WYDZIAŁ
MATEMATYKI
I FIZYKI STOSOWANEJ
POLITECHNIKI RZESZOWSKIEJ

EKONOMETRIA - projekt

Temat pracy:
„Liczba lokali mieszkalnych na Podkarpaciu sprzedanych w ramach transkacji rynkowych w latach 2014 - 2023”

Autorzy:
Mateusz Rzeźnikiewicz oraz Patryk Ryba

Opiekun pracy:
dr inż. Marek Kiczek

Rzeszów, 2025

Spis treści

1 Opis źródeł danych **2**
1.1 Opis i pochodzenie danych 2
1.2 Zmienne objaśniające 2
1.3 Uzasadnienie doboru poszczególnych zmiennych 3
1.4 Zmienna objaśniana 3
1.5 Wykresy zależności zmiennych objaśniających od czasu 4
1.6 Wykres zależności zmiennej objaśnianej od czasu 8

2 Arkusz roboczy **9**
2.1 Analiza doboru zmiennych 9
2.2 Wybór postaci modelu 11
2.3 Estymacja parametrów 13

3 Weryfikacja modelu **14**
3.1 Zbadanie własności koincydencji 14
3.2 Test t-Studenta istotności parametrów 14
3.3 Współczynnik determinacji 15
3.4 Współczynnik zbieżności 15
3.5 Test istotności współczynnika determinacji 16
3.6 Test t dla korelacji 17
3.7 Współczynnik wyrazistości 18
3.8 Variance Inflation Factor 18

4 Testy dotyczące składnika losowego **20**
4.1 Test symetrii składnika losowego 20
4.2 Test serii składnika losowego 21
4.3 Badanie normalności rozkładu składnika losowego 23
4.4 Badanie heteroskedastyczności składnika losowego 24
4.5 Badanie autokorelacji składnika losowego 24

5 Prognozowanie **26**
5.1 Ocena zgodności wartości rzeczywistych i przewidywanych 26
5.2 Określenie przyszłego przebiegu rozważanego zjawiska 27

1 Opis źródeł danych

1.1 Opis i pochodzenie danych

Projekt dotyczy stworzenia modelu ekonometrycznego prognozującego sprzedaż lokali mieszkalnych na rynku pierwotnym i wtórnym w województwie podkarpackim. Model opierać się będzie na danych z Rejestru Cen Nieruchomości (RCN), które uwzględniają zarówno transakcje wolnorynkowe, jak i sprzedaż przetargową. W ramach projektu, rynek pierwotny i wtórny będą definiowane w następujący sposób:

- **Rynek pierwotny** – obejmuje transakcje z osobami prawnymi oraz sprzedaż lokali mieszkalnych, w których cena za metr kwadratowy wynosi co najmniej 2000 zł. Dotyczy to nowych mieszkań lub takich, które jeszcze nie były przedmiotem obrotu.
- **Rynek wtórny** – dotyczy transakcji rynkowych przeprowadzanych na rynku wolnym, w tym sprzedaży mieszkań używanych, a także sprzedaży lokali mieszkalnych w ramach przetargów.

Dane dotyczą transakcji kupna i sprzedaży lokali mieszkalnych w budynkach wielorodzinnych (mieszkania w blokach) na terenie Podkarpacia. Zaktualizowane informacje za rok 2014 zostały opublikowane w grudniu 2021 roku.

Pełen dostęp do danych z Banku Danych Lokalnych znajduje się pod tym odnośnikiem, a dane demograficzne, które mogą wspierać analizę, dostępne są na stronie Baza Demografia - Główny Urząd Statystyczny.

Na podstawie zebranych danych opracujemy model ekonometryczny, który umożliwi określenie zależności między różnymi zmiennymi. Ostatecznie stworzymy uproszczoną, matematyczną reprezentację rzeczywistości, która będzie prognozować przyszłe trendy sprzedaży lokali mieszkalnych w tym regionie.

1.2 Zmienne objaśniające

Zmienne objaśniające X_i dla $i \in \{1, \ldots, 8\}$ to czynniki wpływające na wartość zmiennej objaśnianej Y. Analiza tych zmiennych pozwala oszacować zależności występujące pomiędzy nimi a Y. W projekcie przyjęto następujące zmienne objaśniające:

- $\mathbf{X}_1$ – Stopa bezrobocia rejestrowanego (%), dane z dnia 27.03.2025r.
- $\mathbf{X}_2$ – Mieszkania oddane do użytkowania (liczba mieszkań), dane z dnia 20.03.2025r.
- $\mathbf{X}_3$ – Średnia cena za 1 m^2 lokali mieszkalnych (zł), dane z dnia 31.01.2025r.
- $\mathbf{X}_4$ – Inflacja (okres poprzedni = 100), dane z dnia 06.03.2025r.
- $\mathbf{X}_5$ – Przeciętne miesięczne wynagrodzenie brutto (zł), dane z dnia 16.04.2025r.
- $\mathbf{X}_6$ – Aktywność ekonomiczna ludności w wieku 15-89 lat według BAEL (%), dane z dnia 26.02.2025r.
- $\mathbf{X}_7$ – Liczba rozwodów (ilość rozwiedzionych par), dane z dnia 18.06.2024r.
- $\mathbf{X}_8$ – Liczba urodzeń (ilość osób), dane z dnia 20.05.2024r.

1.3 Uzasadnienie doboru poszczególnych zmiennych

Dobór zmiennych objaśniających został dokonany na podstawie ich potencjalnego wpływu na popyt i podaż mieszkań na rynku nieruchomości. Poniżej przedstawiamy krótkie uzasadnienie dla każdej z wybranych zmiennych:

- **Stopa bezrobocia rejestrowanego ($\mathbf{X}_1$):** wyższy poziom bezrobocia może obniżać zdolność kredytową i ograniczać popyt na mieszkania, dlatego oczekuje się, że zmienna ta będzie miała wpływ ujemny na sprzedaż lokali.
- **Liczba mieszkań oddanych do użytkowania ($\mathbf{X}_2$):** większa liczba nowo oddanych mieszkań zwiększa podaż, co może prowadzić do większej liczby transakcji na rynku pierwotnym.
- **Średnia cena 1 m^2 mieszkania ($\mathbf{X}_3$):** wyższe ceny mogą ograniczać dostępność mieszkań dla nabywców, szczególnie przy rosnących stopach procentowych. Może to prowadzić do zmniejszenia liczby zawieranych transakcji.
- **Inflacja ($\mathbf{X}_4$):** wzrost inflacji wpływa na wartość pieniądza i może zmniejszać realne dochody, ograniczając możliwości zakupu nieruchomości. Z drugiej strony, wysoka inflacja może skłaniać inwestorów do lokowania kapitału w nieruchomości.
- **Przeciętne miesięczne wynagrodzenie brutto ($\mathbf{X}_5$):** wyższe wynagrodzenia mogą zwiększać zdolność kredytową i siłę nabywczą mieszkańców, sprzyjając zakupom mieszkań.
- **Aktywność ekonomiczna ludności ($\mathbf{X}_6$):** większy udział aktywnych zawodowo mieszkańców może świadczyć o większym zapotrzebowaniu na mieszkania, szczególnie w miastach rozwijających się.
- **Liczba rozwodów ($\mathbf{X}_7$):** rozwody mogą prowadzić do konieczności zakupu dodatkowego mieszkania (np. przez jednego z małżonków), co potencjalnie zwiększa popyt.
- **Liczba urodzeń ($\mathbf{X}_8$):** większa liczba urodzeń może przekładać się na większe potrzeby mieszkaniowe rodzin i wzrost zapotrzebowania na większe mieszkania.

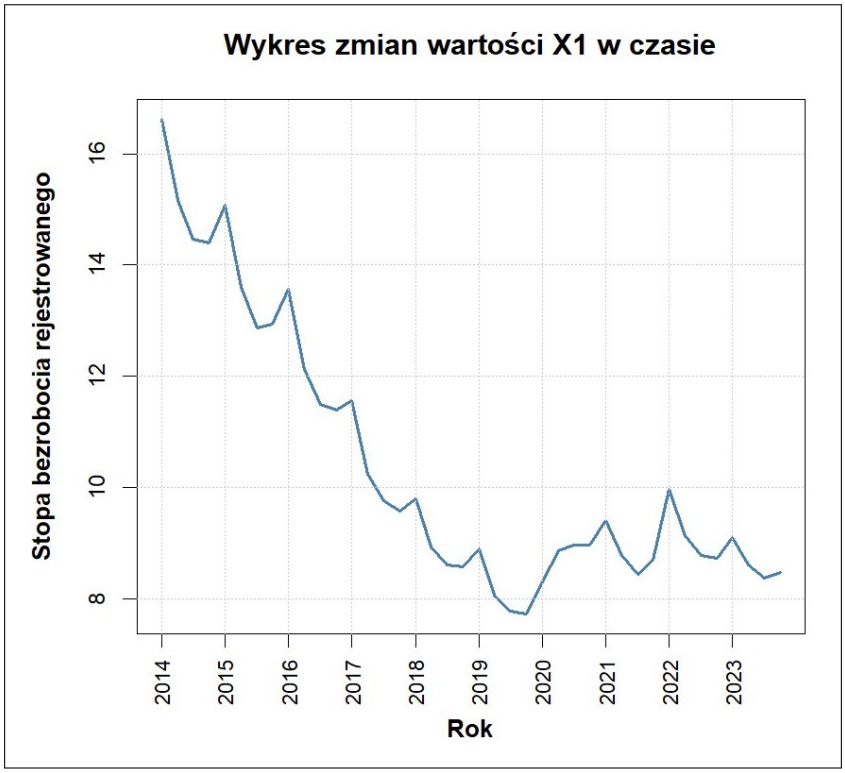
1.4 Zmienna objaśniana

Zmienna objaśniana Y, wybrana do budowy naszego modelu, to liczba lokali mieszkalnych sprzedanych w ramach transakcji rynkowych. Reprezentuje ona całkowitą liczbę mieszkań w blokach sprzedanych na terenie Podkarpacia w latach 2014–2023, w rozbiciu na okresy kwartalne, na podstawie aktów notarialnych zawartych w Rejestrze Cen Nieruchomości (RCN).

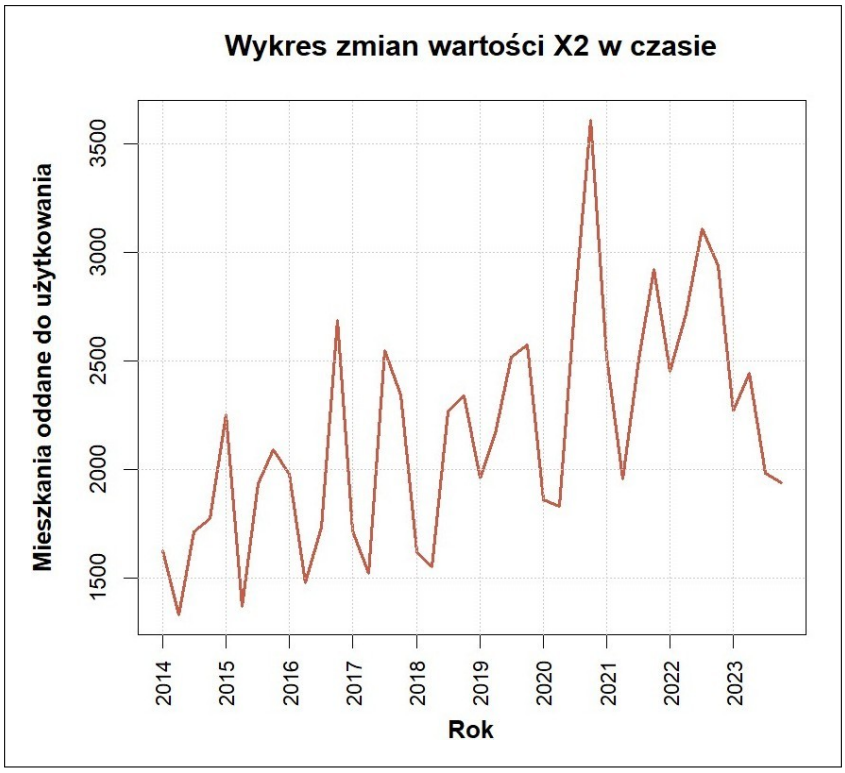
Wartości dla tej zmiennej zostały przedstawione w postaci sumarycznej liczby sprzedanych mieszkań bez podziału na metraż, rok budowy czy stan techniczny lokali.

Zmienna ta jest traktowana jako miara aktywności rynku mieszkaniowego i pełni kluczową rolę w modelu prognostycznym, umożliwiając ocenę wpływu wybranych czynników ekonomiczno-społecznych na dynamikę sprzedaży mieszkań.

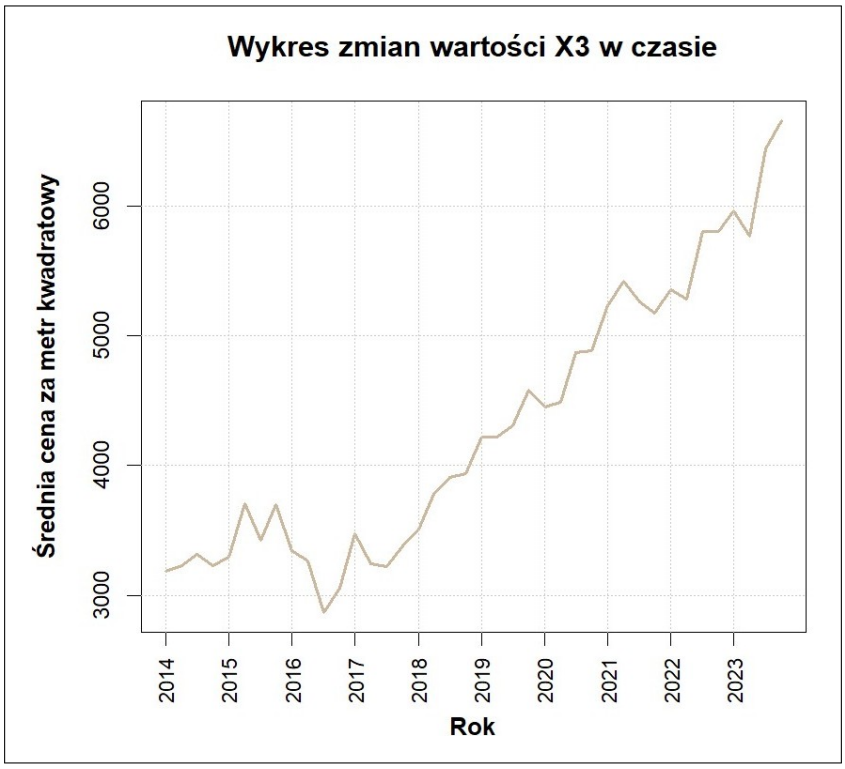
1.5 Wykresy zależności zmiennych objaśniających od czasu



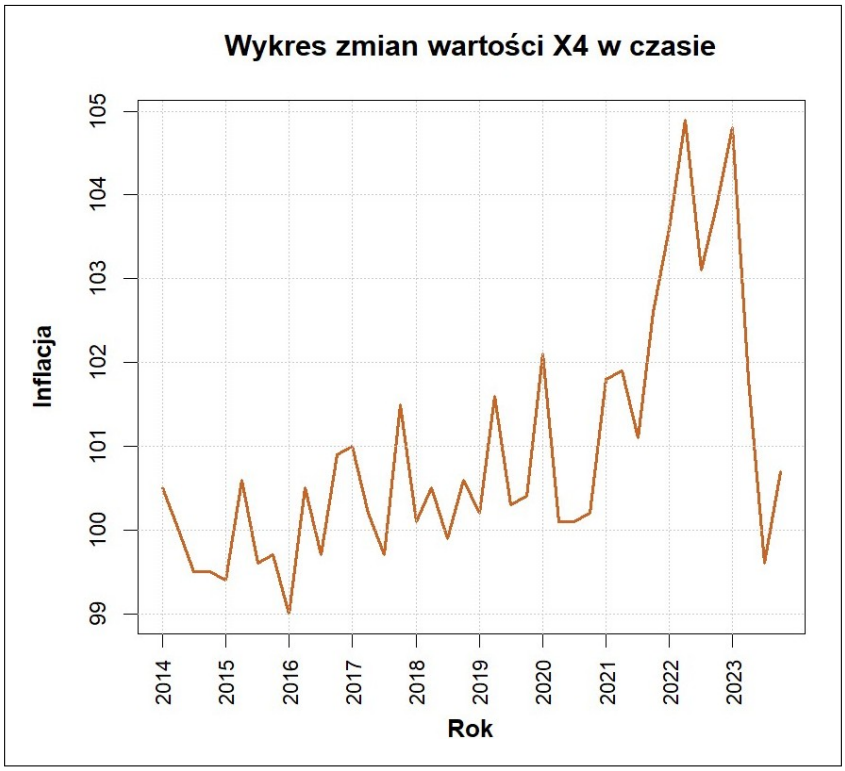
Rysunek 1: Wykres zależności zmiennej X_1 od czasu



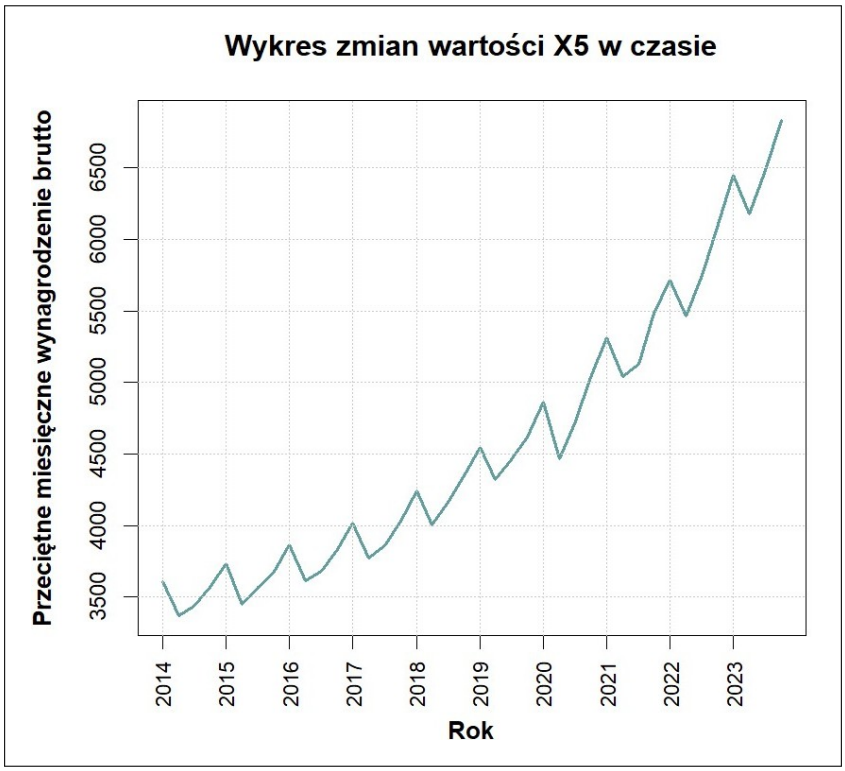
Rysunek 2: Wykres zależności zmiennej X_2 od czasu



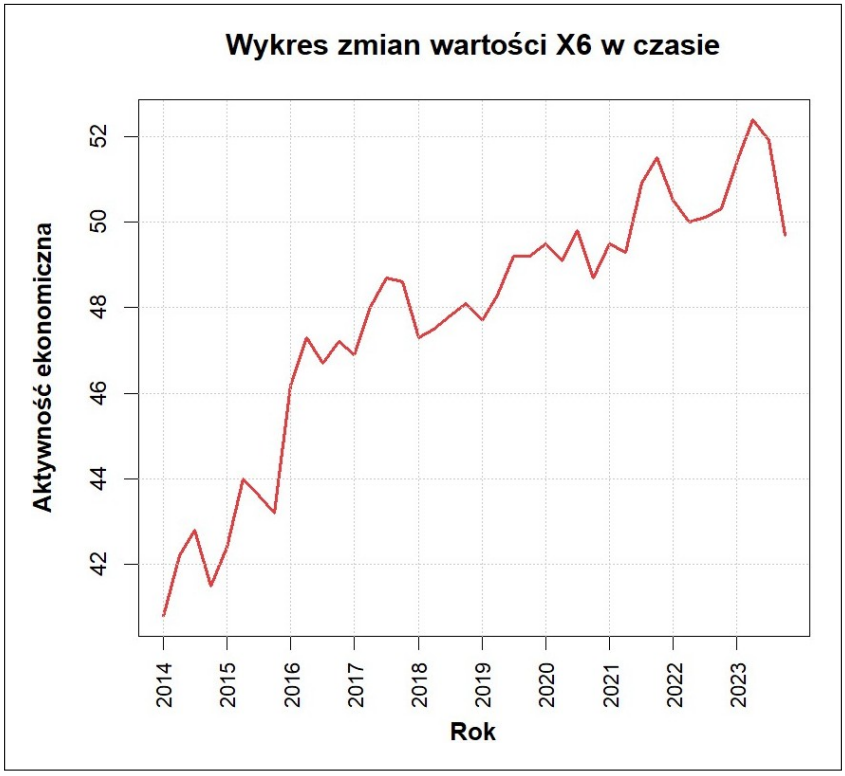
Rysunek 3: Wykres zależności zmiennej X_3 od czasu



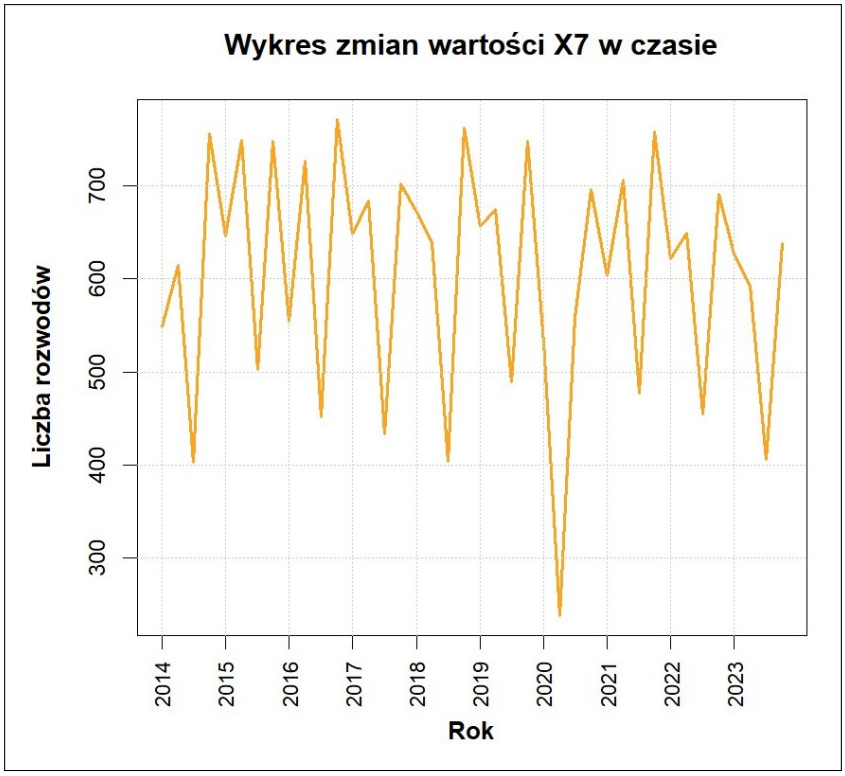
Rysunek 4: Wykres zależności zmiennej X_4 od czasu



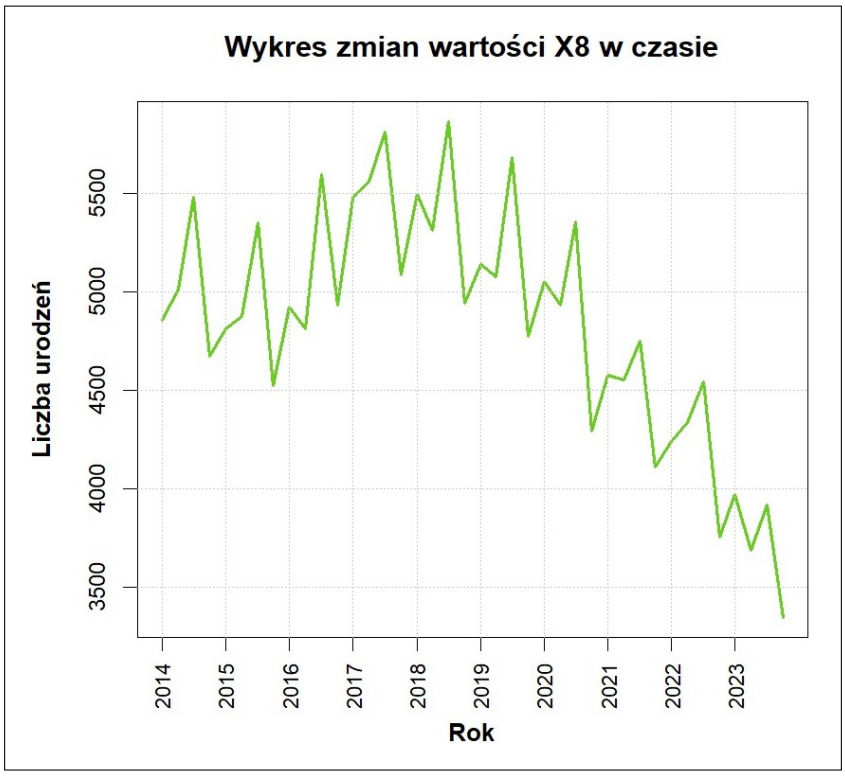
Rysunek 5: Wykres zależności zmiennej X_5 od czasu



Rysunek 6: Wykres zależności zmiennej X_6 od czasu

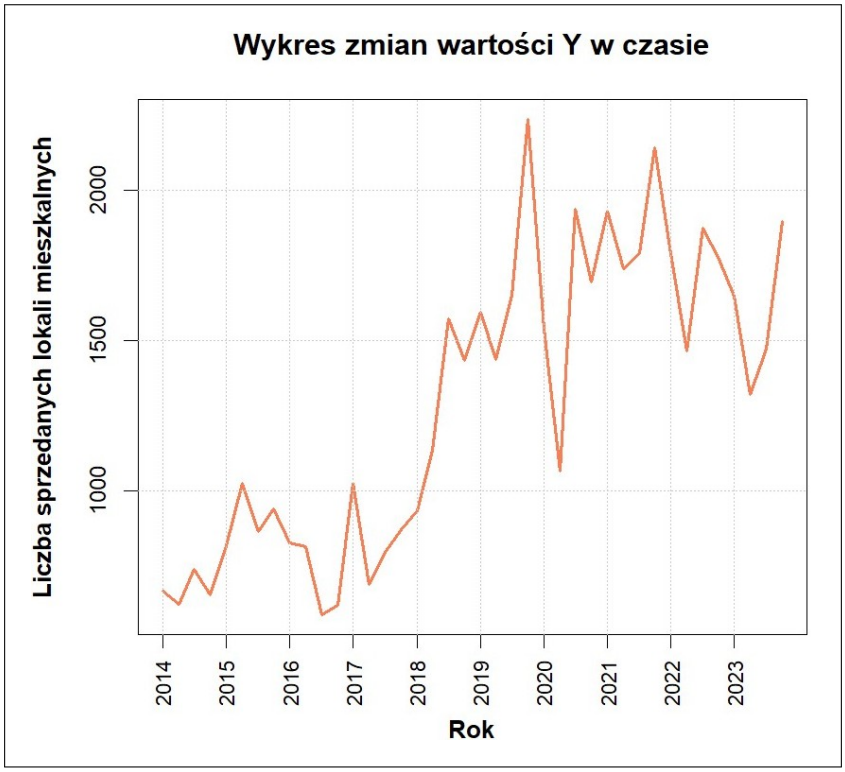


Rysunek 7: Wykres zależności zmiennej X_7 od czasu



Rysunek 8: Wykres zależności zmiennej X_8 od czasu

1.6 Wykres zależności zmiennej objaśnianej od czasu



Rysunek 9: Wykres zależności zmiennej objaśnianej Y od czasu

2 Arkusz roboczy

2.1 Analiza doboru zmiennych

Metoda eliminacji zmiennych Quasi-stałych

Współczynnik zmienności (V) mierzy względne zróżnicowanie danej zmiennej, obliczając stosunek odchylenia standardowego (S) do średniej arytmetycznej ($\bar{X}$). Jest to narzędzie pozwalające na eliminację zmiennych o zbyt małej różnorodności, które nie wnoszą istotnych informacji do analizy. Wzór na V przedstawia się następująco:

$$V = \frac{S}{\bar{X}} \cdot 100\%. \tag{1}$$

Przyjmujemy wartość krytyczną 20%, eliminując zmienne, dla których współczynnik zmienności jest niższy niż ta granica.

	X1	X2	X3	X4	X5	X6	X7	X8
Odchylenie standardowe	2,39	509,63	1 047,23	1,45	956,63	2,98	124,71	596,62
Średnia arytmetyczna	10,37	2 171,53	4 305,75	100,93	4 570,44	47,75	605,85	4 837,58
Współczynnik zmienności	23,05%	23,47%	24,32%	1,44%	20,93%	6,24%	20,58%	12,33%

Rysunek 10: Współczynnik zmienności V

Na podstawie otrzymanych wyników możemy określić, że w budowaniu modelu będą brały udział poniższe zmienne objaśniające:

- Stopa bezrobocia rejestrowanego
- Mieszkania oddane do użytkowania
- Średnia cena za 1 m^2 lokali mieszkalnych
- Przeciętne miesięczne wynagrodzenie brutto
- Liczba rozwodów

Podtrzymujemy dla nich odpowiednie oznaczenia, kolejno X_1, X_2, X_3, X_5 oraz X_7.

Metoda Hellwiga - metoda wskaźników pojemności informacyjnej

W przypadku gdy mamy m potencjalnych zmiennych objaśniających, to liczba wszystkich kombinacji jest równa:

$$L = 2^m - 1. \tag{2}$$

W naszym przypadku $m = 5$, a więc mamy $2^5 - 1 = 31$ kombinacji.

Dla wszystkich otrzymanych kombinacji definiujemy tzw. indywidualną pojemność nośników informacji, która określona jest wzorem:

$$h_{lj} = \frac{r_j^2}{1 + \sum_{i=1, i \neq j}^{m_l} |r_{ij}|}, \quad j = 1, 2, \ldots, m_l, \quad h_{lj} \in [0, 1], \tag{3}$$

gdzie:

- l – numer kombinacji ($l = 1, 2, \ldots, 2^m - 1$),
- m_l – liczba zmiennych w l-tej kombinacji,
- j – numer zmiennej w rozpatrywanej kombinacji,
- r_j – współczynnik korelacji potencjalnej zmiennej objaśniającej o numerze j ze zmienną objaśnianą (element wektora R_0),
- r_{ij} – współczynnik korelacji między i-tą i j-tą potencjalną zmienną objaśniającą (element macierzy R).

Wskaźnik h_{lj} mierzy ilość informacji, jaką wnosi zmienna X_j o zmiennej objaśnianej Y w l-tej kombinacji. W związku z tym h_{lj} wzrasta, jeżeli współczynnik korelacji r_j wzrasta, a maleje im bardziej zmienna X_j jest skorelowana z pozostałymi zmiennymi objaśniającymi.

Dopiero, gdy policzymy indywidualną pojemność nośników informacji dla wszystkich kombinacji, możemy obliczyć pojemność integralną kombinacji nośników informacji według wzoru:

$$H_l = \sum_{j=1}^{m_l} h_{ij}, \quad l = 1, 2, \ldots, 2^m - 1, \quad H_l \in [0, 1], \tag{4}$$

gdzie:

- l – numer kombinacji ($l = 1, 2, \ldots, 2^m - 1$),
- m_l – liczba zmiennych w l-tej kombinacji,
- j – numer zmiennej w rozpatrywanej kombinacji.

Pojemność integralna kombinacji nośników informacji dla l-tej kombinacji jest sumą indywidualnych pojemności nośników informacji, które wchodzą w skład tej kombinacji. Jest ona kryterium wyboru odpowiedniej kombinacji zmiennych objaśniających. Z reguły wybieramy tę kombinację, gdzie H_l jest największe.

2.2 Wybór postaci modelu

Zastosujemy metodę Hellwiga, którą dokładnie opisaliśmy wcześniej, aby dobrać odpowiednie zmienne objaśniające do modelu ekonometrycznego. Zaczynamy od zaprezentowania macierzy korelacji pomiędzy zmiennymi objaśniającymi i zmienną objaśnianą, a także wektora korelacji każdej zmiennej objaśniającej z Y. Wykonamy to w języku R.

```
            [,1]       [,2]        [,3]        [,4]        [,5]
[1,]  1.00000000 -0.5005087 -0.64135226 -0.65614332  0.07350073
[2,] -0.50050871  1.0000000  0.48457591  0.50239517  0.06129660
[3,] -0.64135226  0.4845759  1.00000000  0.96629862 -0.08407002
[4,] -0.65614332  0.5023952  0.96629862  1.00000000 -0.05921334
[5,]  0.07350073  0.0612966 -0.08407002 -0.05921334  1.00000000
```

Rysunek 11: Macierz korelacji R między zmiennymi objaśniającymi

```
[1] -0.74261174  0.60927873  0.81196612  0.74958931  0.05125525
```

Rysunek 12: Wektor korelacji R_0 każdej zmiennej objaśniającej z Y

Następnie przechodzimy do wyliczenia indywidualnej pojemności nośników informacji, aby końcowo uzyskać pojemność integralną kombinacji. Zaprezentujemy dwa podejścia. Pierwsze w Microsoft Excel, a drugie w języku R.

Lp.	X1	X2	X3	X5	X7	H			H	
1	0,551472	0	0	0	0	0,551472192	X1		0,754596481	X1X2X3
2	0	0,371221	0	0	0	0,371220574	X2		0,73935123	X1X2X3X5
3	0	0	0,659289	0	0	0,659288986	X3		0,737660775	X1X3
4	0	0	0	0,561884	0	0,561884139	X5		0,73081084	X1X2X3X7
5	0	0	0	0	0,002627	0,0026271	X7		0,723785616	X1X2X3X5X7
6	0,367523	0,247396	0	0	0	0,614919966	X1X2		0,707120483	X1X3X5
7	0,335986	0	0,401674	0	0	0,737660775	X1X3		0,705958166	X1X3X7
8	0,332986	0	0	0,339273	0	0,672258444	X1X5		0,701356403	X1X2X5
9	0,513714	0	0	0	0,002447	0,516161074	X1X7		0,694144066	X2X3
10	0	0,250052	0,444092	0	0	0,694144066	X2X3		0,68921096	X1X3X5X7
11	0	0,247086	0	0,373992	0	0,621078084	X2X5		0,683432712	X2X3X5
12	0	0,34978	0	0	0,002475	0,352255604	X2X7		0,682675236	X1X2X5X7
13	0	0	0,335294	0,285757	0	0,621051713	X3X5		0,672258444	X1X5
14	0	0	0,608161	0	0,002423	0,610584257	X3X7		0,665769985	X2X3X5X7
15	0	0	0	0,530473	0,00248	0,532953293	X5X7		0,662722057	X2X3X7
16	0,257473	0,187005	0,310118	0	0	0,754596481	X1X2X3		0,659288986	X3
17	0,255708	0,185341	0	0,260308	0	0,701356403	X1X2X5		0,648716064	X1X5X7
18	0,350361	0,237687	0	0	0,002315	0,590363307	X1X2X7		0,621078084	X2X5
19	0,240032	0	0,252829	0,21426	0	0,707120483	X1X3X5		0,621051713	X3X5
20	0,321586	0	0,382103	0	0,002269	0,705958166	X1X3X7		0,614919966	X1X2
21	0,318836	0	0	0,327561	0,002319	0,648716064	X1X5X7		0,610584257	X3X7
22	0	0,186827	0,269002	0,227604	0	0,683432712	X2X3X5		0,601247934	X3X5X7
23	0	0,240137	0,420292	0	0,002294	0,662722057	X2X3X7		0,599555792	X2X5X7
24	0	0,2374	0	0,359811	0,002345	0,599555792	X2X5X7		0,590363307	X1X2X7
25	0	0	0,321547	0,277404	0,002298	0,601247934	X3X5X7		0,561884139	X5
26	0,197095	0,149236	0,213208	0,179812	0	0,73935123	X1X2X3X5		0,551472192	X1
27	0,248931	0,181403	0,298321	0	0,002155	0,73081084	X1X2X3X7		0,532953293	X5X7
28	0,24728	0,179837	0	0,253358	0,0022	0,682675236	X1X2X5X7		0,516161074	X1X7
29	0,232591	0	0,244932	0,209529	0,002159	0,68921096	X1X3X5X7		0,371220574	X2
30	0	0,181236	0,26008	0,222272	0,002181	0,665769985	X2X3X5X7		0,352255604	X2X7
31	0,19205	0,145647	0,207565	0,176468	0,002056	0,723785616	X1X2X3X5X7		0,0026271	X7

Rysunek 13: Podejście nr 1 - Microsoft Excel

```
             X1        X2        X3        X5          X7         H
 [1,] 0.2574734 0.1870049 0.3101182 0.0000000 0.000000000 0.7545965
 [2,] 0.1970948 0.1492356 0.2132085 0.1798123 0.000000000 0.7393512
 [3,] 0.3359865 0.0000000 0.4016743 0.0000000 0.000000000 0.7376608
 [4,] 0.2489310 0.1814034 0.2983211 0.0000000 0.002155362 0.7308108
 [5,] 0.1920499 0.1456466 0.2075653 0.1764684 0.002055504 0.7237856
 [6,] 0.2400319 0.0000000 0.2528287 0.2142599 0.000000000 0.7071205
 [7,] 0.3215857 0.0000000 0.3821030 0.0000000 0.002269494 0.7059582
 [8,] 0.2557075 0.1853412 0.0000000 0.2603077 0.000000000 0.7013564
 [9,] 0.0000000 0.2500516 0.4440925 0.0000000 0.000000000 0.6941441
[10,] 0.2325909 0.0000000 0.2449322 0.2095288 0.002159052 0.6892110
[11,] 0.0000000 0.1868274 0.2690015 0.2276038 0.000000000 0.6834327
[12,] 0.2472800 0.1798375 0.0000000 0.2533575 0.002200232 0.6826752
[13,] 0.3329858 0.0000000 0.0000000 0.3392727 0.000000000 0.6722584
[14,] 0.0000000 0.1812364 0.2600802 0.2222725 0.002180926 0.6657700
[15,] 0.0000000 0.2401366 0.4202918 0.0000000 0.002293676 0.6627221
[16,] 0.0000000 0.0000000 0.6592890 0.0000000 0.000000000 0.6592890
[17,] 0.3188357 0.0000000 0.0000000 0.3275611 0.002319297 0.6487161
[18,] 0.0000000 0.2470858 0.0000000 0.3739922 0.000000000 0.6210781
[19,] 0.0000000 0.0000000 0.3352944 0.2857573 0.000000000 0.6210517
[20,] 0.3675235 0.2473965 0.0000000 0.0000000 0.000000000 0.6149200
[21,] 0.0000000 0.0000000 0.6081609 0.0000000 0.002423368 0.6105843
[22,] 0.0000000 0.0000000 0.3215466 0.2774035 0.002297856 0.6012479
[23,] 0.0000000 0.2374001 0.0000000 0.3598111 0.002344558 0.5995558
[24,] 0.3503614 0.2376868 0.0000000 0.0000000 0.002315039 0.5903633
[25,] 0.0000000 0.0000000 0.0000000 0.5618841 0.000000000 0.5618841
[26,] 0.5514722 0.0000000 0.0000000 0.0000000 0.000000000 0.5514722
[27,] 0.0000000 0.0000000 0.0000000 0.5304731 0.002480237 0.5329533
[28,] 0.5137138 0.0000000 0.0000000 0.0000000 0.002447227 0.5161611
[29,] 0.0000000 0.3712206 0.0000000 0.0000000 0.000000000 0.3712206
[30,] 0.0000000 0.3497802 0.0000000 0.0000000 0.002475369 0.3522556
[31,] 0.0000000 0.0000000 0.0000000 0.0000000 0.002627100 0.0026271
```

Rysunek 14: Podejście nr 2 - język R

Po zastosowaniu metody eliminacji zmiennych Quasi-stałych wiemy, że nasz model może potencjalnie składać się ze zmiennych X_1, X_2, X_3, X_5 albo X_7. Metoda Hellwiga pozwala nam wybrać podzbiór tych zmiennych. Widzimy, że wyniki z obu podejść się pokrywają i są one dodatkowo posortowane w porządku malejącym. Wartości trzech pierwszych kombinacji różnią się od siebie o mniej niż 0.02. Z uwagi na ten fakt, wybierzemy kombinację z trzeciego wiersza, ponieważ zależy ona od mniejszej liczby zmiennych niż kombinacje z pierwszego i drugiego.

Na podstawie wybranych zmiennych model ekonometryczny wygląda następująco:

$$\hat{Y} = \alpha_0 + \alpha_1 X_1 + \alpha_3 X_3. \tag{5}$$

2.3 Estymacja parametrów

Estymacja parametrów strukturalnych modelu polega na wyznaczeniu wartości współczynników $(\alpha_0, \alpha_1, \alpha_3)$ w równaniu modelu ekonometrycznego, które najlepiej oddają zależności w analizowanym przypadku między zmienną objaśnianą (Y) a wybranymi zmiennymi objaśniającymi (X_1, X_3) na podstawie dostępnych danych. W tym celu wykorzystamy narzędzie do analizy danych w programie Microsoft Excel, a dokładniej funkcję regresji.

	Współczynniki
Przecięcie	945,9762487
X1	-76,46645978
X3	0,264066873

Rysunek 15: Estymacja parametrów

Podstawiając wartości wyznaczonych współczynników możemy oszacować równanie opisujące nasz model w następującej postaci:

$$\hat{Y} = 945.976248724304 - 76.4664597809042X_1 + 0.264066873182064X_3. \tag{6}$$

Warto w tym miejscu zdefiniować także czym jest błąd oszacowania parametrów. Znajdziemy go w naszej funkcji regresji pod nazwą błąd standardowy. Mierzy on rozrzut oszacowanych wartości parametrów w modelu regresji wokół ich rzeczywistych wartości. Jest to miara precyzji oszacowania, gdzie mniejszy błąd standardowy wskazuje na bardziej wiarygodne oszacowanie parametrów modelu. Poniżej przedstawiamy błędy dla naszych parametrów.

	Błąd standardowy
Przecięcie	405,3806214
X1	22,04372322
X3	0,050310545

Rysunek 16: Błędy oszacowań parametrów

Dzięki przeprowadzonej estymacji uzyskujemy konkretny model, który może posłużyć do dalszych analiz oraz prognoz. Proces ten stanowi fundament budowy każdego modelu regresyjnego, ponieważ pozwala przełożyć dane empiryczne na formalny opis badanych zależności.

3 Weryfikacja modelu

3.1 Zbadanie własności koincydencji

Model jest koincydentny, gdy wszystkie znaki stojące w wektorze R_0 przy zmiennych objaśniających X_i, czyli znaki współczynników r_i, są zgodne ze znakami stojącymi przy tych zmiennych w oszacowanych parametrach α_i, co opisuje poniższa zależność:

$$\forall i = 1, 2, \ldots, k \quad \operatorname{sgn}(r_i) = \operatorname{sgn}(\alpha_i), \tag{7}$$

gdzie k to liczba zmiennych objaśniających.

Widzimy, że w naszym przypadku model jest koincydentny, co wskazuje na to, że znaki interesujących nas zmiennych objaśniających zgadzają się ze znakami parametrów stojących przy tych zmiennych.

3.2 Test t-Studenta istotności parametrów

Test t-Studenta służy do oceny, czy zmienne objaśniające w modelu ekonometrycznym wywierają istotny wpływ na zmienną objaśnianą. Narzędzie to pomaga również zweryfikować prawidłowość doboru zmiennych objaśniających oraz sprawdzić, czy wyraz wolny różni się statystycznie od zera.

Jest to kluczowy etap selekcji zmiennych *a posteriori*. Proces postępowania wygląda następująco:

1. Jeśli wszystkie parametry są statystycznie istotne, model uznaje się za poprawny.
2. W przypadku, gdy którykolwiek parametr okazuje się nieistotny, należy dokonać modyfikacji modelu poprzez eliminację takich zmiennych.
3. Proces usuwania odbywa się iteracyjnie, zaczynając od zmiennej o największym poziomie istotności (najwyższym *p-value*).
4. Wyraz wolny jest zawsze zachowywany w modelu i nie podlega eliminacji.

Formułujemy następujące hipotezy dla $\alpha = 0.05$ dla zmiennych X_i:

$H_0: \quad \alpha_i = 0 \quad$ (zmienna X_i jest nieistotna dla rozpatrywanego modelu),
$H_1: \quad \alpha_i \neq 0 \quad$ (zmienna X_i ma statystycznie istotny wpływ na zmienną objaśnianą).

Obliczamy statystykę:

$$t_i^* = \frac{|a_i|}{S(\alpha_i)}, \tag{8}$$

która ma rozkład z $n - (k + 1)$ stopniami swobody. Wiemy, że α_i to parametr stojący przy zmiennej objaśniającej X_i oraz, że $S(\alpha_i)$ to błąd standardowy, określający precyzję oszacowania tego parametru. Wykorzystamy do obliczeń program Microsoft Excel, który automatycznie wykonuje obliczenia za pomocą funkcji regresji. Należy jednak zauważyć drobną nieścisłość — Excel przyjmuje wartość α_i we wzorze (8) bez uwzględnienia znaku minus. Mimo to, w naszej dalszej analizie znak minus zostanie oczywiście uwzględniony.

	t Stat
Przecięcie	2,333550739
X1	-3,468854105
X3	5,248738111

Rysunek 17: Wyznaczenie statystyki t_i^*

Ze względu na dużą liczebność próby ($n = 40$), wartość krytyczna t_{kr} nie została odczytana z tradycyjnych tablic rozkładu t-Studenta, lecz została wyznaczona z wykorzystaniem języka R, co pozwala na precyzyjne obliczenie wartości dla dowolnych stopni swobody.

$$t_{kr} = t_{\frac{\alpha}{2};n-k-1} = t_{\frac{0.05}{2};40-2-1} = t_{0.025;37} = 2.0262.$$

Wynika z tego, że $t_1^*, t_2^* > t_{kr}$, co prowadzi do odrzucenia hipotezy H_0 na rzecz hipotezy alternatywnej H_1. Obie zmienne są istotne statystycznie, więc możemy ocenić model pozytywnie.

3.3 Współczynnik determinacji

Współczynnik determinacji R^2 jest jedną z miar jakości dopasowania modelu do danych uczących. Określa on, jak dobrze zmienna objaśniana jest wytłumaczona przez zmienne objaśniające. Współczynnik determinacji przyjmuje wartości z przedziału $[0; 1]$, jeśli w modelu występuje wyraz wolny, a do estymacji parametrów wykorzystano metodę najmniejszych kwadratów. Zdefiniowany jest jako:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \geq 0, \tag{9}$$

gdzie

- y_i – i-ta obserwacja zmiennej y,
- $\hat{y}_i$ – wartość teoretyczna zmiennej objaśnianej (na podstawie modelu),
- $\overline{y}$ – średnia arytmetyczna empirycznych wartości zmiennej objaśnianej.

Wartość współczynnika R^2 dla naszego modelu wynosi w przybliżeniu 0.74.
Oznacza to, że model można ocenić pozytywnie, ponieważ dobrze dopasowuje się do danych, a jego predykcje są stosunkowo dokładne, co sprawia, że jest użyteczny w przewidywaniu ogólnych trendów zmiennej objaśnianej.

3.4 Współczynnik zbieżności

Współczynnik zbieżności φ^2 określa, jaka część zaobserwowanego w próbie zróżnicowania zmiennej objaśnianej nie pasuje do modelu (mieści się w jego błędzie). Również przyjmuje wartości z przedziału $[0; 1]$. Wyraża się on wzorem:

$$\varphi^2 = 1 - R^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \geq 0, \tag{10}$$

Statystyki regresji	
Wielokrotność R	0,861917222
R kwadrat	0,742901297
Dopasowany R kwadrat	0,72900407
Błąd standardowy	255,6619175
Obserwacje	40

Rysunek 18: Współczynnik determinacji R^2

gdzie y_i, $\hat{y}_i$ oraz $\overline{y}$ są określone jak w części poprzedniej.

Wartość współczynnika φ^2 dla naszego modelu wynosi około 0.26. Biorąc pod uwagę ten fakt, stwierdzamy, że nasz model jest dobrze dopasowany do danych, wyjaśniając znaczną część zmienności Y, podczas gdy pozostała część mieści się w błędzie modelu.

R kwadrat	0,742901297
1 - R kwadrat	0,257098703

Rysunek 19: Współczynnik zbieżności φ^2

3.5 Test istotności współczynnika determinacji

Gdy zachodzi potrzeba sprawdzenia, czy grupa wybranych zmiennych objaśniających ma istotny wpływ na zmienną objaśnianą, wykorzystuje się test F^*.

Formułujemy następujące hipotezy dla $\alpha = 0.05$:

$H_0: \quad \beta_1 = \beta_2 = 0$ (obie zmienne objaśniające są nieistotne statystycznie),

$H_1: \quad \beta_1 \neq 0 \vee \beta_2 \neq 0$ (istnieje zmienna objaśniająca, która jest istotna).

Obliczamy statystykę:

$$F^* = \frac{R^2}{1 - R^2} \frac{n - (k+1)}{k}, \tag{11}$$

gdzie:

- R^2 – współczynnik determinacji,
- n – liczba obserwacji,
- k – liczba zmiennych w modelu.

R kwadrat	0,742901297
1 - R kwadrat	0,257098703
n	40
k	2
n - k - 1	37
F*	53,45680022

Rysunek 20: Wyznaczenie statystyki F^*

Ze względu na dużą liczbę obserwacji wartość krytyczną F_{kr} obliczamy w języku R, korzystając z funkcji wyznaczającej wartości z rozkładu Fishera-Snedecora.

$$F_{kr} = F_{\alpha;k;n-(k+1)} = F_{0.05;2;40-(2+1)} = F_{0.05;2;37} = 3.252.$$

Nietrudno zauważyć, że $F^* > F_{kr}$. Należy zatem odrzucić hipotezę zerową H_0 na korzyść hipotezy alternatywnej H_1. Oznacza to, że istnieje taka zmienna objaśniająca (X_1 lub X_3), która ma statystycznie istotny wpływ na zmienną objaśnianą Y.

3.6 Test t dla korelacji

Test t dla współczynnika korelacji służy do oceny, czy zależność pomiędzy dwiema zmiennymi jest statystycznie istotna, a więc czy zaobserwowany współczynnik korelacji różni się istotnie od zera i nie wynika jedynie z przypadkowych wahań w próbie.

Formułujemy następujące hipotezy dla $\alpha = 0.05$:

$$\begin{aligned} H_0: &\quad \rho = 0 \quad \text{(brak korelacji między zmiennymi)}, \\ H_1: &\quad \rho \neq 0 \quad \text{(istnieje korelacja między zmiennymi)}. \end{aligned}$$

Wyliczamy statystykę testową dla testu istotności korelacji:

$$t^* = R \cdot \sqrt{\frac{n-k-1}{1-R^2}}, \tag{12}$$

gdzie n, k oraz R^2 określone są jak w poprzedniej części.

R	0,861917222
n - k - 1	37
1 - R kwadrat	0,257098703
t*	10,33990331

Rysunek 21: Wyznaczenie statystyki testowej t^*

Wyznaczamy wartość krytyczną t_{kr} za pomocą języka R:

$$t_{kr} = t_{\frac{\alpha}{2};n-k-1} = t_{\frac{0.05}{2};40-2-1} = t_{0.025;37} = 2.0262.$$

Zatem obszar krytyczny ma postać:

$$K = (-\infty; -2.0262] \cup [2.0262; +\infty).$$

Widzimy, że wartość statystyki testowej t^* należy do obszaru krytycznego. Odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1. Korelacja między zmiennymi istnieje, a do tego jest istotna.

3.7 Współczynnik wyrazistości

Współczynnik wyrazistości modelu jest definiowany jako:

$$V_e = \frac{S_e}{\overline{Y}}, \tag{13}$$

gdzie:

- S_e – odchylenie standardowe reszt modelu,
- $\overline{Y}$ – średnia arytmetyczna zmiennej objaśnianej Y.

Informuje on, jaki procent średniej wartości zmiennej objaśnianej stanowi odchylenie standardowe reszt modelu. Im mniejsza wartość V_e, tym lepsze dopasowanie modelu do danych empirycznych. Wartość współczynnika wyrazistości na poziomie około 20% wskazuje, iż przeciętne odchylenie reszt względem średniej zmiennej objaśnianej stanowi jedną piątą tej średniej. Wynik ten sugeruje umiarkowaną jakość dopasowania modelu do danych empirycznych – prognozy modelu są relatywnie bliskie wartościom obserwowanym, choć widoczny jest pewien poziom rozproszenia.

Błąd standardowy	255,6619175
Średnia arytmetyczna z Y	1 290
V_e	19,82%

Rysunek 22: Współczynnik wyrazistości V_e

3.8 Variance Inflation Factor

Variance Inflation Factor (VIF) to miara współliniowości między zmiennymi objaśniającymi w modelu regresji. VIF określa, w jakim stopniu wariancja estymatora parametru jest zawyżona z powodu współliniowości.

- VIF = 1: brak współliniowości (zmienna jest ortogonalna względem innych predyktorów),
- VIF $<$ 4: brak istotnej współliniowości,
- $4 \leq$ VIF ≤ 10: możliwość występowania współliniowości (należy zachować ostrożność),
- VIF $>$ 10: silna współliniowość, konieczna modyfikacja modelu.

Wysokie wartości VIF wskazują, że dana zmienna objaśniająca jest silnie skorelowana z innymi, co może negatywnie wpływać na stabilność i interpretację modelu. Jednym ze wzorów definiujących tę miarę jest następujący:

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}. \tag{14}$$

Poniżej przedstawiamy obliczone miary VIF_1 oraz VIF_2 uzyskane za pomocą Microsoft Excel oraz pakietu do analizy danych, korzystającego z funkcji regresji.

PODSUMOWANIE - WYJŚCIE

Statystyki regresji	
Wielokrotność R	0,742611737
R kwadrat	0,551472192
Dopasowany R kwadrat	0,539668829
Błąd standardowy	333,2113988
Obserwacje	40

ANALIZA WARIANCJI

	df	*SS*	*MS*	*F*	*Istotność F*
Regresja	1	5187493,196	5187493,196	46,72161439	4,11137E-08
Resztkowy	38	4219133,779	111029,8363		
Razem	39	9406626,975			

	Współczynniki	*Błąd standardowy*	*t Stat*	*Wartość-p*	*Dolne 95%*	*Górne 95%*	*Dolne 95,0%*	*Górne 95,0%*
Przecięcie	2852,493905	234,5804499	12,15998139	1,14087E-14	2377,610612	3327,377199	2377,610612	3327,377199
X1	-150,6720256	22,04315257	-6,835321089	4,11137E-08	-195,296055	-106,0479962	-195,296055	-106,0479962

R kwadrat	0,551472192
1 - R kwadrat	0,448527808
VIF_1	2,229516168

Rysunek 23: Współczynnik współliniowości VIF_1

PODSUMOWANIE - WYJŚCIE

Statystyki regresji	
Wielokrotność R	0,811966123
R kwadrat	0,659288986
Dopasowany R kwadrat	0,650322906
Błąd standardowy	290,4144688
Obserwacje	40

ANALIZA WARIANCJI

	df	*SS*	*MS*	*F*	*Istotność F*
Regresja	1	6201685,555	6201685,555	73,53146914	2,03958E-10
Resztkowy	38	3204941,42	84340,56367		
Razem	39	9406626,975			

	Współczynniki	*Błąd standardowy*	*t Stat*	*Wartość-p*	*Dolne 95%*	*Górne 95%*	*Dolne 95,0%*	*Górne 95,0%*
Przecięcie	-328,918185	194,3008092	-1,692829723	0,098672391	-722,2596091	64,42323913	-722,2596091	64,42323913
X3	0,37599563	0,043847635	8,575049221	2,03958E-10	0,287230733	0,464760527	0,287230733	0,464760527

R kwadrat	0,659288986
1 - R kwadrat	0,340711014
VIF_2	2,935038662

Rysunek 24: Współczynnik współliniowości VIF_2

Wartości VIF wynoszące około 2.23 lub 2.94 mieszczą się w zakresie od 1 do 4, co wskazuje na brak istotnej współliniowości między predyktorami w analizowanym modelu. Uzyskane wartości VIF wskazują, że wpływ ewentualnej współliniowości na wyniki modelu jest pomijalny, co podnosi wiarygodność oszacowanych parametrów.

4 Testy dotyczące składnika losowego

W tej sekcji zajmiemy się testami dotyczącymi składnika losowego. Składnik zakłócający jest zmienną losową i charakteryzuje się pewnym rozkładem prawdopodobieństwa. Cechy rozkładu składnika losowego są ważnym elementem modelu ekonometrycznego.

4.1 Test symetrii składnika losowego

Test symetrii sprawdza, czy odchylenia od wartości zmiennej uzyskanych za pomocą modelu są równomiernie rozłożone względem zera, co oznacza, że odsetek reszt ujemnych i dodatnich powinien wynosić około 50%. Oznacza to, że model ekonometryczny powinien być osią symetrii analizowanej zmienności, zgodnie z założeniem normalnego i symetrycznego rozkładu składników losowych.

Formułujemy następujące hipotezy dla $\alpha = 0.05$:

$$\begin{aligned} H_0: &\quad \frac{m}{n} = \frac{1}{2} \quad \text{(reszty losowe są symetryczne)}, \\ H_1: &\quad \frac{m}{n} \neq \frac{1}{2} \quad \text{(reszty losowe nie są symetryczne)}. \end{aligned}$$

Statystyka testowa ma postać:

$$t^* = \frac{\left|\frac{m}{n} - \frac{1}{2}\right|}{\sqrt{\frac{\frac{m}{n}\left(1-\frac{m}{n}\right)}{n-1}}}, \tag{15}$$

gdzie:

- m to liczba reszt dodatnich,
- n to liczebność próby.

m	21
n	40
iloraz m przez n	0,525
licznik	0,025
mianownik	0,079963934
t*	0,312640946

Rysunek 25: Wyznaczenie statystyki testowej t^*

Statystyka ta, przy prawdziwości hipotezy zerowej, ma rozkład t-Studenta o $(n-1)$ stopniach swobody. Wartości krytycznej nie odczytujemy z tablic, lecz obliczamy przy użyciu języka R:

$$t_{kr} = t_{\frac{\alpha}{2};n-1} = t_{\frac{0.05}{2};40-1} = t_{0.025;39} = 2.0227.$$

Wynika z tego, że $t_{kr} > t^*$. W związku z tym nie mamy podstaw do odrzucenia hipotezy zerowej H_0, która zakłada, że reszty losowe są symetryczne. Model oceniamy pozytywnie.

4.2 Test serii składnika losowego

Test serii służy do sprawdzania niezależności reszt modelu, czyli weryfikacji, czy wartości reszt nie wykazują jakiejkolwiek struktury, która mogłaby sugerować autokorelację lub inne systematyczne wzorce. Jest to istotne w analizie reszt, ponieważ zakłada się, że reszty powinny być losowe, niezależne i rozkładać się w sposób przypadkowy wokół zera.

Formułujemy następujące hipotezy dla $\alpha = 0.05$:

$$H_0 : \text{składniki resztowe mają charakter losowy,}$$
$$H_1 : \text{składniki resztowe nie mają charakteru losowego.}$$

Wyznaczamy liczbę serii oznaczoną jako r, a także liczbę symboli „+” w ciągu reszt, którą oznaczamy n_1, oraz liczbę symboli „−”, oznaczaną jako n_2. Możemy do tego podejść na dwa sposoby. Poniżej przedstawiamy oba z nich.

Obserwacja	Y	Składniki resztowe	Numer serii
11	589	-235,4838873	1
12	618	-262,9829048	1
2	624	-15,44418797	1
4	654	-43,26709451	1
1	667	152,400187	2
14	689	-330,5651011	3
3	739	23,84767283	4
15	795	-254,1942683	5
10	813	-67,88049939	5
5	818	155,0651537	6
9	826	34,36991971	6
7	864	-1,2256845	7
16	871	-236,2663937	7
17	935	-187,951534	7
8	939	4,687645469	8
13	1 024	45,37237351	8
6	1 026	142,1279729	8
26	1 066	-385,7946098	9
18	1 135	-126,0394766	9
38	1 321	-489,9742854	9
20	1 433	102,7100319	10
22	1 437	-6,518189994	11
34	1 464	-179,4308957	11
39	1 471	-535,8067105	11
25	1 537	50,33371492	12
19	1 572	251,9260319	12
21	1 594	214,7411722	12
37	1 648	-176,2340958	13
23	1 649	160,0961879	14
28	1 694	144,7534206	14
30	1 737	31,4484184	14
36	1 768	-42,54005304	15
33	1 787	188,7881157	16
31	1 788	98,17238804	16
35	1 873	64,46233786	16
40	1 895	-161,7266429	17
29	1 931	321,4183921	18
27	1 935	388,6581562	18
32	2 139	491,9999495	18
24	2 236	671,9472731	18

Rysunek 26: Test serii - sposób nr 1

Obserwacja	Składniki resztowe	Numer serii
1	152,400187	1
2	-15,44418797	2
3	23,84767283	3
4	-43,26709451	4
5	155,0651537	5
6	142,1279729	5
7	-1,2256845	6
8	4,687645469	7
9	34,36991971	7
10	-67,88049939	8
11	-235,4838873	8
12	-262,9829048	8
13	45,37237351	9
14	-330,5651011	10
15	-254,1942683	10
16	-236,2663937	10
17	-187,951534	10
18	-126,0394766	10
19	251,9260319	11
20	102,7100319	11
21	214,7411722	11
22	-6,518189994	12
23	160,0961879	13
24	671,9472731	13
25	50,33371492	13
26	-385,7946098	14
27	388,6581562	15
28	144,7534206	15
29	321,4183921	15
30	31,4484184	15
31	98,17238804	15
32	491,9999495	15
33	188,7881157	15
34	-179,4308957	16
35	64,46233786	17
36	-42,54005304	18
37	-176,2340958	18
38	-489,9742854	18
39	-535,8067105	18
40	-161,7266429	18

Rysunek 27: Test serii - sposób nr 2

Pierwsza metoda polega na uporządkowaniu składników resztowych według wartości zmiennej objaśnianej Y w porządku rosnącym. Druga metoda pomija użycie zmiennej Y, a sortowanie reszt odbywa się wyłącznie na podstawie kolejności obserwacji.

Obszar krytyczny można zapisać w postaci:

$$K = (-\infty; k_1] \cup [k_2; +\infty),$$

gdzie:

- k_1 to wartość odczytana z tablicy dla poziomu ufności $\frac{\alpha}{2}$ oraz liczb n_1 i n_2,
- k_2 to wartość odczytana z tablicy dla poziomu ufności $1 - \frac{\alpha}{2}$ oraz liczb n_1 i n_2.

Na podstawie powyższych zrzutów ekranu możemy ustalić, że $n_1 = 21$ oraz $n_2 = 19$. Z tablic rozkładu serii odczytujemy, że dla zadanych wartości n_1, n_2 oraz poziomu ufności $\alpha = 0.05$, dolna granica wynosi 14, a górna granica 28. Niezależnie od przyjętej metody, liczba serii w obu przypadkach wynosi 18, co oznacza, że mieści się w tym przedziale. W związku z tym nie ma podstaw do odrzucenia hipotezy zerowej H_0. Składniki resztowe mają charakter losowy, więc możemy ocenić model pozytywnie.

4.3 Badanie normalności rozkładu składnika losowego

Test Shapiro-Wilka to statystyczny test normalności, który sprawdza, czy dane pochodzą z rozkładu normalnego. Jest szczególnie przydatny w przypadku małych próbek, ponieważ jest bardziej czuły na odchylenia od normalności w porównaniu do innych testów. Statystyka testowa W może przyjmować wartości od 0 do 1, gdzie W bliskie 1 sugeruje, że dane są zgodne z rozkładem normalnym, a W mniejsze niż 1 wskazuje na odchylenie od normalności.

Formułujemy następujące hipotezy dla $\alpha = 0.05$:

$$H_0 : \text{składniki resztowe mają rozkład normalny,}$$
$$H_1 : \text{składniki resztowe nie mają rozkładu normalnego.}$$

Funkcję testową definiujemy jako:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{16}$$

gdzie:

- $x_{(i)}$ to i-ta najmniejsza obserwacja w posortowanej nierosnąco próbce (tzn. $x_1 \leq x_2 \leq \cdots \leq x_n$),
- a_i to współczynniki, które zależą od liczby próbek n i są obliczane na podstawie rozkładu normalnego,
- x_i to obserwacja próbki,
- $\bar{x}$ to średnia próbki.

```
        Shapiro-Wilk normality test

data:  reszty
W = 0.98331, p-value = 0.8092
```

Rysunek 28: Wyznaczenie statystyki testowej W

Odczytujemy wartość krytyczną W_{kr} z tablic rozkładu Shapiro-Wilka:

$$W_{kr} = W_{\alpha;n} = W_{0.05;40} = 0.94.$$

Na podstawie powyższych wartości wynika, że $W > W_{kr}$. Zatem nie mamy podstaw do odrzucenia hipotezy zerowej H_0. Reszty modelu z X_1 i X_3 są zgodne z założeniem o normalności, więc model oceniamy pozytywnie.

4.4 Badanie heteroskedastyczności składnika losowego

Test Breuscha-Pagana jest używany do sprawdzenia, czy w modelu regresji występuje heteroskedastyczność, czyli czy wariancja składnika losowego jest stała w całym zakresie danych. Jego celem jest weryfikacja, czy zmienność reszt modelu zmienia się w zależności od wartości zmiennych objaśniających, co może wskazywać na problemy z poprawnością założeń modelu regresji.

Formułujemy następujące hipotezy dla $\alpha = 0.05$:

H_0 : homoskedastyczność (wariancja składnika losowego jest stała),

H_1 : heteroskedastyczność (wariancja składnika losowego jest zmienna).

Zbadamy hipotezę o heteroskedastyczności za pomocą języka R.

```
        studentized Breusch-Pagan test

data:  model
BP = 5.5553, df = 2, p-value = 0.06218
```

Rysunek 29: Wyznaczenie statystyki testowej BP

Wartość *p-value* w ostatnim wierszu jest wyższa niż przyjęty poziom istotności $\alpha = 0.05$. Z tego powodu nie mamy podstaw do odrzucenia hipotezy zerowej H_0. Oznacza to, że nie wykazano istotnych dowodów na występowanie heteroskedastyczności w badanym modelu.

4.5 Badanie autokorelacji składnika losowego

Test Durbina-Watsona służy do wykrywania autokorelacji w resztach modeli regresyjnych, czyli zależności między kolejnymi wartościami reszt. Jest wykorzystywany głównie do oceny, czy założenie niezależności reszt w modelu liniowym jest spełnione, co ma kluczowe znaczenie dla poprawności wniosków statystycznych.

Formułujemy następujące hipotezy dla $\alpha = 0.05$:

$H_0 : \quad \rho = 0 \quad$ (brak autokorelacji składników losowych),

$H_1 : \quad \rho > 0 \quad$ (istnieje dodatnia autokorelacja składników losowych).

Jako statystykę testową przyjmujemy statystykę Durbin-Watsona, którą definiujemy jako:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \in [0, 4], \tag{17}$$

gdzie:

- n – liczba obserwacji w zbiorze danych,
- t – numer obserwacji (od 1 do n),
- e_t – reszta modelu dla obserwacji t,
- e_{t-1} – reszta modelu dla poprzedniej obserwacji.

```
        Durbin-Watson test

data:  model
DW = 1.2269, p-value = 0.001742
alternative hypothesis: true autocorrelation is greater than 0
```

Rysunek 30: Wyznaczenie statystyki testowej DW

Z tablic testu Durbin-Watsona odczytujemy dwie wartości krytyczne: d_l (dolną granicę) i d_u (górną granicę).

Reguła decyzyjna testu Durbin-Watsona:

- Jeśli $DW < d_l$, odrzucamy hipotezę zerową H_0 i stwierdzamy obecność autokorelacji składników losowych.
- Jeśli $DW > d_u$, brak podstaw do odrzucenia hipotezy zerowej H_0, czyli nie występuje autokorelacja składników losowych.
- Jeśli $d_l < DW < d_u$, test nie dostarcza jednoznacznej odpowiedzi, czy w modelu występuje autokorelacja składników losowych.

Z tablic Durbin-Watsona dla $k = 2$ i $n = 40$, gdzie k oznacza liczbę zmiennych objaśniających w naszym modelu, odczytujemy wartości $d_l = 1.39083$ oraz $d_u = 1.59999$. Oznacza to, że $DW < d_l$, co prowadzi do wniosku, że hipoteza zerowa o braku autokorelacji zostaje odrzucona, a w modelu występuje istotna dodatnia autokorelacja składnika losowego.

Krótkie podsumowanie dotychczasowych działań

W przeprowadzonej analizie skonstruowano model ekonometryczny opisujący zależność zmiennej objaśnianej (Y) od wybranych zmiennych objaśniających (X_1 oraz X_3). Dokonano weryfikacji koincydencji modelu oraz przetestowano istotność poszczególnych parametrów przy użyciu testu t-Studenta, co pozwoliło ocenić wpływ zmiennych na zmienną zależną. Przeanalizowano jakość dopasowania modelu za pomocą współczynnika determinacji R^2 oraz współczynnika zbieżności, a także zweryfikowano istotność całego modelu przy wykorzystaniu testu F^*.

W dalszej części pracy przeprowadzono szczegółowe badanie własności składnika losowego, analizując m.in. symetrię, losowość, normalność rozkładu, a także występowanie heteroskedastyczności i autokorelacji. Wyniki testów Shapiro-Wilka oraz Breuscha-Pagana nie wykazały naruszeń podstawowych założeń modelu, natomiast test Durbin-Watsona ujawnił istotną dodatnią autokorelację składnika losowego. Przeprowadzona analiza pozwoliła na całościową ocenę poprawności modelu oraz wskazała potencjalne obszary do dalszych usprawnień.

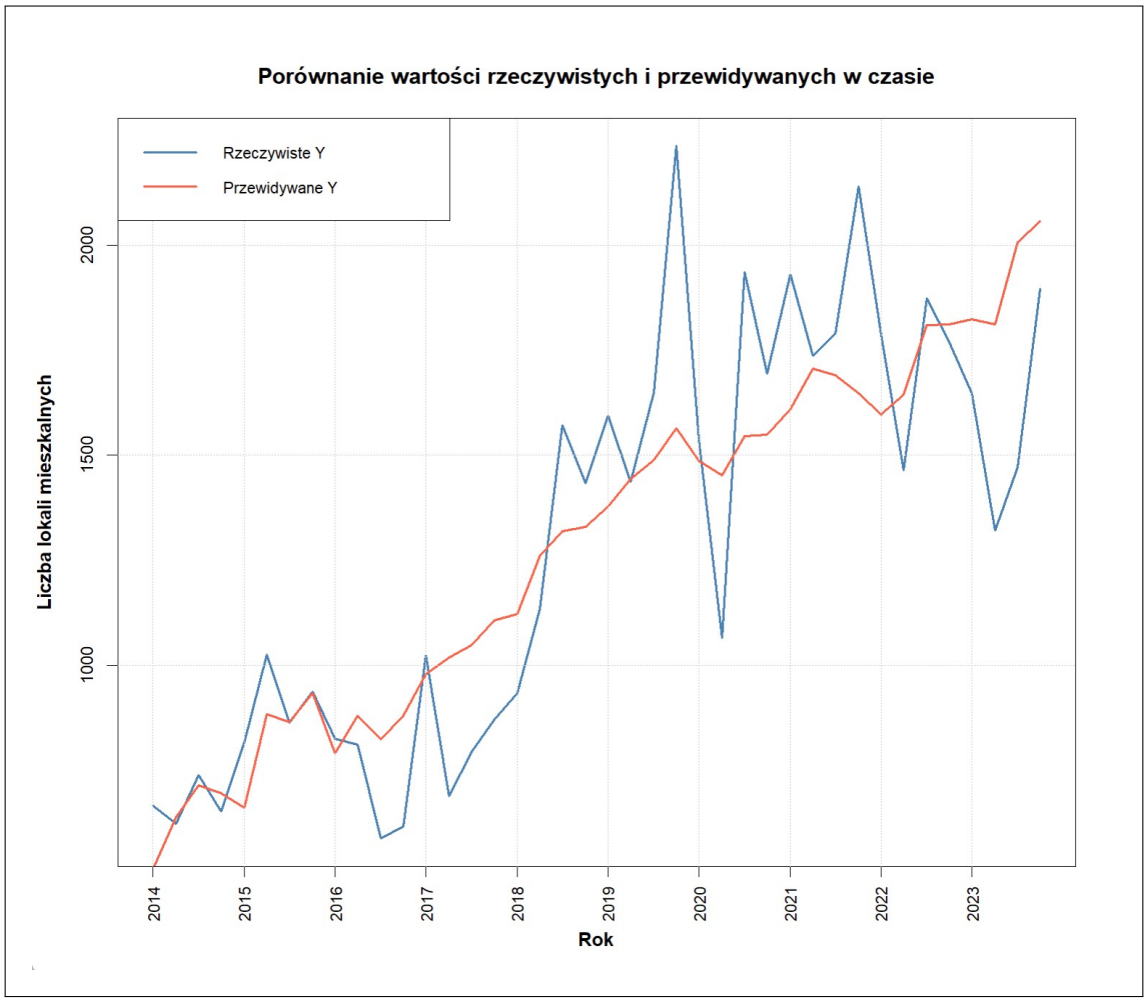
Mimo że model charakteryzuje się dobrym dopasowaniem do danych oraz spełnia większość założeń klasycznego modelu regresji liniowej, wykryta dodatnia autokorelacja składnika losowego wskazuje na potrzebę dalszej weryfikacji i modyfikacji. W celu zwiększenia wiarygodności modelu warto rozważyć jego usprawnienie poprzez zastosowanie metod korygujących autokorelację lub rozszerzenie zbioru danych. Na obecnym etapie model dostarcza użytecznych informacji analitycznych, jednak nie powinien być traktowany jako w pełni wiarygodna podstawa do podejmowania decyzji strategicznych.

5 Prognozowanie

Pomimo wykrycia dodatniej autokorelacji składnika losowego, model cechuje się solidną jakością dopasowania i spełnia kluczowe założenia klasycznej regresji liniowej. Zdecydowaliśmy się zatem na prognozowanie w celu zilustrowania praktycznego zastosowania modelu. Porównanie przewidywań z danymi rzeczywistymi pozwala lepiej ocenić jego możliwości i ograniczenia. Należy jednak pamiętać, że wyniki mają charakter jedynie dydaktyczny, a sam model nie stanowi ostatecznego narzędzia prognostycznego.

5.1 Ocena zgodności wartości rzeczywistych i przewidywanych

W niniejszej części porównamy wartości rzeczywiste zmiennej Y, uzyskane z danych opublikowanych przez GUS, z wartościami prognozowanymi $\hat{Y}$, wyznaczonymi na podstawie skonstruowanego przez nas modelu ekonometrycznego.



Rysunek 31: Wartości rzeczywiste a przewidywane

Na powyższym wykresie widzimy, że wartości przewidywane $\hat{Y}$ dobrze odzwierciedlają ogólny trend wzrostowy liczby lokali mieszkalnych, jednak w niektórych okresach (szczególnie od 2018 roku) model nie oddaje pełnej zmienności danych rzeczywistych, wygładzając ich wahania. Mimo to prognozy pozostają względnie bliskie wartościom empirycznym, co potwierdza umiarkowaną trafność modelu w długoterminowej perspektywie.

5.2 Określenie przyszłego przebiegu rozważanego zjawiska

Modelowanie ekonometryczne nie powinno być jedynie ćwiczeniem teoretycznym, lecz powinno zmierzać do formułowania konkretnych i praktycznych wniosków. Pomimo ograniczonej akceptacji modelu, zdecydowaliśmy się przeprowadzić prognozy w celu oceny, na ile uzyskane wyniki mogą być uznane za wiarygodne oraz przydatne w analizie przyszłego przebiegu badanego zjawiska.

ZADANIE: Określ liczbę lokali mieszkalnych sprzedanych w ramach transakcji rynkowych w pierwszym kwartale 2024 roku. Do prognozy wykorzystaj utworzony model ekonometryczny oraz program Microsoft Excel.

Lp.	Rok	Kwartał	Y	Przewidywane Y	X1	X3
1	2014	I	667	515	16,63	3 182
2		II	624	639	15,17	3 232
3		III	739	715	14,47	3 316
4		IV	654	697	14,4	3 228
5	2015	I	818	663	15,07	3 292
6		II	1 026	884	13,6	3 703
7		III	864	865	12,87	3 421
8		IV	939	934	12,93	3 700
9	2016	I	826	792	13,57	3 345
10		II	813	881	12,13	3 266
11		III	589	824	11,5	2 870
12		IV	618	881	11,4	3 055
13	2017	I	1 024	979	11,57	3 474
14		II	689	1 020	10,23	3 241
15		III	795	1 049	9,77	3 220
16		IV	871	1 107	9,57	3 382
17	2018	I	935	1 123	9,8	3 508
18		II	1 135	1 261	8,93	3 779
19		III	1 572	1 320	8,6	3 907
20		IV	1 433	1 330	8,57	3 937
21	2019	I	1 594	1 379	8,9	4 218
22		II	1 437	1 444	8,07	4 221
23		III	1 649	1 489	7,77	4 306
24		IV	2 236	1 564	7,73	4 579
25	2020	I	1 537	1 487	8,3	4 451
26		II	1 066	1 452	8,87	4 484
27		III	1 935	1 546	8,97	4 871
28		IV	1 694	1 549	8,97	4 882
29	2021	I	1 931	1 610	9,4	5 235
30		II	1 737	1 706	8,77	5 416
31		III	1 788	1 690	8,43	5 258
32		IV	2 139	1 647	8,7	5 174
33	2022	I	1 787	1 598	9,97	5 357
34		II	1 464	1 643	9,13	5 285
35		III	1 873	1 809	8,77	5 806
36		IV	1 768	1 811	8,73	5 802
37	2023	I	1 648	1 824	9,1	5 961
38		II	1 321	1 811	8,6	5 766
39		III	1 471	2 007	8,37	6 441
40		IV	1 895	2 057	8,47	6 659
41	2024	I				

Oszacowane parametry	
	Współczynniki
Przecięcie	945,9762487
X1	-76,46645978
X3	0,264066873

Rysunek 32: Zadanie prognostyczne - część I

Za pomocą funkcji `FORECAST.LINEAR` w programie Microsoft Excel obliczamy prognozowane wartości zmiennych objaśniających dla pierwszego kwartału roku 2024. Funkcja ta wykorzystuje dane historyczne do oszacowania wartości zmiennych X_1 oraz X_3, które następnie zostaną wykorzystane w równaniu modelu do wyznaczenia prognozowanej wartości zmiennej objaśnianej Y. W poprzednich obliczeniach model bazował na rzeczywistych danych dla X_1 i X_3. Tym razem, w celach dydaktycznych zastosowaliśmy wartości prognozowane, aby zilustrować praktyczne działanie modelu.

Lp.	Rok	Kwartał	Y	Przewidywane Y	X1	X3
1	2014	I	667	515	16,63	3 182
2		II	624	639	15,17	3 232
3		III	739	715	14,47	3 316
4		IV	654	697	14,4	3 228
5	2015	I	818	663	15,07	3 292
6		II	1 026	884	13,6	3 703
7		III	864	865	12,87	3 421
8		IV	939	934	12,93	3 700
9	2016	I	826	792	13,57	3 345
10		II	813	881	12,13	3 266
11		III	589	824	11,5	2 870
12		IV	618	881	11,4	3 055
13	2017	I	1 024	979	11,57	3 474
14		II	689	1 020	10,23	3 241
15		III	795	1 049	9,77	3 220
16		IV	871	1 107	9,57	3 382
17	2018	I	935	1 123	9,8	3 508
18		II	1 135	1 261	8,93	3 779
19		III	1 572	1 320	8,6	3 907
20		IV	1 433	1 330	8,57	3 937
21	2019	I	1 594	1 379	8,9	4 218
22		II	1 437	1 444	8,07	4 221
23		III	1 649	1 489	7,77	4 306
24		IV	2 236	1 564	7,73	4 579
25	2020	I	1 537	1 487	8,3	4 451
26		II	1 066	1 452	8,87	4 484
27		III	1 935	1 546	8,97	4 871
28		IV	1 694	1 549	8,97	4 882
29	2021	I	1 931	1 610	9,4	5 235
30		II	1 737	1 706	8,77	5 416
31		III	1 788	1 690	8,43	5 258
32		IV	2 139	1 647	8,7	5 174
33	2022	I	1 787	1 598	9,97	5 357
34		II	1 464	1 643	9,13	5 285
35		III	1 873	1 809	8,77	5 806
36		IV	1 768	1 811	8,73	5 802
37	2023	I	1 648	1 824	9,1	5 961
38		II	1 321	1 811	8,6	5 766
39		III	1 471	2 007	8,37	6 441
40		IV	1 895	2 057	8,47	6 659
41	2024	I		2 004	6,73	5 954

Oszacowane parametry	
	Współczynniki
Przecięcie	945,9762487
X1	-76,46645978
X3	0,264066873

Rysunek 33: Zadanie prognostyczne - część II

Po wykonaniu powyższej procedury otrzymaliśmy informację, że przewidywana liczba lokali mieszkalnych sprzedanych w ramach transakcji rynkowych w pierwszym kwartale 2024 roku może wynieść około 2 004. Tym samym kończymy zadanie prognostyczne. Pozostaje czekać na publikację danych i porównać uzyskaną prognozę z wartościami empirycznymi. Choć model został odrzucony, przeszedł szereg testów pozytywnie i wydaje się mieć duży potencjał, szczególnie jeśli zostanie wzbogacony o dane z kolejnych lat lub zastosowane zostaną metody korygujące autokorelację.

Literatura

[1] Gruszczyński, M. (2012). *Ekonometria i badania operacyjne*. Polskie Wydawnictwo Ekonomiczne.

[2] Etapy modelowania ekonometrycznego, dostęp: https://etrapez.pl/wp-content/uploads/domowe/ke/Etapy%20Modelowania%20Ekonometrycznego.pdf.

[3] Metoda Hellwiga, dostęp: https://pl.wikipedia.org/wiki/Metoda_Hellwiga.