyniki testu wskazują, że nie ma podstaw do odrzucenia hipotezy zerowej, ponieważ p = 0,06735 i jest większe niż przyjęty poziom istotności α = 0,05 (statystyka testowa wynosi 5,396 i jest mniejsza niż $\chi^2_{0,05,\,2}$ = 5,991. Można stwierdzić, że z prawdopodobieństwem 95% rozkład składników losowych reprezentuje rozkład normalny. Wyniki testu przedstawione są w poniższej Tabeli 5. jak i w formie histogramu na Rysunku 1.

Tabela 5. Wyniki testu Doornika-Hansena na normalność rozkładu reszt (obserwacje 1-44, liczba przedziałów = 7, średnia = 0,00000000000136578, odch.std. = 3329,39).

Przedziały	średnia	liczba	częstość	skumlowana	
< -3905,8	-5011,8	4	10,81%	10,81%	***
-3905,8 - -1693,6	-2799,7	8	21,62%	32,43%	*******
-1693,6 - 518,56	-587,52	10	27,03%	59,46%	*********
518,56 - 2730,7	1624,6	9	24,32%	83,78%	********
2730,7 - 4942,9	3836,8	3	8,11%	91,89%	**
4942,9 - 7155,0	6049,0	2	5,41%	97,30%	*
>= 7155,0	8261,1	1	2,70%	100,00%	

Brakujące obserwacje = 7 (15,91%)

Hipoteza zerowa: dystrybuanta empiryczna posiada rozkład normalny.Test Doornika-Hansena (1994) - transformowana skośność i kurtoza.: Chi-kwadrat(2) = 5,396 z wartością p 0,06735

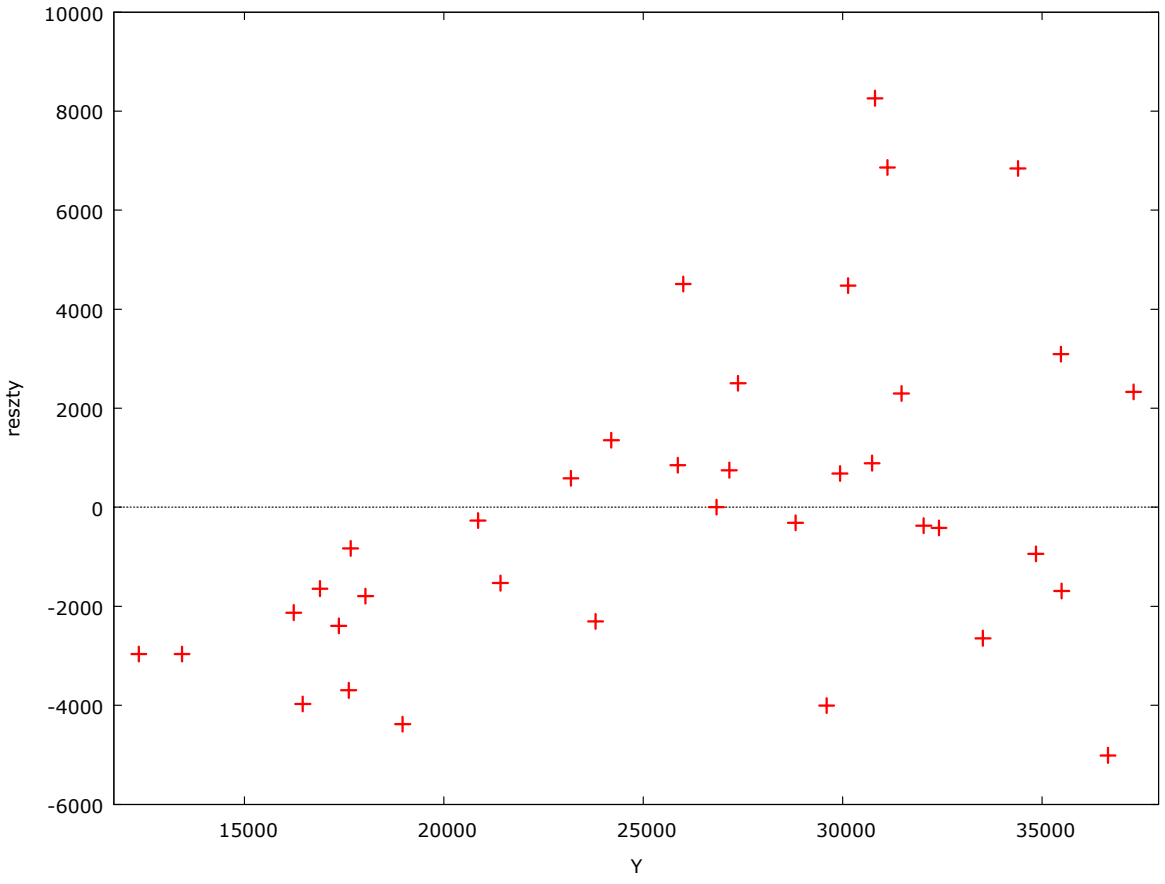


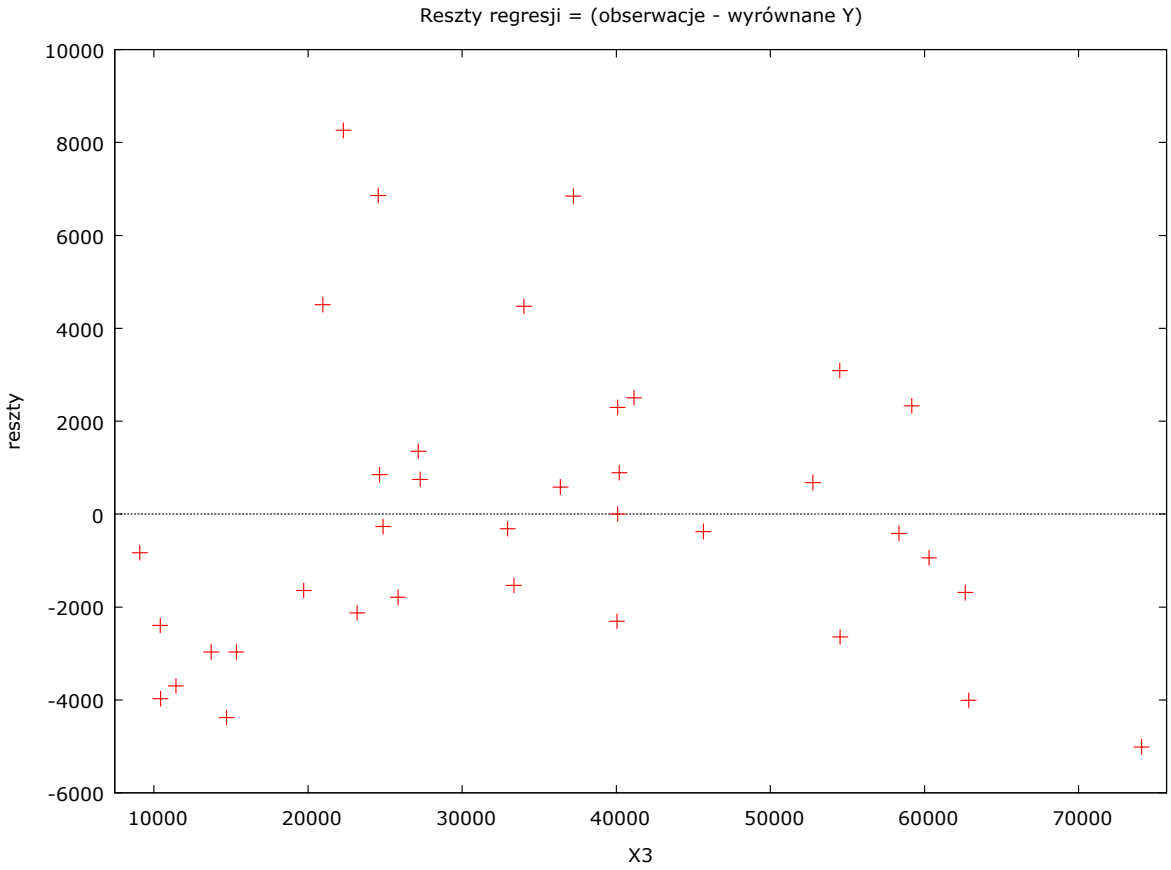
Rysunek 1. Wyniki testu Doornika-Hansena na normalność rozkładu reszt (obserwacje 1-44, liczba przedziałów = 7, średnia = 1,36578e-012, odch.std. = 3329,39).

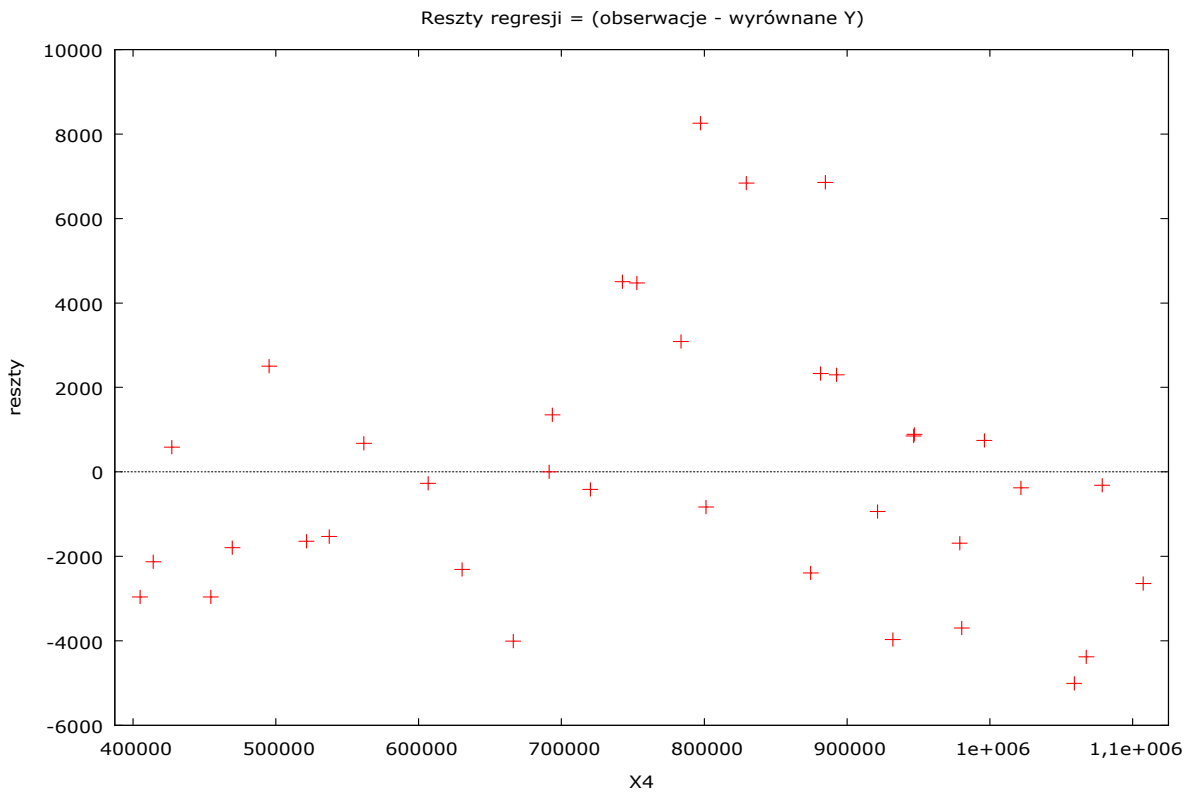
Wyniki wskazują, że rozkład reszt jest normalny, ale i również, że metoda najmniejszych kwadratów (KMNK) zapewnia, że średnia reszt jest równa zero.

Ominięty zostanie kolejny element weryfikacji statystycznej budowanego modelu ekonometrycznego polegający na sprawdzeniu **autokorelacji** (i jej usuwaniu) w programie Gretl, ponieważ nie jest ona dostępna do wykonania przez program dla takiego układu danych i wyników.

Można jednakże próbować ocenić jakość reszt na podstawie wykresu (rozkład losowości i autokorelację). Losowość reszt można ocenić na podstawie ich wykresu dla kolejnych obserwacji – wykresu reszt pochodzących z przedstawianego modelu. Wydaje się, iż reszty te spełniają założenia regresji. Rozrzut reszt wokół zera jest niezależny od wartości zmiennych X, czy Y; reszty nie rosną też wraz ze wzrostem X lub Y, czyli ten warunek o nieskorelowaniu reszt można uznać za spełniony.


10000
8000
6000
4000
2000
0
-2000
-4000
-6000
reszty
15000
20000
25000
30000
35000
Y


Reszty regresji = (obserwacje - wyrównane Y)
10000
8000
6000
4000
2000
0
-2000
-4000
-6000
reszty
10000
20000
30000
40000
50000
60000
70000
X3



W dalszej kolejności oceny składnika losowego należy przeprowadzić **badanie heteroskedastyczności**. W celu sprawdzenia heterodskedastyczności wykonano test White'a oraz test Breuscha-Pagana. Hipotezą zerową obu tych testów jest brak heteroskedastyczności. Wyniki testów przedstawiono w poniższych tabelach.

Tabela 6. Wyniki testu White'a na heteroskedastyczność reszt (zmienność wariancji resztowej)

Test White'a na heteroskedastyczność reszt (zmienność wariancji resztowej)
Estymacja KMNK, wykorzystane obserwacje 2005:1-2015:4 (N = 37)
Liczba pominiętych niekompletnych obserwacji: 7
Zmienna zależna (Y): uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	−3,37386e+07	3,53349e+07	−0,9548	0,3471
X3	−817,350	878,886	−0,9300	0,3596
X4	160,730	99,1734	1,621	0,1152
sq_X3	0,00240499	0,00900771	0,2670	0,7912
X2_X3	0,000574402	0,000896665	0,6406	0,5265
sq_X4	−0,000111289	7,23700e-05	−1,538	0,1342

Wsp. determ. R-kwadrat = 0,118291

Statystyka testu: TR^2 = 4,376765,
z wartością p = P(Chi-kwadrat(5) > 4,376765) = 0,496539

Tabela 7. Wyniki testu Breuscha-Pagana na heteroskedastyczność

Test Breuscha-Pagana na heteroskedastyczność
Estymacja KMNK, wykorzystane obserwacje 2005:1-2015:4 (N = 37)
Liczba pominiętych niekompletnych obserwacji: 7
Zmienna zależna (Y): standaryzowane uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	0,451242	0,986796	0,4573	0,6504
X3	−1,33499e-05	1,44429e-05	−0,9243	0,3618
X4	1,31142e-06	1,19446e-06	1,098	0,2800

Wyjaśniona suma kwadr. = 3,81947

Statystyka testu: LM = 1,909737,
z wartością p = P(Chi-kwadrat(2) > 1,909737) = 0,384863

W obu testach wartość p jest większa od poziomu istotności 0,05, nie ma więc podstaw do odrzucenia hipotezy zerowej. Zatem w estymowanym modelu nie występuje heteroskedastyczność, czyli składniki losowe modelu charakteryzuje jednorodność.

Kolejnym etapem weryfikacji modelu jest **testowanie współliniowości zmiennych modelu**. Z przeprowadzonego w tym celu **testu VIF** wynika, że w modelu nie występuje współliniowość zmiennych, ponieważ wartość czynnika jest mniejsza od 10, a nawet znajduje się w pobliżu minimalnej wartości czynnika.

Tabela 8. Wyniki testu VIF

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

X3	1,044
X4	1,044

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

Przeprowadzono również **test Ramsey'a RESET**. Hipotezą zerową tego testu jest liniowość modelu. Wszystkie wartości p są mniejsze od poziomu istotności 0,05, zatem hipotezę o liniowości modelu należy odrzucić.

Tabela 9. Wyniki testu Ramsey'a RESET

Test RESET na specyfikację (kwadrat i sześcian zmiennej)
Statystyka testu: F = 6,257798,
z wartością p = P(F(2,32) > 6,2578) = 0,00508

Test RESET na specyfikację (tylko kwadrat zmiennej)

Statystyka testu: F = 12,693451,
z wartością p = P(F(1,33) > 12,6935) = 0,00114

Test RESET na specyfikację (tylko sześcian zmiennej)
Statystyka testu: F = 12,200963,
z wartością p = P(F(1,33) > 12,201) = 0,00138

Istotnym elementem oceny estymowanego modelu jest **zbadanie efektu katalizy**. Może się bowiem okazać, że wysokie wartości współczynnika determinacji R^2 i rzeczywistej siły powiązań zmiennych objaśniających w prezentowanym powyżej modelu nie uzasadniają takiego wyniku. Zmienna X_i z pary zmiennych $\{X_i, X_j\}$, $i < j$, jest katalizatorem, jeśli $r_{ij} < 0$ lub $r_{ij} > r_i\ r_j$. Wyniki z obliczeń mających na celu zbadanie efektu katalizy zostały przestawione poniżej.

Tabela 9. Wyniki badania efektu katalizy

Współczynniki korelacji liniowej dla obserwacji z próby 2005:2-2015:3
(obserwacje z brakującymi danymi będą pominięte)
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,3044 dla n = 42

Y	X3	X4	
1,0000	0,8275	0,5694	Y
	1,0000	0,2042	X3
		1,0000	X4

R0	
X3	0,8275
X4	0,5694

R Korelacja x3 z x4	
1	0,2042
0,2042	1

Korelacja X3 z Y/korelacja X4 z Y	
r3/4	r34
1,453284159	0,2042

Z macierzy korelacji wynika, że korelacja zmiennej x3 z x4 nie jest mniejsza od 0. Korelacja X3 z Y podzielona przez korelację X4 z Y wynosi: 1,45328, nie jest więc mniejsza niż korelacja X3 z X4 (0,5694). Z tego wynika, że zmienna X3 nie jest katalizatorem, nie zaburza więc wyniku.

Ostatnim elementem oceny estymowanego modelu będzie **zbadanie 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. Zestawienie współczynników dla szacowanego modelu zamieszczono poniżej.

Tabela 9. Zestawienie współczynników dla szacowanego modelu

	Współczynnik	Korelacja
X3	0,310203	0,8275
X4	0,011653	0,5694

Widać zatem, że w modelu zachodzi zjawisko koincydencji, ponieważ spełniony jest warunek, że dla każdej zmiennej znak współczynnika jest równy znakowi współczynnika korelacji.

Można zatem przejść do sformułowania **ostatecznej postaci modelu i jego interpretacji**. W oparciu o wcześniejsze obliczenia wiadomo, że dla nowego modelu współczynniki determinacji R^2 i skorygowanego R^2 wynoszą odpowiednio 0,798177 i 0,786305, czyli różnica między współczynnikami jest niska i wynosi 0,011872. Można zatem uznać, że model wyjaśnia prawie 80% zmienności i nie jest przeparametryzowany. Współczynnik ten jest na poziomie niższym niż w pierwszym modelu, jednakże wyniki testu F oceniającego łączną istotność modelu, stopień wyjaśniania zmienności zmiennej zależnej jest wysoce istotny. Zmniejszyły się niestety kryteria informacyjne w porównaniu do pierwszego modelu. Dodatkowo, jeżeli w oparciu o powyżej przedstawione argumenty, uzna się, że występuje normalność rozkładu reszt i zakłada się, że nie ma autokorelacji, to można wnioskować, że uzyskany estymator KMNK jest nieobciążony i efektywny, średnie błędy szacunku doszacowane, oraz zarówno wartości statystyk t-Studenta jak i współczynnik determinacji nie są przeszacowane.

Ostateczną postać **analityczną modelu** można zapisać jako:

$$Y = 6343{,}62 + 0{,}310203\, x_3 + 0{,}0116532\, x_4$$

Mając na uwadze, że:

- Y: Wartość dodana brutto ogółem (ceny bieżące) - budownictwo (mln zł) – zmienna zależna (objaśniana)
- X3: Nakłady inwestycyjne w przemyśle (od początku roku do końca okresu) (w mln zł) – zmienna niezależna (objaśniająca)
- X4: Podaż pieniądza ogółem M3 (stan w końcu okresu) (mln zł) – zmienna niezależna (objaśniająca);

powyższy model można zinterpretować, następująco:

- wzrost nakładów inwestycyjnych w przemyśle (X3) o 1 zł powoduje wzrost wartości dodanej brutto w budownictwie (Y) o 0,310203 zł, **ceteris paribus**;
- wzrost podaży pieniądza M3 (X4) o 1 zł, powoduje wzrost wartości dodanej brutto w budownictwie (Y) o 0,0116532 zł , **ceteris paribus**

* * * * * * * * * * * * * *

Powyżej przedstawiony model został wyestymowany w oparciu o dane niestacjonarne. Jednakże większość danych odnoszących się do procesów makroekonomicznych prezentowanych jest przez zmienne w większości niestacjonarne. Konstruowane zatem na ich podstawie modele mogą być oparte na korelacjach pozornych. W takim wypadku, jeżeli nawet hipotezy zostaną postawione poprawnie (specyfikacja równań będzie w sensie ekonomicznym właściwa), to wnioskowanie statystyczne nie może być przeprowadzone w sposób konwencjonalny. Warto zatem przyjrzeć się **modelowi oszacowanemu dla danych stacjonarnych** oraz interpretacji wyników uzyskanych w takim modelu.

Jeśli zmienna jest niestacjonarna to należy wprowadzić do modelu pierwsze różnice dla zmiennych niestacjonarnych. W celu sprawdzenia stacjonarności zmiennych stosowany jest test ADF (Ho: proces jest niestacjonarny, H1: proces jest stacjonarny) jak i KPSS (H0: proces jest stacjonarny, H1: proces nie jest stacjonarny). Na podstawie przeprowadzonych testów stwierdzono, że wśród analizowanych zmiennych tylko zmienna X10 jest stacjonarna, pozostałe zmienne (wraz ze zmienną objaśnianą) zostały sprowadzone do stacjonarności poprzez kolejne dodawanie różnic. Dla tak przygotowanych zmiennych należy ponownie

przeprowadzić wszystkie testy, które nie będą obarczone ewentualnym błędem z powodu niestacjonarności.

Tabela 10. Pierwszy wstępny model z estymacja parametrów metodą KMNK dla danych stacjonarnych (ujęcie kwartalne od 2005 do 2015 roku włącznie)

Model 1: Estymacja KMNK, wykorzystane obserwacje 2005:1-2015:4 (N = 23)
Liczba pominiętych niekompletnych obserwacji: 21
Zmienna zależna (Y): d_d_Y

	Współczynnik	*Błąd stand.*	*t-Studenta*	*wartość p*	
const	30133,8	25741,8	1,1706	0,2645	
X1	−41,7015	36,1682	−1,1530	0,2714	
d_d_X2	−344,449	149,804	−2,2993	0,0402	**
d_d_X3	0,272935	0,0666779	4,0933	0,0015	***
d_d_X4	−0,0799799	0,0446558	−1,7910	0,0985	*
d_d_d_d_X5	0,026851	0,00507852	5,2872	0,0002	***
d_d_X6	−23,8756	7,00041	−3,4106	0,0052	***
d_d_X7	−50,7249	77,2884	−0,6563	0,5240	
d_d_X8	899,725	244,684	3,6771	0,0032	***
d_d_X9	−703,662	140,863	−4,9954	0,0003	***
X10	−78,2288	94,1355	−0,8310	0,4222	

Średn.aryt.zm.zależnej	449,5826	Odch.stand.zm.zależnej	18213,02
Suma kwadratów reszt	59538640	Błąd standardowy reszt	2227,454
Wsp. determ. R-kwadrat	0,991841	Skorygowany R-kwadrat	0,985043
F(10, 12)	145,8852	Wartość p dla testu F	6,02e-11
Logarytm wiarygodności	−202,4520	Kryt. inform. Akaike'a	426,9039
Kryt. bayes. Schwarza	439,3944	Kryt. Hannana-Quinna	430,0452

Wyłączając stałą, największa wartość p jest dla zmiennej 27 (d_d_X7)

Współczynniki determinacji R^2 = 0,991841 i skorygowany R^2 = 0,985043, czyli biorąc pod uwagę R^2 99,18% zmienności zostaje wyjaśnionych przez powyższy model.

Istotność zmiennych została sprawdzona testem t-Studenta. Na podstawie wyników należy stwierdzić, że największa wartość p jest ponownie dla zmiennej X7, co pozwala na stwierdzenie, że powinna zostać wykluczona również i z tego modelu. W tym modelu wyraz wolny nie jest wskazany jako istotny statystycznie (p > 0,2645), należałoby się zatem zastanowić nad możliwością jego ewentualnego usunięcia z modelu. Zmienne X3, X5, X6, X8, X9 wskazane jako najbardziej istotne, a następnie w kolejności jako istotne należy uznać zmienną X2 i X4.

Wyniki dla testu F (wartość p dla testu F jest bardzo mała, mniejsza od poziomu istotności, który można przyjąć na poziomie 1% (α = 0,01)), co upoważnia do odrzucenia hipotezy zerowej o nieistotności wszystkich zmiennych i przyjęciu hipotezy alternatywnej, że co najmniej jedna zmienna w modelu jest istotna. W tym modelu widać, że wiele zmiennych zostało wskazanych za istotne.

Również i w tym przypadku, budując ostateczny model ekonometryczny należy przeprowadzić dalszy proces doboru zmiennych objaśniających. W tym przypadku generowanie statystyk opisowych nie wniesie dużo do analizy, ponieważ dane dotyczą

zmiany przyrostów określonych zmiennych. W związku z tym kolejnym krokiem jest **zbadanie korelacji między zmiennymi stacjonarnymi**, przeprowadzając test istotności dla współczynnika korelacji liniowej Pearsona. Przy założeniu, że hipoteza zerowa ma postać H_0: $R_{ij} = 0$ (przy H_1: ~ H_0), wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) wyznaczona ze statystyki r* wynosi 0,3494 (dla n = 32). Wyniki współczynników korelacji dla zmiennych z próby przedstawia poniższa Tabela 11.

Tabela 11. Współczynniki korelacji liniowej dla obserwacji z próby (ujęcie kwartalne od 2005 do 2015 roku włącznie; obserwacje z brakującymi danymi zostały pominięte)

Współczynniki korelacji liniowej dla obserwacji z próby 2007:4-2015:3
(obserwacje z brakującymi danymi będą pominięte)
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,3494 dla n = 32

d_d_Y	X1	d_d_X2	d_d_X3	d_d_X4	
1,0000	0,2033	0,6612	0,9430	0,4429	d_d_Y
	1,0000	0,3661	0,2385	0,1608	X1
		1,0000	0,8119	0,5617	d_d_X2
			1,0000	0,5493	d_d_X3
				1,0000	d_d_X4

d_d_d_d_X5	d_d_X6	d_d_X7	d_d_X8	d_d_X9	
0,8755	0,9343	0,8598	-0,3662	-0,7340	d_d_Y
-0,0174	0,2002	0,2350	-0,2660	-0,3264	X1
0,4159	0,6835	0,7446	-0,6073	-0,8122	d_d_X2
0,7627	0,9342	0,9647	-0,7578	-0,9016	d_d_X3
0,3996	0,4593	0,6332	-0,6693	-0,7072	d_d_X4
1,0000	0,8402	0,7047	-0,1935	-0,5037	d_d_d_d_X5
	1,0000	0,8619	-0,3538	-0,7206	d_d_X6
		1,0000	-0,6270	-0,8805	d_d_X7
			1,0000	0,8764	d_d_X8
				1,0000	d_d_X9

X10	
-0,0623	d_d_Y
-0,0621	X1
-0,2122	d_d_X2
-0,1875	d_d_X3
-0,3782	d_d_X4
-0,1566	d_d_d_d_X5
-0,1186	d_d_X6
-0,1818	d_d_X7
0,3564	d_d_X8
0,2728	d_d_X9
1,0000	X10

Na podstawie powyższych wyników, należy stwierdzić, że współczynniki korelacji $|r_{ij}| > r^*$ również i w tym przypadku (w porównaniu do danych stacjonarnych) można uznać, za istotnie różniące się od zera. Ponadto, z macierzy korelacji wyników dostępnych za określone wyżej kwartały wynika, że większość zmiennych objaśniających jest dość istotnie związana ze zmienną objaśnianą.

- Podobnie jak w poprzednim modelu (dla zmiennych stacjonarnych) zmienna X3 jest najsilniej skorelowana ze zmienną objaśnianą, następnie X6 i X5, najmniej X1 i ona być może wykluczona z modelu.
- Zmienna X3 jest dość silnie skorelowana z X2, zatem X2 zostanie wykluczona z dalszych rozważań.
- Zmienna X6 nie zostaje wzięta do rozważań ponieważ jest silnie skorelowana z X3, która jest istotna statystycznie i najbardziej skorelowana z Y.
- Zmienna X5 jest nieco mniej skorelowana z X3 i można uwzględnić ją w rozważaniach.
- Zmienna X7 charakteryzuje się wysoką korelacją z Y, jednak również i z X3 i stosunkowo wysoko z X5, nie zostaje zatem wzięta pod uwagę.
- Podobnie zmienna X9 – jest wysoko skorelowana z Y, ale również wysoko z X3.
- Warto zwrócić uwagę na X4, która ma relatywnie dużą istotność statystyczną w modelu, mniejszą korelację z Y, ale niską z X3 i akceptowalną z X2.

Na podstawie powyższej analizy w dalszych etapach budowania modelu **wstępnie wybrane zostają zmienne X3, X4, X5**.

Ponieważ jednak wystąpiły wątpliwości co do możliwości uwzględnienia zmiennych X3 i X5 jednocześnie w modelu, dokonana została weryfikacja wyboru zmiennych, w oparciu o **metodę pojemności integralnej Hellwiga**. Analiza przeprowadzona w Gretlu dla 10 zmiennych wskazała za zmienne X3, X4 i X7 jako najlepsze, przy pojemności 0,96087135. W wynikach KMNK zmienna X7 jest wskazana jako mało istotna, jest ponadto bardzo skorelowana z X3. Z kolei zmienna X5 jest bardziej istotna w modelu niż X4, jednak o wiele bardziej skorelowana z X3 niż X4 Dlatego **postanowiono zbudować model w oparciu o zmienne X3 i X4**.

Wyestymowano zatem **drugi model dla zmiennych stacjonarnych**, uwzględniający oprócz zmiennej objaśnianej już dwie zmienne objaśniające: X3 i X4. Również i w tym przypadku należy zbadać ich istotność oraz sprawdzić istotność wyrazu wolnego. **Wyniki estymacji parametrów** dla **modelu uwzględniającego wybrane zmienne**, wykonanego **klasyczną metodą najmniejszych kwadratów (KMNK)**, zaprezentowano poniżej w Tabeli 12.

Tabela 12. Drugi model z estymacja parametrów metodą KMNK (ujęcie kwartalne od 2005 do 2015 roku włącznie; obserwacje z brakującymi danymi (13) zostały pominięte, zmienna zależna (Y): Y)

Model 2: Estymacja KMNK, wykorzystane obserwacje 2005:1-2015:4 (N = 25)
Liczba pominiętych niekompletnych obserwacji: 19
Zmienna zależna (Y): d_d_Y

	Współczynnik	*Błąd stand.*	*t-Studenta*	*wartość p*	
const	617,797	1113,82	0,5547	0,5847	
d_d_X3	0,397818	0,0294372	13,5141	<0,0001	***
d_d_X4	−0,135801	0,0599274	−2,2661	0,0336	**

Średn.aryt.zm.zależnej	726,1360	Odch.stand.zm.zależnej	17465,56
Suma kwadratów reszt	6,64e+08	Błąd standardowy reszt	5492,866
Wsp. determ. R-kwadrat	0,909334	Skorygowany R-kwadrat	0,901092
F(2, 22)	110,3244	Wartość p dla testu F	3,40e-12
Logarytm wiarygodności	−249,1557	Kryt. inform. Akaike'a	504,3114
Kryt. bayes. Schwarza	507,9680	Kryt. Hannana-Quinna	505,3256

W tym wypadku (w porównaniu do modelu dla danych stacjonarnych), $\mathbf{R^2}$ jest o wiele wyższy w porównaniu do modelu opartego na danych niestacjonarnych i stworzonego dla tych samych zmiennych; model stracił niewiele jeśli chodzi o wartość R^2 w porównaniu do pierwszego z 10 zmiennymi (dla zmiennych stacjonarnych). Dla nowego modelu z dwoma zmiennymi współczynniki determinacji R^2 = 0,909334, skorygowany R^2 = 0,901092, zatem różnica między współczynnikami wynosi 0,008242, czyli niewielka i można uznać, że model nie jest przeparametryzowany i wyjaśnia prawie 90% zmienności.

W drugim oszacowanym modelu według **testu F** oceniającego łączną istotność modelu, stopień wyjaśniania zmienności zmiennej zależnej jest istotny. Z kolei na podstawie **testu t-Studenta** uznać należy, że istotność zmiennych X3 i X4 w modelu jest na bardzo wysokim poziomie ($p < \alpha = 0,01$), niestety wyraz wolny nadal nie jest wskazywany jako istotny. Zmniejszyły się niestety kryteria informacyjne w porównaniu do pierwszego modelu.

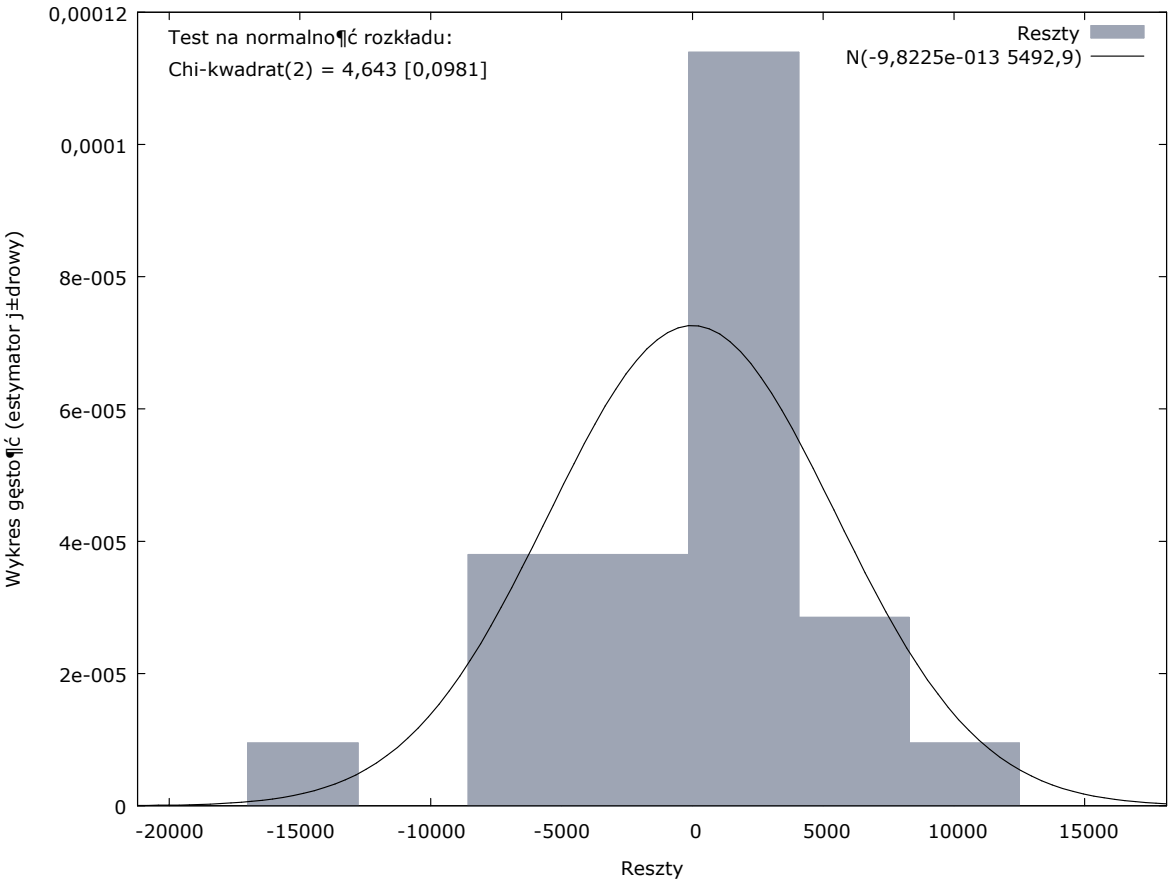
W ramach kolejnych etapów budowania modelu sprawdzone zostaną również pewne własności dotyczące rozkładu reszt – składnika losowego. W tym celu zostanie dokonana ocena **normalności reszt składnika losowego** wykorzystując w programie Gretl **test Doornika-Hansena**. Również i drugim rozważanym modelu, dla zmiennych stacjonarnych, jedynie z dwoma zmiennymi objaśniającymi X3 i X4, wyniki testu wskazują, że nie ma podstaw do odrzucenia hipotezy zerowej, ponieważ p = 0,09813 i jest większe niż przyjęty poziom istotności $\alpha = 0,05$ (statystyka testowa wynosi 4,643 i jest mniejsza niż $\chi^2_{0,05,\,2}$ = 5,991. Można zatem stwierdzić, że z prawdopodobieństwem 95% rozkład składników losowych dla zmiennych stacjonarnych reprezentuje rozkład normalny. Wyniki testu przedstawione są w poniższej Tabeli 13. jak i w formie histogramu na rysunku.

Tabela 13. Wyniki testu Doornika-Hansena na normalność rozkładu reszt (obserwacje obserwacje 1-44, liczba przedziałów = 7, średnia = -9,82254e-013, odch.std. = 5492,87)

Przedziały	średnia	liczba	częstość	skumlowana	
< -12784,	-14891,	1	4,00%	4,00%	*
-12784, - -8570,2	-10677,	0	0,00%	4,00%	
-8570,2 - -4356,6	-6463,4	4	16,00%	20,00%	*****
-4356,6 - -143,04	-2249,8	4	16,00%	36,00%	*****
-143,04 - 4070,5	1963,8	12	48,00%	84,00%	**********
4070,5 - 8284,1	6177,3	3	12,00%	96,00%	****
>= 8284,1	10391,	1	4,00%	100,00%	*

Brakujące obserwacje = 19 (43,18%)

Hipoteza zerowa: dystrybuanta empiryczna posiada rozkład normalny.Test Doornika-Hansena (1994) - transformowana skośność i kurtoza.:
Chi-kwadrat(2) = 4,643 z wartością p 0,09813

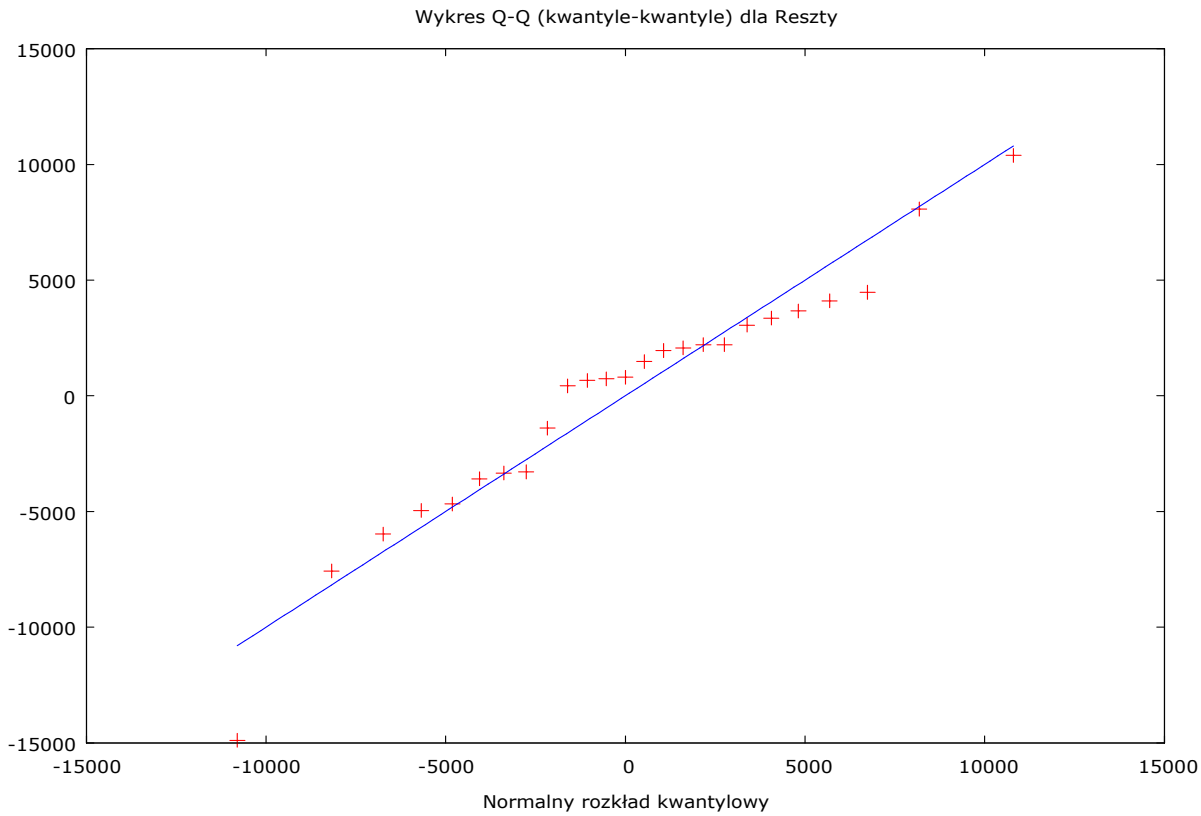


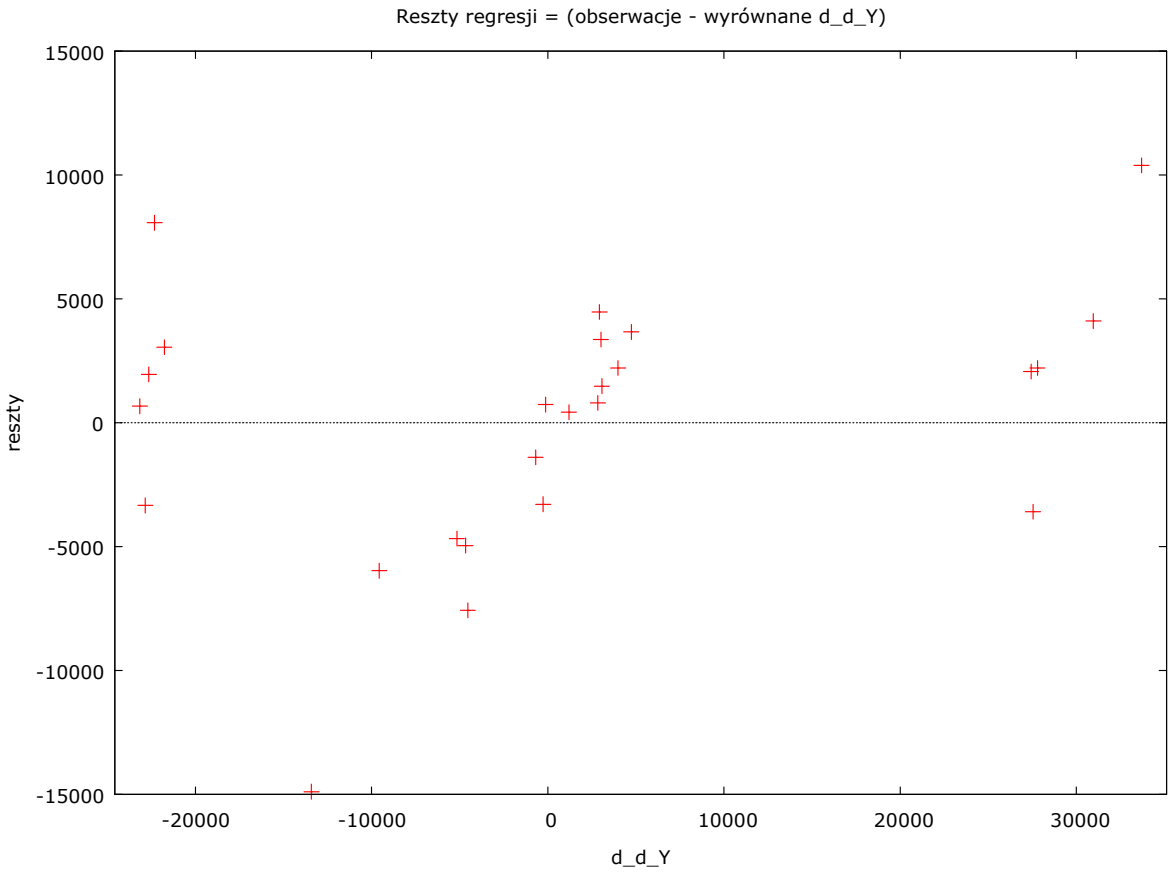
Rysunek 2. Wyniki testu Doornika-Hansena na normalność rozkładu reszt (obserwacje obserwacje 1-44, liczba przedziałów = 7, średnia = -9,82254e-013, odch.std. = 5492,87).

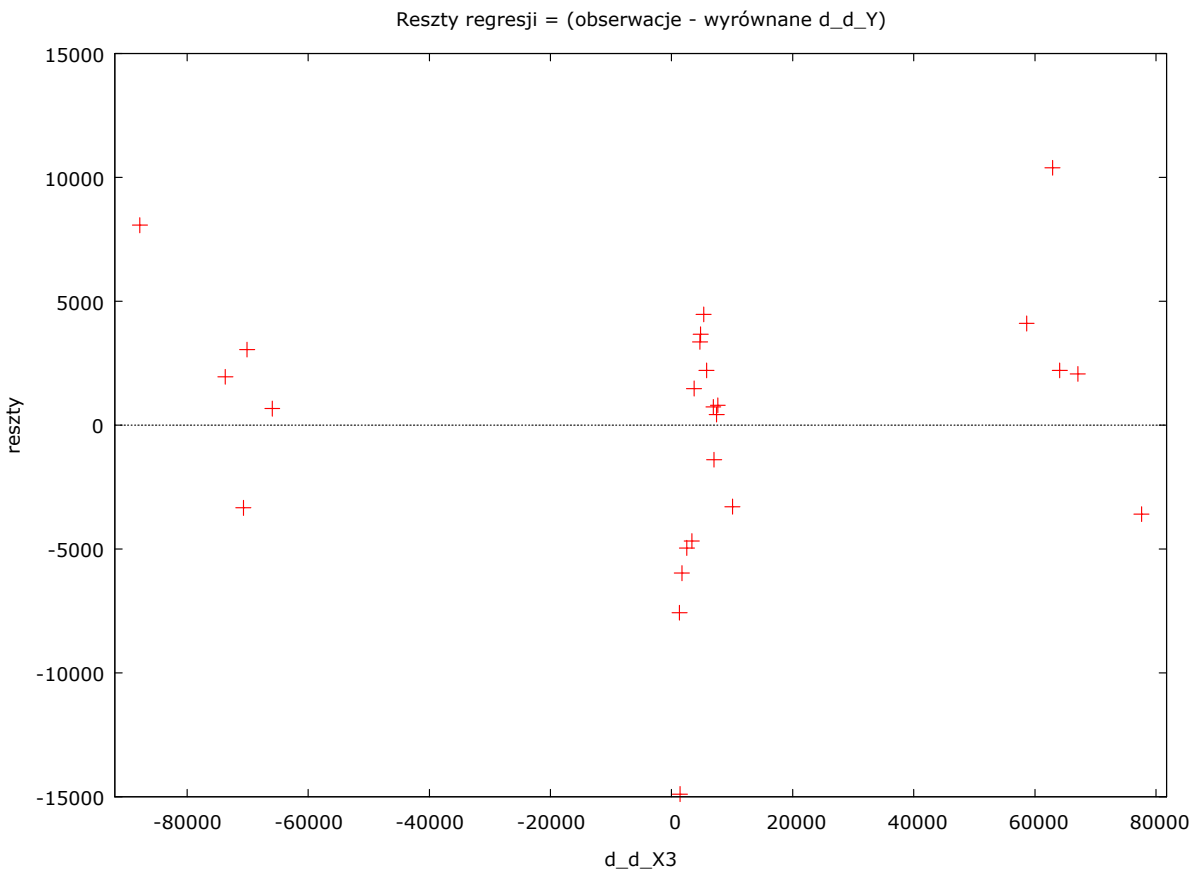
Na podstawie powyższego histogramu automatycznie wygenerowanego przez program razem ze statystyką testową, można mieć wątpliwości czy jest rozkładem normalnym. Być może zmieniając długość przedziałów klasowych dla histogramu, czytelność byłaby kompatybilna z wynikami przeprowadzonego testu wskazującego na normalność rozkładu reszt.

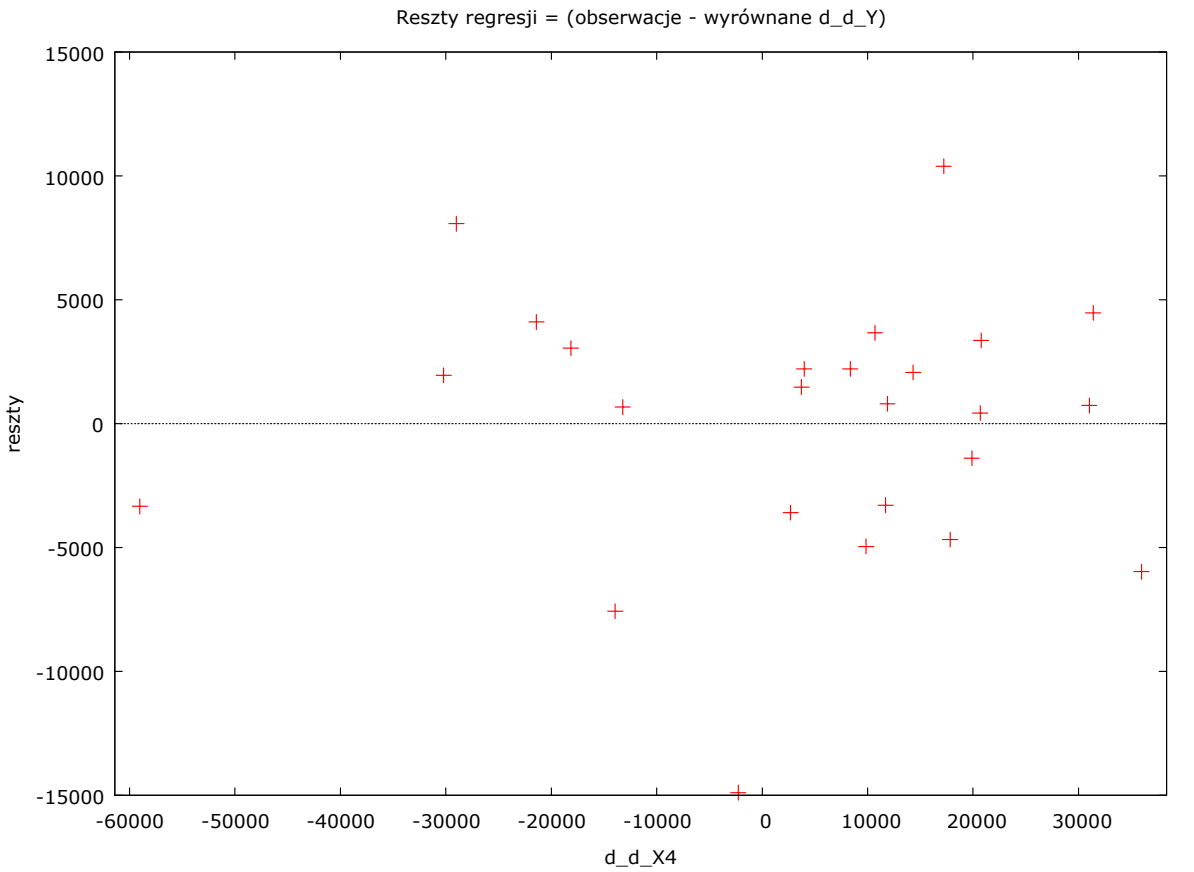
Również i w tym wypadku zostanie ominięty kolejny element weryfikacji statystycznej budowanego modelu ekonometrycznego polegający na sprawdzeniu **autokorelacji** (i jej usuwaniu) w programie Gretl, ponieważ nie jest ona dostępna do wykonania przez program dla takiego układu danych i wyników.

Niemniej jednak można próbować ocenić jakość reszt na podstawie wykresu (rozkład losowości i autokorelację). Również i w przypadku zmiennych stacjonarnych wydaje się, iż reszty te spełniają założenia regresji. Rozrzut reszt wokół zera jest niezależny od wartości zmiennych X, czy Y; reszty nie rosną też wraz ze wzrostem X lub Y, czyli ten warunek o nieskorelowaniu reszt można uznać za spełniony. Można dodać wykres kwantyle-kwantyle, który pozwala domniemywać, że autokorelacja nie występuje.


Wykres Q-Q (kwantyle-kwantyle) dla Reszty
15000
10000
5000
0
-5000
-10000
-15000
-15000
-10000
-5000
0
5000
10000
15000
Normalny rozkład kwantylowy


Reszty regresji = (obserwacje - wyrównane d_d_Y)
15000
10000
5000
0
-5000
-10000
-15000
reszty
-20000
-10000
0
10000
20000
30000
d_d_Y


Reszty regresji = (obserwacje - wyrównane d_d_Y)
15000
10000
5000
0
-5000
-10000
-15000
reszty
-80000
-60000
-40000
-20000
0
20000
40000
60000
80000
d_d_X3


Reszty regresji = (obserwacje - wyrównane d_d_Y)
15000
10000
5000
0
-5000
-10000
-15000
reszty
-60000
-50000
-40000
-30000
-20000
-10000
0
10000
20000
30000
d_d_X4

W dalszej kolejności oceny składnika losowego należy przeprowadzić **badanie heteroskedastyczności**. W celu sprawdzenia heterodskedastyczności wykonano test White'a oraz test Breuscha-Pagana. Hipotezą zerową obu tych testów jest brak heteroskedastyczności. Wyniki testów przedstawiono w poniższych tabelach.

Tabela 14. Wyniki testu White'a na heteroskedastyczność reszt (zmienność wariancji resztowej)

Test White'a na heteroskedastyczność reszt (zmienność wariancji resztowej)
Estymacja KMNK, wykorzystane obserwacje 2005:1-2015:4 (N = 25)
Liczba pominiętych niekompletnych obserwacji: 19
Zmienna zależna (Y): uhat^2

współczynnik	błąd standardowy	t-Studenta	wartość p	
const	4,24113e+07	1,92332e+07	2,205	0,0400 **
d_d_X3	227,247	349,153	0,6509	0,5229
d_d_X4	−622,603	755,944	−0,8236	0,4204
sq_d_d_X3	−0,00627641	0,00719536	−0,8723	0,3939
X2_X3	0,0170857	0,0232482	0,7349	0,4714
sq_d_d_X4	−0,0219587	0,0234105	−0,9380	0,3600

Wsp. determ. R-kwadrat = 0,068389

Statystyka testu: TR^2 = 1,709716,
z wartością p = P(Chi-kwadrat(5) > 1,709716) = 0,887673

Tabela 15. Wyniki testu Breuscha-Pagana na heteroskedastyczność reszt

Test Breuscha-Pagana na heteroskedastyczność
Estymacja KMNK, wykorzystane obserwacje 2005:1-2015:4 (N = 25)
Liczba pominiętych niekompletnych obserwacji: 19
Zmienna zależna (Y): standaryzowane uhat^2

współczynnik	błąd standardowy	t-Studenta	wartość p	
const	1,02666	0,378502	2,712	0,0127 **
d_d_X3	4,03837e-06	1,00035e-05	0,4037	0,6903
d_d_X4	−9,55874e-06	2,03648e-05	−0,4694	0,6434

Wyjaśniona suma kwadr. = 0,87394

Statystyka testu: LM = 0,436970,
z wartością p = P(Chi-kwadrat(2) > 0,436970) = 0,803736

W obu testach wartość p jest większa od poziomu istotności 0,05, nie ma więc podstaw do odrzucenia hipotezy zerowej. Zatem w estymowanym modelu nie występuje heteroskedastyczność, czyli składniki losowe modelu charakteryzuje jednorodność.

Kolejnym etapem weryfikacji modelu jest **testowanie współliniowości zmiennych modelu**. Z przeprowadzonego w tym celu **testu VIF** wynika, że w modelu nie występuje współliniowość zmiennych, ponieważ wartość czynnika jest mniejsza od 10, a nawet znajduje się w pobliżu minimalnej wartości czynnika.

Tabela 16. Wyniki testu VIF

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_d_X3	1,431
d_d_X4	1,431

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

Przeprowadzono również **test Ramsey'a RESET**. Hipotezą zerową tego testu jest liniowość modelu. Wszystkie wartości p są mniejsze od poziomu istotności 0,05, zatem hipotezę o liniowości modelu należy odrzucić.

Tabela 17. Wyniki testu Ramsey'a RESET

Test RESET na specyfikację (kwadrat i sześcian zmiennej)
Statystyka testu: F = 6,070026,
z wartością p = P(F(2,20) > 6,07003) = 0,00871

Test RESET na specyfikację (tylko kwadrat zmiennej)
Statystyka testu: F = 7,460853,
z wartością p = P(F(1,21) > 7,46085) = 0,0125

Test RESET na specyfikację (tylko sześcian zmiennej)
Statystyka testu: F = 2,152495,
z wartością p = P(F(1,21) > 2,15249) = 0,157

Istotnym elementem oceny estymowanego modelu jest **zbadanie efektu katalizy**. Należy ocenić czy wysokie wartości współczynnika determinacji R^2 i rzeczywistej siły powiązań zmiennych objaśniających w prezentowanym powyżej modelu dla zmiennych stacjonarnych nie uzasadniają takiego wyniku. Zmienna X_i z pary zmiennych $\{X_i, X_j\}$, $i < j$, jest katalizatorem, jeśli $r_{ij} < 0$ lub $r_{ij} > r_i\ r_j$. Wyniki z obliczeń mających na celu zbadanie efektu katalizy zostały przestawione poniżej.

Tabela 18. Wyniki badania efektu katalizy

Współczynniki korelacji liniowej dla obserwacji z próby 2005:4-2015:3
(obserwacje z brakującymi danymi będą pominięte)
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,3120 dla n = 40

d_d_Y	d_d_X3	d_d_X4	
1,0000	0,9424	0,4379	d_d_Y
	1,0000	0,5490	d_d_X3
		1,0000	d_d_X4

R0	
d_d_X3	0,9424
d_d_X4	0,4379

R Korelacja d_d_X3 z d_d_X4	
1	0,549
0,549	1

Korelacja d_d_X3 z Y/korelacja d_d_X4 z Y	
r3/4	r34
2,15209	0,549

Z macierzy korelacji wynika, że korelacja zmiennej X3 z X4 nie jest mniejsza od 0. Korelacja X3 z Y podzielona przez korelację X4 z Y wynosi: 2,15209, nie jest więc mniejsza niż korelacja X3 z X4 (0,5694). Z tego wynika, że zmienna X3 nie jest katalizatorem, nie zaburza więc wyniku.

Ostatnim elementem oceny estymowanego modelu będzie **zbadanie 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. Zestawienie współczynników dla szacowanego modelu zamieszczono poniżej.

Tabela 19. Zestawienie współczynników dla szacowanego modelu

	Współczynnik	Korelacja
X3	0,397818	0,8275
X4	−0,135801	0,5694

W tym modelu (w przeciwieństwie do modelu dla danych niestacjonarnych) nie zachodzi zjawisko koincydencji, ponieważ nie jest spełniony warunek, że dla każdej zmiennej znaj współczynnika jest równy znakowi współczynnika korelacji.

Można zatem przejść do sformułowania **ostatecznej postaci modelu i jego interpretacji**. W oparciu o wcześniejsze obliczenia wiadomo, że dla nowego modelu współczynniki determinacji R^2 i skorygowanego R^2 wynoszą odpowiednio 0,798177 i 0,786305, czyli różnica między współczynnikami jest niska i wynosi 0,011872. Można zatem uznać, że model wyjaśnia prawie 80% zmienności i nie jest przeparametryzowany. Współczynnik ten jest na poziomie niższym niż w pierwszym modelu, jednakże wyniki testu F oceniającego łączną istotność modelu, stopień wyjaśniania zmienności zmiennej zależnej jest wysoce istotny. Zmniejszyły się niestety kryteria informacyjne w porównaniu do pierwszego modelu. Dodatkowo, jeżeli w oparciu o powyżej przedstawione argumenty, uzna się, że występuje normalność rozkładu reszt i zakłada się, że nie ma autokorelacji, to można wnioskować, że uzyskany estymator KMNK jest nieobciążony i efektywny, średnie błędy szacunku doszacowane, oraz zarówno wartości statystyk t-Studenta jak i współczynnik determinacji nie są przeszacowane.

Ostateczną postać **analityczną modelu** można zapisać jako:

$$Y = 617{,}797 + 0{,}397818\ X_3 - 0{,}135801\ X_4$$

Interpretacja dla modelu zbudowanego w oparciu o zmienne stacjonarne, będzie nieco inna niż w przypadku modelu wyestymowanego w oparciu o dane niestacjonarne.

Mając zatem na uwadze, że:

- Y: Wartość dodana brutto ogółem (ceny bieżące) - budownictwo (mln zł) – zmienna zależna (objaśniana)
- X3: Nakłady inwestycyjne w przemyśle (od początku roku do końca okresu) (w mln zł) – zmienna niezależna (objaśniająca)
- X4: Podaż pieniądza ogółem M3 (stan w końcu okresu) (mln zł) – zmienna niezależna (objaśniająca);

powyższy model można zinterpretować, następująco:

- wzrost przyrostu nakładów inwestycyjnych w przemyśle (X3) o 1 zł powoduje wzrost przyrostu wartości dodanej brutto w budownictwie (Y) o 0,310203 zł, **ceteris paribus**;
- spadek przyrostu podaży pieniądza M3 (X4) o 1 zł, powoduje wzrost przyrostu wartości dodanej brutto w budownictwie (Y) o 0,135801 zł , **ceteris paribus**.

Należy zwrócić uwagę, że w tym przypadku mamy do czynienia z danymi stanowiącymi przyrosty i dlatego interpretacja modelu zmieniła się w porównaniu do interpretacji w modelu dla danych niestacjonarnych.

Dodatkową kwestię stanowi dylemat obecności wyrazu wolnego w wyestymowanym powyżej modelu. Być może model należy oszacować bez wyrazu wolnego, gdyż w oparciu o KMNK okazał się statystycznie nieistotny.

Wydaje się, że lepszy jest model z wyrazem wolnym, a ten występuje w modelu dla zmiennych niestacjonarnych. Poza tym, model dla danych niestacjonarnych w sposób poprawny z obowiązującą teorią ekonomii wyjaśniałby zmiany w wartości brutto w budownictwie – wzrost podaży pieniądza napędza inwestycje, co prowadzi do wzrostów w przemyśle budowlanym (ceteris paribus).