



POLITECHNIKA RZESZOWSKA  
im. Ignacego Łukasiewicza  
WYDZIAŁ MATEMATYKI I FIZYKI STOSOWANEJ

Projektowanie modeli łączenia źródeł danych

Witkowski Jakub

Rzeszów, 2023

## Spis treści

1. Wstęp .....	3
2. Dane .....	3
3. Metoda eliminacji zmiennych Quasi-stałych.....	4
4. Metoda analizy macierzy współczynników korelacji .....	5
5. Metoda Hellwiga.....	6
6. Estymacja.....	8
Metoda najmniejszych kwadratów .....	8
7. Weryfikacja.....	10
8. Testowanie reszt modelu.....	12
9. Heteroskedastyczność .....	16
10. Prognozy.....	17
11. Wnioski .....	19

## 1. Wstęp

Zadaniem było wykonanie modelu ekonometrycznego. Przedstawiony będzie wpływ na ceny mieszkań.

## 2. Dane

Dane zostały pobrane ze stat.gov.pl 28.01.2023 roku w zakresie 2005-2021

Y - cena mieszkania za m<sup>2</sup> [zł]

X1 – mieszkania oddane do użytkowania [szt.]

X2 – Ludność [os.]

X3 – Ilość małżeństw

X4 – Cena szkła okiennego ciągniętego za 1m<sup>2</sup> [zł]

X5 – Cena cegły palonej za 1m<sup>2</sup> [zł]

X6 – Liczba zarejestrowanych bezrobotnych

X7 – Przeciętne wynagrodzenie brutto

Dane w Excelu:

Rok	t	Y	X1	X2	X3	X4	X5	X6	X7
2005	1	2439,25	37 641	38 157 055	206 916	39,12	0,73	2 773 000	2 306
2006	2	2545,25	42 029	38 125 479	226 181	41,98	0,75	2 309 410	2 413
2007	3	2816	51 077	38 115 641	248 702	44,35	1,34	1 746 573	2 603
2008	4	3316,25	74 512	38 135 876	257 744	45,24	1,48	1 473 752	2 849
2009	5	3891,5	79 590	38 167 329	250 794	47,47	1,33	1 892 680	3 139
2010	6	4360,25	59 324	38 529 866	228 337	49,72	1,23	1 954 706	3 300
2011	7	3858,25	56 925	38 538 447	206 471	52,17	1,21	1 982 676	3 481
2012	8	3996,25	74 367	38 533 299	203 850	54,10	1,20	2 136 815	3 628
2013	9	4025,25	65 723	38 495 659	180 396	55,91	1,18	2 157 883	3 761
2014	10	4033,5	68 928	38 478 602	188 488	57,34	1,17	1 825 180	3 899
2015	11	3969,5	74 425	38 437 239	188 832	57,87	1,15	1 563 339	4 017
2016	12	4054	91 516	38 432 992	193 455	58,84	1,14	1 335 155	4 182
2017	13	4170	105 027	38 433 558	192 576	60,39	1,14	1 081 746	4 454
2018	14	4237,5	112 317	38 411 148	192 443	62,64	1,19	968 888	4 816
2019	15	4461,25	131 435	38 382 576	183 371	63,32	1,29	866 374	5 168
2020	16	4891,5	142 691	38 088 564	145 045	65,85	1,37	1 046 432	5 394
2021	17	5134,25	141 941	37 907 704	168 324	67,54	1,52	895 203	5537

Y - Cena mieszkania za m <sup>2</sup>
X1 - Mieszkania oddane do użytkowania
X2 - Ludność
X3 - Ilość małżeństw
X4 - Cena szkła okiennego ciągniętego za m <sup>2</sup>
X5 - Cena cegły palonej za m <sup>2</sup>
X6 - Bezrobotni zarejestrowani
X7 - Przeciętne wynagrodzenie brutto

### 3. Metoda eliminacji zmiennych Quasi-stałych

Pierwszym krokiem jest policzenie średniej i odchylenia standardowego dla zmiennych objaśniających:

średnia	3894,102941	82909,88235	38315943,18	203642,6471	54,34411765	1,201176471	1647636	3820,4
odchylenie stand	718,5600647	32106,36122	192740,9548	29456,09341	8,322062851	0,201432586	540469,629	982,1119955

Następnie licząc ze wzoru:

$$V = \frac{S}{\bar{x}}$$

liczymy współczynnik zmienności. Ustawiamy wartość krytyczną na 15% i usuwamy zmienne, które mają współczynnik zmienności mniejszy od wartości krytycznej.

vj		38,72%	0,50%	14,46%	15,31%	16,77%	32,80%	25,71%
wartość kryt.	15%	zostaje	odpada	odpada	zostaje	zostaje	zostaje	zostaje

Odpadają zmienne X2 (ludność) i X3 (ilość małżeństw). Co oznacza, że w modelu zostają X1 (mieszkania oddane do użytkowania), X4 (cena szkła okiennego ciągniętego za 1m<sup>2</sup>), X5 (Cena cegły palonej za 1m<sup>2</sup>), X6 (zarejestrowani bezrobotni), X7 (przeciętne wynagrodzenie brutto).

Rok	t	Y	X1	X4	X5	X6	X7
2005	1	2439,25	37 641	39,12	0,73	2 773 000	2 306
2006	2	2545,25	42 029	41,98	0,75	2 309 410	2 413
2007	3	2816	51 077	44,35	1,34	1 746 573	2 603
2008	4	3316,25	74 512	45,24	1,48	1 473 752	2 849
2009	5	3891,5	79 590	47,47	1,33	1 892 680	3 139
2010	6	4360,25	59 324	49,72	1,23	1 954 706	3 300
2011	7	3858,25	56 925	52,17	1,21	1 982 676	3 481
2012	8	3996,25	74 367	54,10	1,20	2 136 815	3 628
2013	9	4025,25	65 723	55,91	1,18	2 157 883	3 761
2014	10	4033,5	68 928	57,34	1,17	1 825 180	3 899
2015	11	3969,5	74 425	57,87	1,15	1 563 339	4 017
2016	12	4054	91 516	58,84	1,14	1 335 155	4 182
2017	13	4170	105 027	60,39	1,14	1 081 746	4 454
2018	14	4237,5	112 317	62,64	1,19	968 888	4 816
2019	15	4461,25	131 435	63,32	1,29	866 374	5 168
2020	16	4891,5	142 691	65,85	1,37	1 046 432	5 394
2021	17	5134,25	141 941	67,54	1,52	895 203	5537

#### 4. Metoda analizy macierzy współczynników korelacji

Tworzymy tabelę korelacji za pomocą dodatku Analiza danych. Zaznaczamy zmienne, które weszły do modelu w poprzednim kroku:

Po stworzeniu tabeli, transponujemy ją i składamy tabelę w całość:

Korelacja	Y	X1	X4	X5	X6	X7
Y	1	0,82187	0,90335	0,63749	-0,70967	0,898724
X1	0,821870385	1	0,880556	0,581532	-0,89697	0,939181
X4	0,903350211	0,880556	1	0,480401	-0,81026	0,98417
X5	0,637489999	0,581532	0,480401	1	-0,6388	0,49135
X6	-0,709669283	-0,89697	-0,81026	-0,6388	1	-0,84139
X7	0,898724155	0,939181	0,98417	0,49135	-0,84139	1

Pierwszy wiersz oznaczony innym kolorem przedstawia wektor korelacji, a pozostałe wiersze przedstawiają macierz korelacji.

Następnym krokiem jest policzenie  $r^*$ :

$$r^* = \sqrt{\frac{T^2}{T^2 + n - 2}}$$

gdzie T – jest to wartość z tablicy t-Studenta dla  $n-2$  stopni swobody i  $\alpha = 0.05$ .

Wartość krytyczna	
alfa	0,05
T	2,13145
	0,232465
$r^*$	0,482146

Następnie liczymy war. Bezwzględna tablicy korelacji:

Korelacja war. Bezw.	Y	X1	X4	X5	X6	X7
Y	1	0,82187	0,90335	0,63749	0,709669	0,898724
X1	0,821870385	1	0,880556	0,581532	0,896966	0,939181
X4	0,903350211	0,880556	1	0,480401	0,810263	0,98417
X5	0,637489999	0,581532	0,480401	1	0,638798	0,49135
X6	0,709669283	0,896966	0,810263	0,638798	1	0,841386
X7	0,898724155	0,939181	0,98417	0,49135	0,841386	1

Kolejnym krokiem jest wyrzucenie z modelu zmiennych z wiersza Y, które mają mniejszą wartość od naszego  $r^* = 0,482146$ . Jak widać żadna z wartości nie jest mniejsza od  $r^*$ , więc wszystkie zmienne pozostają w modelu.

Następnie wybieramy zmienną, której korelacja z Y jest największa. W tym przypadku ta zmienna to X4 (cena szkła okiennego ciągniętego za 1m<sup>2</sup>). Następnie nastąpi eliminacja zmiennych, których korelacja z X4 jest większa od krytycznej (0,482146). W tym przypadku eliminujemy zmienne X1, X6, X7. Zatem postać modelu będzie wyglądać następująco:

$$y = a_4 * X4 + a_5 * X5 + b$$

## 5. Metoda Hellwiga

Metodą Hellwiga pozwala na wybranie kombinacji zmiennych, które mają największą wartość pojemności informacyjnej.

Używając wcześniej stworzonej tablicy korelacji z wartościami bezwzględnymi tworzymy tablicę wszystkich możliwych kombinacji naszych zmiennych:

Kombinacja	x1	x4	x5	x6	x7	H	
x1	0,67547093	0,816042	0,406393	0,50363	0,807705107	0,67547093	
x4						0,816041604	
x5						0,406393499	
x6						0,503630492	
x7						0,807705107	
x1x4	0,359186844	0,433936	0,256962	0,265493	0,416518641	0,79312322	
x1x5	0,427099018					0,684060875	
x1x6	0,356079681					0,621572357	
x1x7	0,348327913					0,764846554	
x4x5	0,55123					0,825745943	
x4x6	0,450786	0,274516	0,278208	0,407074505	0,728994684		
x4x7	0,411276				0,818350513		
x5x6	0,247983				0,307317	0,555299737	
x5x7	0,2725				0,541593339	0,814093798	
x6x7	0,273506				0,438639677	0,71214583	
x1x4x5	0,274348787	0,34564	0,197093	0,186032	0,276294226	0,817082408	
x1x4x6	0,243191968	0,303269	0,732492593				
x1x4x7	0,239551035	0,284859	0,800703771				
x1x5x6	0,272532362	0,183033	0,198611			0,654176262	
x1x5x7	0,267968142	0,196052	0,332316322			0,796336852	
x1x6x7	0,238165006	0,356247	0,191768	0,183917	0,290482107	0,712564481	
x4x5x6	0,205642			0,753656574			
x4x5x7	0,331109			0,206108	0,326276948	0,863493854	
x4x6x7				0,292024	0,189931	0,285856996	0,767812044
x5x6x7				0,190782	0,203062	0,346247954	0,740091539
x1x4x5x6	0,201089635	0,257327	0,150475	0,150516	0,236537567	0,759408339	
x1x4x5x7	0,198593773	0,243949	0,159165	0,141923		0,214544855	0,838245772
x1x4x6x7	0,181739295	0,222053	0,149868	0,149129		0,24685988	0,760260061
x1x5x6x7	0,197640236	0,249186	0,155674	0,153058		0,243511582	0,743496768
x4x5x6x7	0,157150764		0,127313	0,120272		0,189776434	0,801429185
x1x4x5x6x7						0,79089417	

Wybieramy kombinację, dla której wartość H jest największa. W tym przypadku są to zmienne X4, X5 i X7.

Postać modelu:

$$y = a_0 + a_1 * X4 + a_2 * X5 + a_3 * X7 + \varepsilon$$

Po wykonaniu 2 metod mamy do wyboru dwie postacie modelu:

$$y = a_1 * X4 + a_2 * X5 + b$$

$$y = a_0 + a_1 * X4 + a_2 * X5 + a_3 * X7 + \varepsilon$$

Model z jedną lub dwiema zmiennymi może okazać się niewystarczający, więc wybieramy model, który otrzymaliśmy po metodzie Hellwiga.

## 6. Estymacja

Metoda najmniejszych kwadratów

Dla tej metody tworzymy tabelę Y i tabelę X:

Y		X			
2439,25		39,12	0,73	2 306	1
2545,25		41,98	0,75	2 413	1
2816		44,35	1,34	2 603	1
3316,25		45,24	1,48	2 849	1
3891,5		47,47	1,33	3 139	1
4360,25		49,72	1,23	3 300	1
3858,25		52,17	1,21	3 481	1
3996,25		54,10	1,20	3 628	1
4025,25		55,91	1,18	3 761	1
4033,5		57,34	1,17	3 899	1
3969,5		57,87	1,15	4 017	1
4054		58,84	1,14	4 182	1
4170		60,39	1,14	4 454	1
4237,5		62,64	1,19	4 816	1
4461,25		63,32	1,29	5 168	1
4891,5		65,85	1,37	5 394	1
5134,25		67,54	1,52	5537	1

Następnie transponujemy tabelę X, która będzie potrzebna do dalszych obliczeń:

XT																
39,12	41,98	44,35	45,24	47,47	49,72	52,17	54,10	55,91	57,34	57,87	58,84	60,39	62,64	63,32	65,85	67,54
0,73	0,75	1,34	1,48	1,33	1,23	1,21	1,20	1,18	1,17	1,15	1,14	1,14	1,19	1,29	1,37	1,52
2 306	2 413	2 603	2 849	3 139	3 300	3 481	3 628	3 761	3 899	4 017	4 182	4 454	4 816	5 168	5 394	5537
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Następnie liczymy iloczyn macierzy X i XT:

XT*X			
51383,178	1123,3972	3666221,442	923,85
1123,3972	25,2178	79665,026	20,42
3666221,4	79665,026	264520002,2	64946,8
923,85	20,42	64946,8	17

Macierz odwrotną do XTX:

XTX <sup>-1</sup>			
0,027053183	0,004671913	-0,000226081	-0,61207356
0,004671913	1,911948441	-0,000231641	-1,665517205
-0,00022608	-0,000231641	1,96973E-06	0,005039244
-0,61207356	-1,665517205	0,005039244	16,07007221



Iloczyn  $XY$  i  $Y$ :

$XY$
3689400,1
81086,193
263691527
66199,75

Współczynnik  $a$  liczymy ze wzoru:

$$a = (X^T X)^{-1} X^T Y$$

I otrzymujemy:

	A
$X_4$	54,10029882
$X_5$	930,5846168
$X_7$	0,112598803
Wyraz wolny	-593,8988757

Zatem model wygląda następująco:

$$y = -593,8989 + 54,1 * X_4 + 930,585 * X_5 + 0,113 * X_7$$

Używając ponownie dodatku Analiza danych robimy regresję z naszej tabeli:

PODSUMOWANIE - WYJŚCIE								
Statystyki regresji								
Wielokrotność	0,933071463							
R kwadrat	0,870622355							
Dopasowany	0,840765975							
Błąd standard	295,5598806							
Obserwacje	17							
ANALIZA WARIANCJI								
	df	SS	MS	F	Istotność F			
Regresja	3	7641962,273	2547320,758	29,1603458	4,83056E-06			
Resztkowy	13	1135623,359	87355,64302					
Razem	16	8777585,632						
	Współczynniki	standardow	t Stat	Wartość-p	Dołne 95%	Górne 95%	Dołne 95,0%	Górne 95,0%
Przecięcie	-593,8988757	1184,825511	-0,50125429	0,624572438	-3153,558773	1965,761021	-3153,558773	1965,761021
$X_4$	54,10029882	48,61325124	1,112871438	0,285919114	-50,92224544	159,1228431	-50,92224544	159,1228431
$X_5$	930,5846168	408,680175	2,277048591	0,040338819	47,6847761	1813,484457	47,6847761	1813,484457
$X_7$	0,112598803	0,414809899	0,27144676	0,790310179	-0,7835435	1,008741106	-0,7835435	1,008741106

## 7. Weryfikacja

Obliczamy miary dopasowania z naszej tabeli ze zmiennymi i tabelki A:

Rok	t	Y	X4	X5	X7	Y^	e	e^2	$(y_k - \bar{y}_k)^2$	$(\hat{y}_k - \bar{y}_k)^2$
2005	1	2439,25	39,12	0,73	2 306	2461,48442	-22,23442412	494,369616	2116597,08	2052395,81
2006	2	2545,25	41,98	0,75	2 413	2646,87104	-101,621043	10326,83638	1819404,257	1555587,41
2007	3	2816	44,35	1,34	2 603	3345,52745	-529,5274477	280399,3179	1162305,952	300935,072
2008	4	3316,25	45,24	1,48	2 849	3551,65787	-235,4078655	55416,86315	333914,0216	117268,63
2009	5	3891,5	47,47	1,33	3 139	3565,36749	326,1325078	106362,4126	6,775302768	108066,995
2010	6	4360,25	49,72	1,23	3 300	3612,16311	748,0868898	559633,9947	217293,0804	79490,068
2011	7	3858,25	52,17	1,21	3 481	3746,47753	111,7724667	12493,08431	1285,433391	21793,2609
2012	8	3996,25	54,10	1,20	3 628	3858,13729	138,1127121	19075,12124	10434,02163	1293,52818
2013	9	4025,25	55,91	1,18	3 761	3952,42278	72,82722277	5303,804376	17199,55104	3401,20333
2014	10	4033,5	57,34	1,17	3 899	4036,01899	-2,518993192	6,3453267	19431,54001	20140,166
2015	11	3969,5	57,87	1,15	4 017	4059,36712	-89,86711798	8076,098895	5684,716479	27312,2483
2016	12	4054	58,84	1,14	4 182	4121,11736	-67,11736417	4504,740573	25567,06942	51535,5485
2017	13	4170	60,39	1,14	4 454	4235,5997	-65,59970175	4303,32087	76119,18707	116620,038
2018	14	4237,5	62,64	1,19	4 816	4444,61537	-207,1153716	42896,77716	117921,54	303063,937
2019	15	4461,25	63,32	1,29	5 168	4614,09682	-152,8468152	23362,1489	321655,7863	518391,179
2020	16	4891,5	65,85	1,37	5 394	4850,86467	40,63533001	1651,230045	994800,8929	915393,007
2021	17	5134,25	67,54	1,52	5537	5097,96098	36,28902341	1316,89322	1537964,728	1449274,17

Gdzie:

$Y^{\wedge}$ :

$$y = -593,8989 + 54,1 * X4 + 930,585 * X5 + 0,113 * X7$$

$e = Y - Y^{\wedge}$

Kolejne el. SSE -  $e^2$  – policzone wcześniej e do kwadratu

Kolejne el. SST -  $(y_k - \bar{y}_k)^2$

Kolejne el. SSR -  $(\hat{y}_k - \bar{y}_k)^2$

Dalej liczymy sumę kolejno SSE, SST i SSR:

1135623,359	8777585,632	7641962,27
SSE	SST	SSR

Kolejnym krokiem jest policzenie wsp. Determinacji  $R^2$ , który jest opisany wzorem:

$$R^2 = \frac{SSR}{SST}$$

I wynosi:

Statystyki regresji	
Wielokrotność R	0,933071463
R kwadrat	0,870622355
Dopasowany R kw	0,840765975
Błąd standardowy	295,5598806
Obserwacje	17

Model wyjaśnia 87% zmienności badanej cechy, więc model można zweryfikować pozytywnie.

Następnie liczymy wsp. Zbieżności  $\varphi^2 = 1 - R^2$ :

Wsp. Zbieżności
0,129377645

**Współczynnik zmienności** mówi nam jaki procent zmienności objaśnej zmiennej nie pasuje do modelu. W tym przypadku jest to ok. 13%.

**Błąd standardowy**, który pokazuje ile parametr może się zmieniać w różnych badaniach tego zjawiska:

Błąd standardowy
295,5598806

**Współczynnik wyrazistości** mówi jaką część średniej wartości zmiennej Y stanowi jej odchylenie standardowe reszt dla modelu. Im mniejsza wartość tym lepszy model. Wyrażany wzorem:

$$w = \frac{s}{\bar{y}} \cdot 100\%.$$

W tym przypadku wynosi:

Wyrazistość
8%

Oznacza to, że 8% części średniej wartości Y stanowi jej odchylenie.

### Badanie koincydencji

Korelacja	Y		Współczynniki
Y	1	Przecięcie	-593,8988757
X4	0,903350211	X4	54,10029882
X5	0,637489999	X5	930,5846168
X7	0,898724155	X7	0,112598803

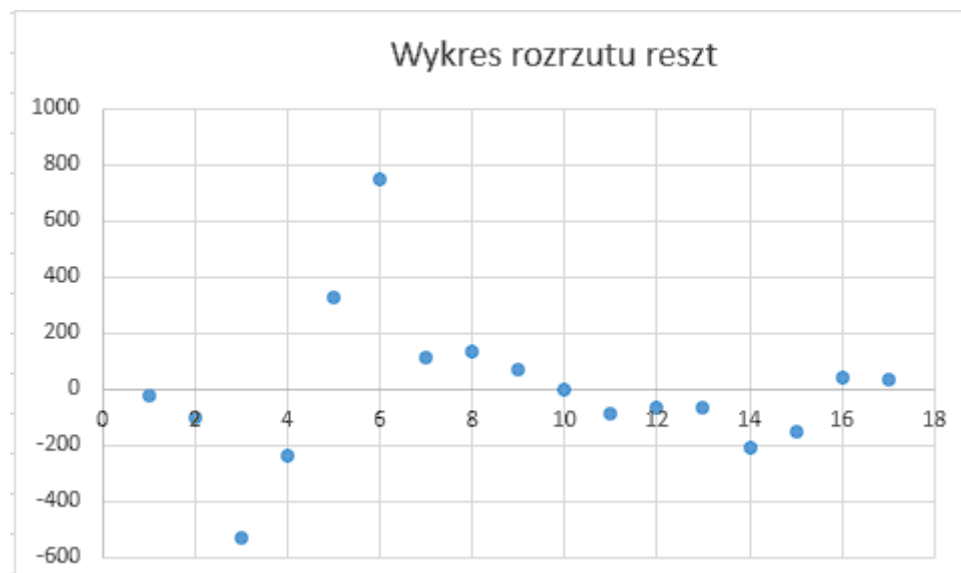
Znaki dla wszystkich zmiennych się zgadzają, więc szacunki parametrów modelu poprawnie wskazują kierunek zależności między Y, a X4,X5,X7 zgodnie z zależnością wynikającą z danych empirycznych.

## 8. Testowanie reszt modelu

Reszty modelu to estymatory składnika losowego. Obliczone reszty w naszym modelu:

$\hat{y}$	e
2461,48442	-22,23442412
2646,87104	-101,621043
3345,52745	-529,5274477
3551,65787	-235,4078655
3565,36749	326,1325078
3612,16311	748,0868898
3746,47753	111,7724667
3858,13729	138,1127121
3952,42278	72,82722277
4036,01899	-2,518993192
4059,36712	-89,86711798
4121,11736	-67,11736417
4235,5997	-65,59970175
4444,61537	-207,1153716
4614,09682	-152,8468152
4850,86467	40,63533001
5097,96098	36,28902341

Wykres rozrzutu reszt:



**Badanie składnika losowego – symetryczność:**

Badanie to ma na celu ocenić trafność wyboru postaci analitycznej modelu.

Sprawdamy ile z policzonych reszt jest dodatnia:

$\hat{Y}$	e	czy >0?
2461,484	-22,2344	ujemne
2646,871	-101,621	ujemne
3345,527	-529,527	ujemne
3551,658	-235,408	ujemne
3565,367	326,1325	dodatnie
3612,163	748,0869	dodatnie
3746,478	111,7725	dodatnie
3858,137	138,1127	dodatnie
3952,423	72,82722	dodatnie
4036,019	-2,51899	ujemne
4059,367	-89,8671	ujemne
4121,117	-67,1174	ujemne
4235,6	-65,5997	ujemne
4444,615	-207,115	ujemne
4614,097	-152,847	ujemne
4850,865	40,63533	dodatnie
5097,961	36,28902	dodatnie

Liczba dodatnich reszt wynosi 7, a ujemnych 10.

Biorąc pod uwagę tablicę testu serii dla  $n=17$  liczba reszt dodatnich mieści się w przedziale  $\langle 5, 13 \rangle$ . Rozkład reszt jest symetryczny.

### Test normalności rozkładu składnika losowego Shapiro-Wilka

$H_0$ - składnik losowy ma rozkład normalny

$H_1$ - składnik losowy nie ma rozkładu normalnego

Rok	t	e	u rosnąco	u malejąco	a	(um-ur)*a	(ur-średnia)^2
2007	3	-529,5274480	-2,04878214	2,89440531	0,4968	2,455775524	4,197510005
2008	4	-235,4078660	-0,91081101	1,26183158	0,3273	0,711105921	0,829577483
2018	14	-207,1153720	-0,80134519	0,53436863	0,254	0,339271311	0,642154807
2019	15	-152,8468150	-0,59137600	0,43245621	0,1988	0,203537844	0,349726083
2006	2	-101,6210430	-0,39317958	0,28177408	0,1524	0,102862937	0,154590518
2015	11	-89,8671180	-0,34770274	0,15722119	0,1109	0,055996064	0,120897497
2016	12	-67,1173642	-0,25968221	0,14004500	0,0725	0,029006323	0,067435073
2017	13	-65,5997018	-0,25381026	-0,00974618	0,0359	0,0087619	0,064419865
2005	1	-22,2344241	-0,08602669	-0,08602669	0	0	0,007400665
2014	10	-2,5189932	-0,00974618	-0,25381026	0	0	9,49964E-05
2021	17	36,2890234	0,14004500	-0,25968221	0	0	0,019713444
2020	16	40,6353300	0,15722119	-0,34770274	0	0	0,024718369
2013	9	72,8272228	0,28177408	-0,39317958	0	0	0,079396393
2011	7	111,7724667	0,43245621	-0,59137600	0	0	0,187018
2012	8	138,1127121	0,53436863	-0,80134519	0	0	0,285549373
2009	5	326,1325078	1,26183158	-0,91081101	0	0	1,592217846
2010	6	748,0868898	2,89440531	-2,04878214	0	0	8,377579584
	średnia	4,29412E-07					
	odch.stand	258,459618			suma	3,906317824	17
					W	0,897606997	
	n	7					
	alfa	0,05					
	obszar kryt.	(0;0,892)					
	W	0.897606997					

Obszar krytyczny:  $(0; 0,892)$  i  $W > W^*$ .

Nie ma podstaw do odrzucenia hipotezy zerowej.

## Test serii

$H_0$ - wektor reszt ma charakter losowy

$H_1$ - wektor reszt nie ma charakteru losowego

W tabeli zostały ponumerowane nr. serii, a następnie posortowane rosnąco:

Rok	t	Y	e	Nr serii	Nr serii dla posortowanych
2021	17	5134,25	36,28902	4	1
2020	16	4891,5	40,63533	4	1
2019	15	4461,25	-152,847	3	2
2010	6	4360,25	748,0869	2	3
2018	14	4237,5	-207,115	3	4
2017	13	4170	-65,5997	3	4
2016	12	4054	-67,1174	3	4
2014	10	4033,5	-2,51899	3	4
2013	9	4025,25	72,82722	2	5
2012	8	3996,25	138,1127	2	5
2015	11	3969,5	-89,8671	3	6
2009	5	3891,5	326,1325	2	7
2011	7	3858,25	111,7725	2	7
2008	4	3316,25	-235,408	1	8
2007	3	2816	-529,527	1	8
2006	2	2545,25	-101,621	1	8
2005	1	2439,25	-22,2344	1	8

Przedział jest równy  $\langle 5,13 \rangle$ , a liczba serii wynosi 8, więc mieści się w przedziale.

Nie ma podstaw do odrzucenia hipotezy zerowej.

### Badanie autokorelacji Durbina-Watsona

Wypisujemy nasze obserwacje i jakie przypadają im reszty.

Rok	t	e	e(t-1)	e-e(t-1)	(e-e(t-1))^2	e^2
2005	1	-22,2344				494,3685
2006	2	-101,621	-22,2344	-79,3866	6302,23226	10326,83
2007	3	-529,527	-101,621	-427,906	183103,5448	280398,8
2008	4	-235,408	-529,527	294,119	86505,98616	55416,93
2009	5	326,1325	-235,408	561,5405	315327,7331	106362,4
2010	6	748,0869	326,1325	421,9544	178045,5157	559634
2011	7	111,7725	748,0869	-636,3144	404896,0156	12493,09
2012	8	138,1127	111,7725	26,3402	693,806136	19075,12
2013	9	72,82722	138,1127	-65,28548	4262,193899	5303,804
2014	10	-2,51899	72,82722	-75,34621	5677,051361	6,345311
2015	11	-89,8671	-2,51899	-87,34811	7629,692321	8076,096
2016	12	-67,1174	-89,8671	22,7497	517,5488501	4504,745
2017	13	-65,5997	-67,1174	1,5177	2,30341329	4303,321
2018	14	-207,115	-65,5997	-141,5153	20026,58013	42896,62
2019	15	-152,847	-207,115	54,268	2945,015824	23362,21
2020	16	40,63533	-152,847	193,48233	37435,41202	1651,23
2021	17	36,28902	40,63533	-4,34631	18,89041062	1316,893

Na koniec liczymy sumę 2 ostatnich kolumn:

suma	1253389,522	1135623
------	-------------	---------

Statystyka testu to iloraz pierwszej i drugiej wartości sumy i wynosi:

stat. Testu	1,103702268
-------------	-------------

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

Sprawdzamy war. kryt. Dla rozkładu Durbina-Watsona. Odczytujemy przedział dla  $\alpha=0,05$ ,  $n=17$  i  $k=3$ .

dL	0,9
dG	1,71

Przedział wynosi  $\langle 0,9;1,71 \rangle$

Statystyka mieści się w naszym przedziale, zatem nie ma podstaw do odrzucenia hipotezy zerowej.

## 9. Heteroskedastyczność

### Test Goldfelda-Quandt

$H_0$ - homoskedastyczność wariancji reszt

$H_1$ - heteroskedastyczność

Dzielimy reszty na dwie próby, obliczamy kwadraty reszt i wyliczamy liczebność każdej z prób, sumę kwadratów oraz wariancję.

Rok	t	e	e^2	Rok	t	e	e^2
2005	1	-22,2344	494,3685434	2013	9	72,82722	5303,804
2006	2	-101,621	10326,82764	2014	10	-2,51899	6,345311
2007	3	-529,527	280398,8437	2015	11	-89,8671	8076,096
2008	4	-235,408	55416,92646	2016	12	-67,1174	4504,745
2009	5	326,1325	106362,4076	2017	13	-65,5997	4303,321
2010	6	748,0869	559634,01	2018	14	-207,115	42896,62
2011	7	111,7725	12493,09176	2019	15	-152,847	23362,21
2012	8	138,1127	19075,1179	2020	16	40,63533	1651,23
				2021	17	36,28902	1316,893
	n1	8					
	suma1	1044202			n1	9	
	wariancja1	130521,2			suma1	91421,26	
					wariancja1	10153,92	

Nie ma podstaw do odrzucenia hipotezy zerowej.

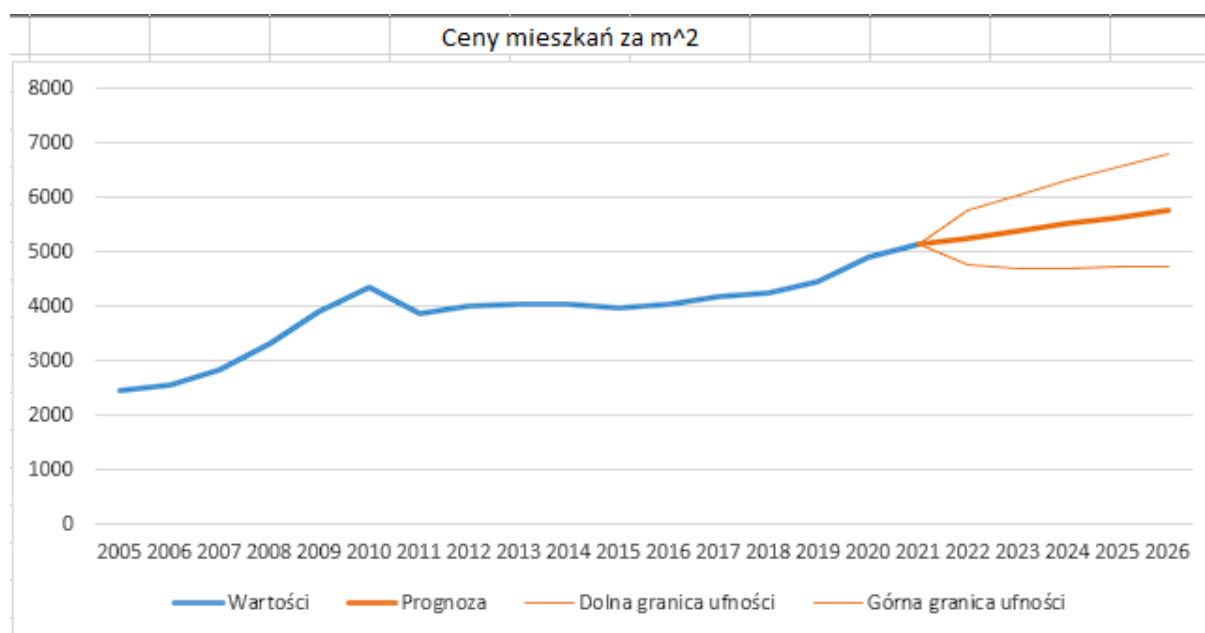


## 10. Prognozy

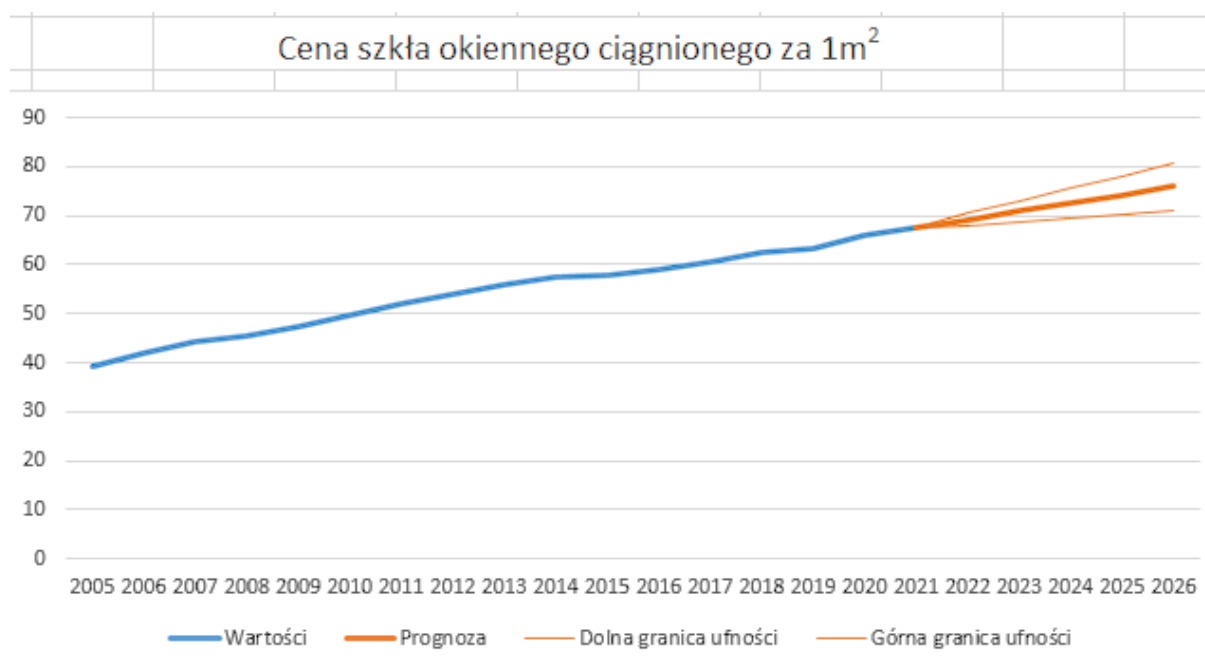
Prognozy wyglądają następująco:

Rok	t	Y	X4	X5	X7
2005	1	2439,25	39,12	0,73	2 306
2006	2	2545,25	41,98	0,75	2 413
2007	3	2816	44,35	1,34	2 603
2008	4	3316,25	45,24	1,48	2 849
2009	5	3891,5	47,47	1,33	3 139
2010	6	4360,25	49,72	1,23	3 300
2011	7	3858,25	52,17	1,21	3 481
2012	8	3996,25	54,10	1,20	3 628
2013	9	4025,25	55,91	1,18	3 761
2014	10	4033,5	57,34	1,17	3 899
2015	11	3969,5	57,87	1,15	4 017
2016	12	4054	58,84	1,14	4 182
2017	13	4170	60,39	1,14	4 454
2018	14	4237,5	62,64	1,19	4 816
2019	15	4461,25	63,32	1,29	5 168
2020	16	4891,5	65,85	1,37	5 394
2021	17	5134,25	67,54	1,52	5537
2022	18	5250,268	69,22088	1,525707	5626,922
2023	19	5380,63	70,90179	1,545173	5825,476
2024	20	5510,993	72,58271	1,564639	6024,031
2025	21	5641,355	74,26362	1,584105	6222,585
2026	22	5771,717	75,94453	1,60357	6421,139

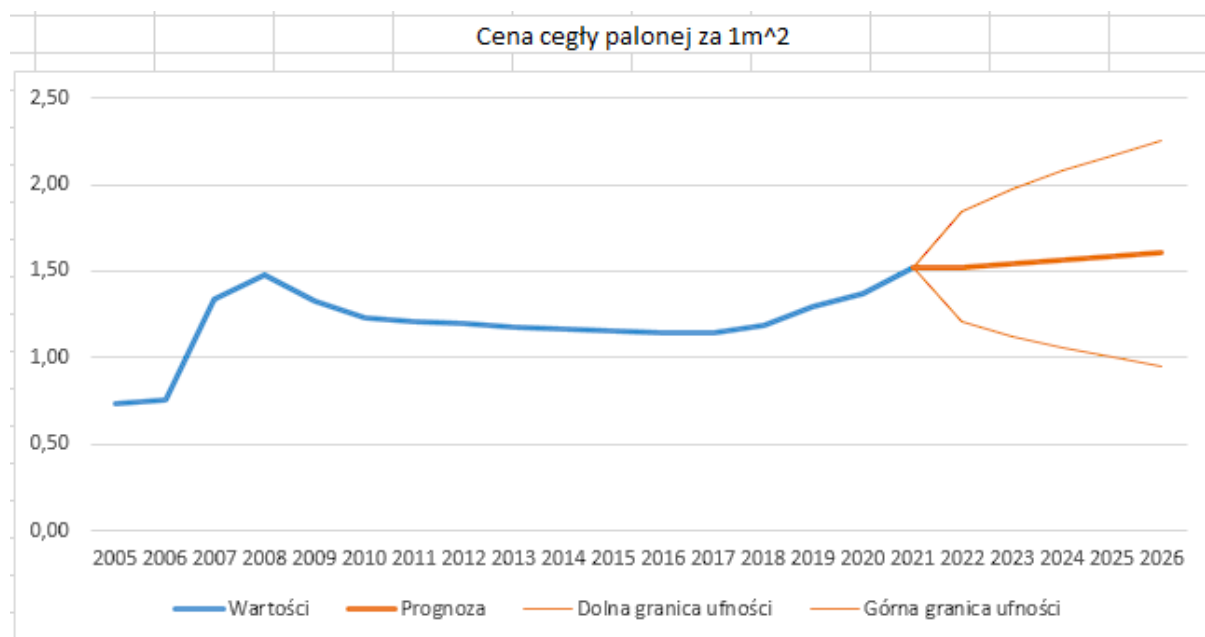
Dla cen mieszkań za m<sup>2</sup>:



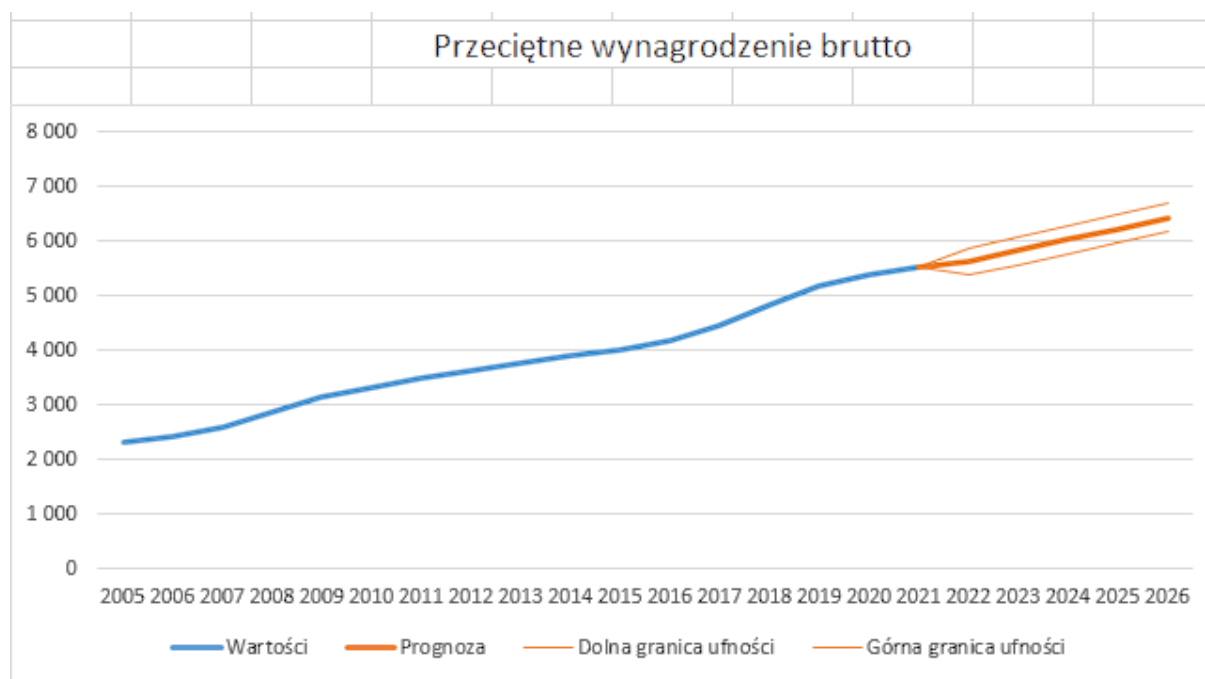
Dla cen szkła okiennego ciągnionego:



Dla ceny cegły palonej:



Dla przeciętnego wynagrodzenia brutto:



## 11. Wnioski

Po przeprowadzeniu wielu testów model ma bardzo dobre dopasowanie na poziomie 87%. Jak można było zauważyć ceny mieszkań zależą od cen materiałów budowlanych, oraz od wynagrodzenia obywateli. Model jest symetryczny, ma rozkład normalny, wektor reszt ma charakter losowy i ma homoskedastyczną wariancję reszt.