

**WYŻSZA SZKOŁA INFORMATYKI STOSOWANEJ
I ZARZĄDZANIA**

pod auspicjami Polskiej Akademii Nauk



**WYDZIAŁ INFORMATYCZNYCH
TECHNIK ZARZĄDZANIA**

STUDIA II STOPNIA (MAGISTERSKIE)

**METODY
PROGNOZOWANIA I SYMULACJI**

MODEL EKONOMETRYCZNY

Autorzy:	Marcin Górowski	UZN02K3
	Sławomir Chodkiewicz	UZN02K3

WARSZAWA, 2008 r.

1. Postać modelu

Postać modelu:
$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + a_3X_{3t} + a_4X_{4t} + \xi_t$$

gdzie: Y_t - emigracja zagraniczna ludności na pobyt stały w tys.

X_{1t} - ludność miast w mln.

X_{2t} - ludność wiejska w mln.

X_{3t} - przyrost naturalny w tys.

X_{4t} - małżeństwa w tys.

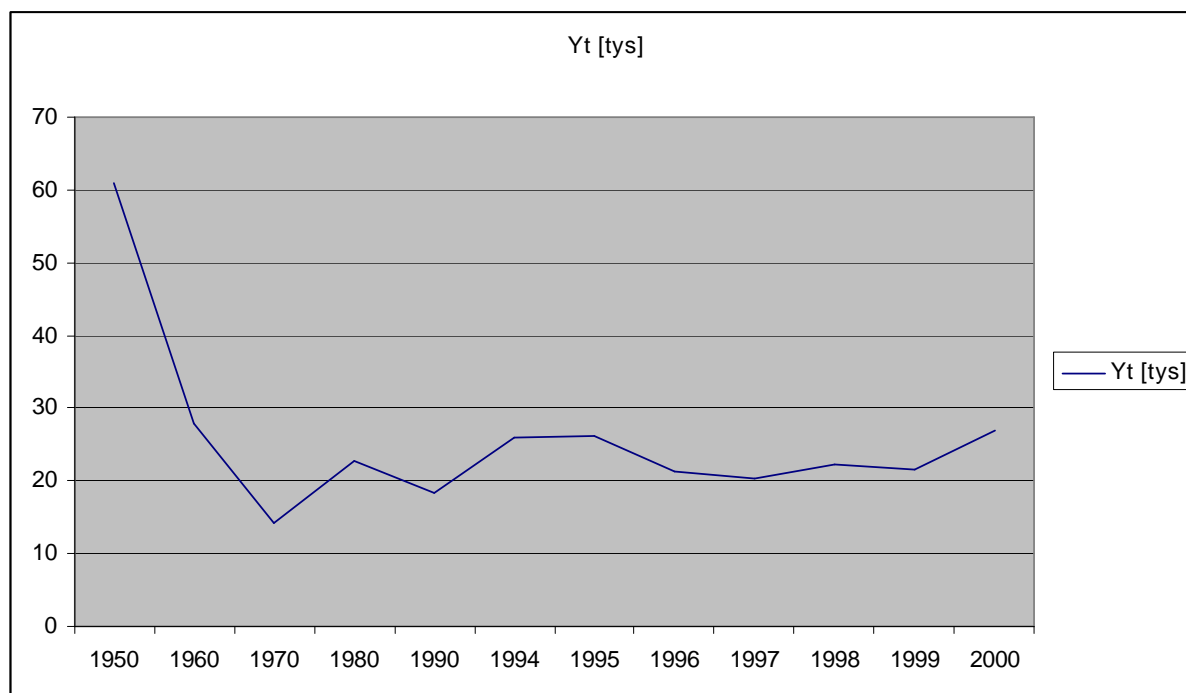
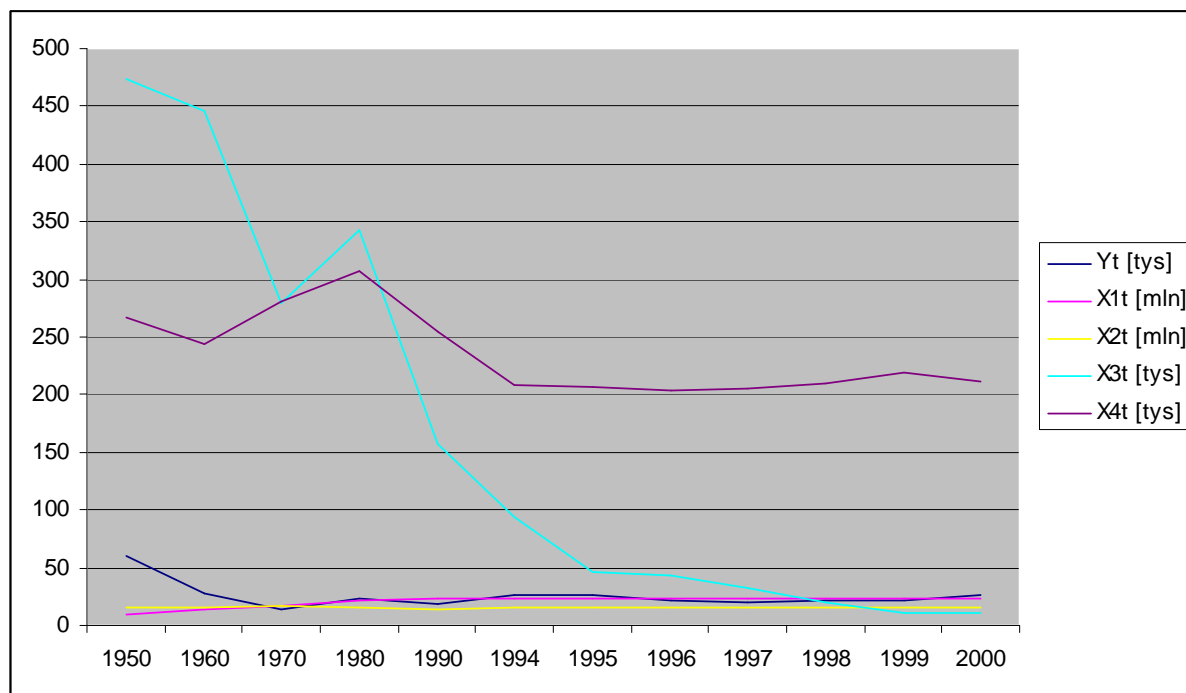
a_0, a_1, a_2, a_3, a_4 – parametry strukturalne

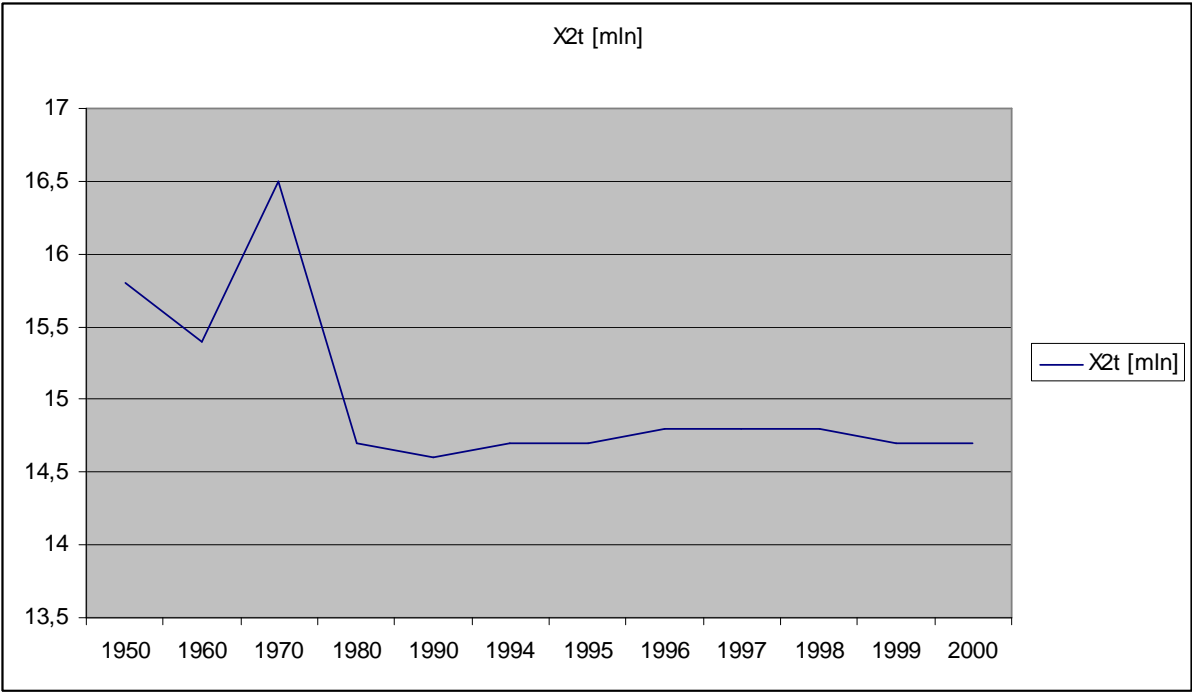
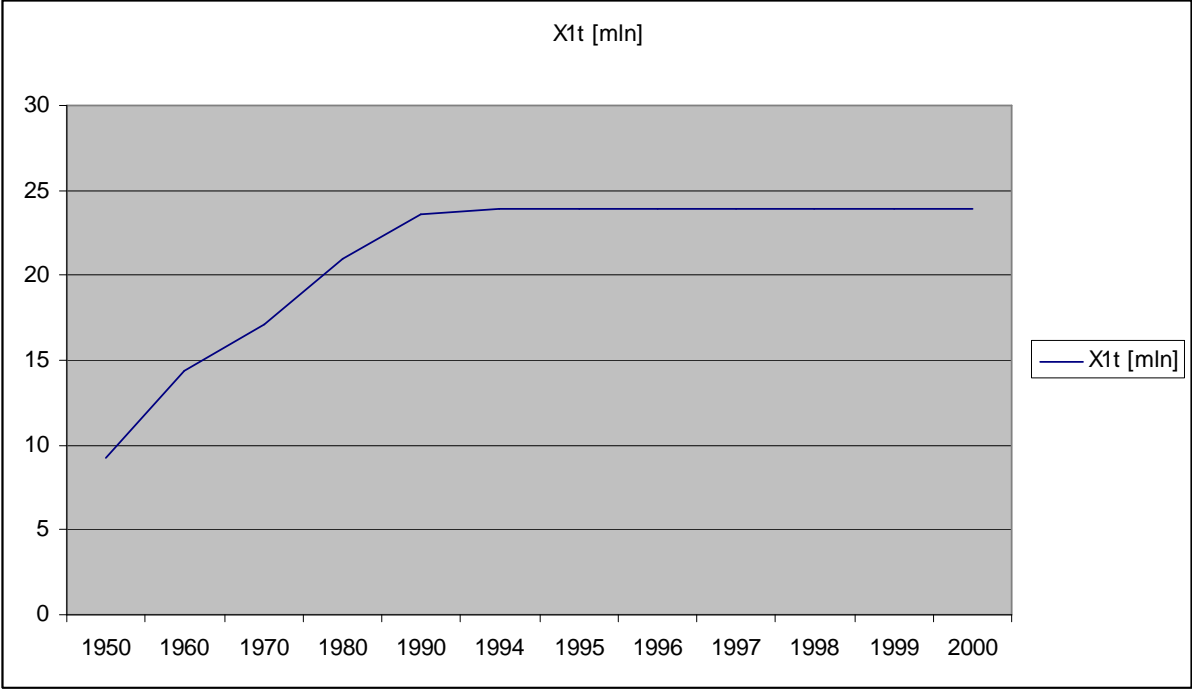
1.1 Dane

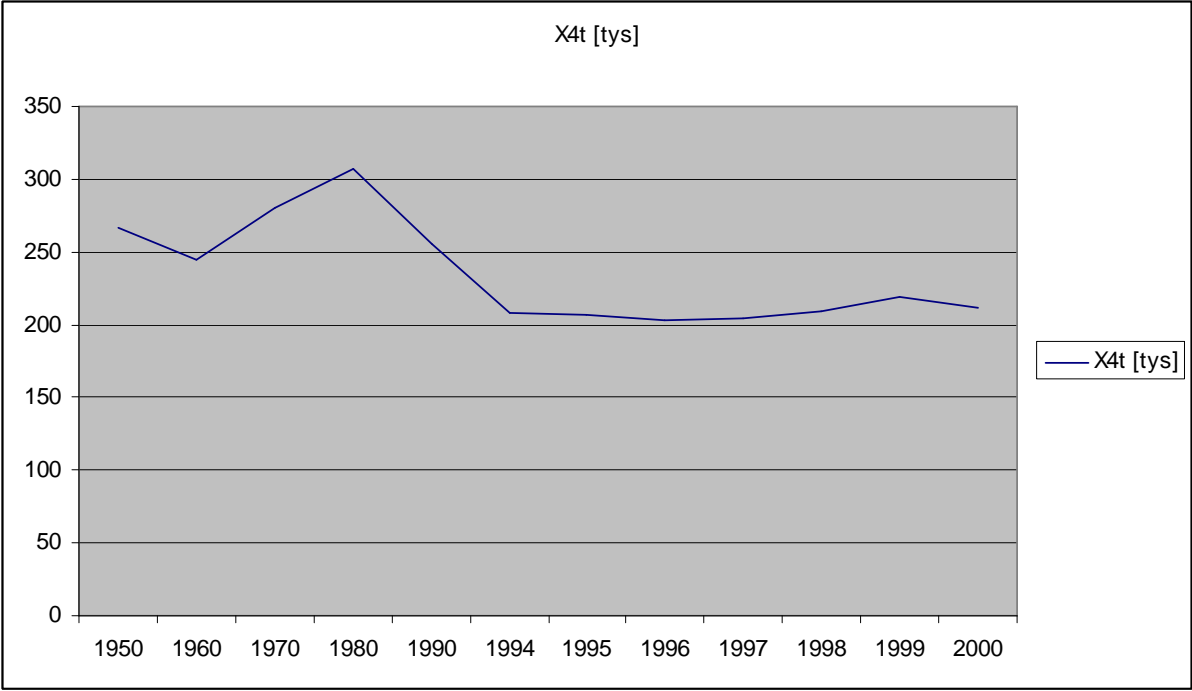
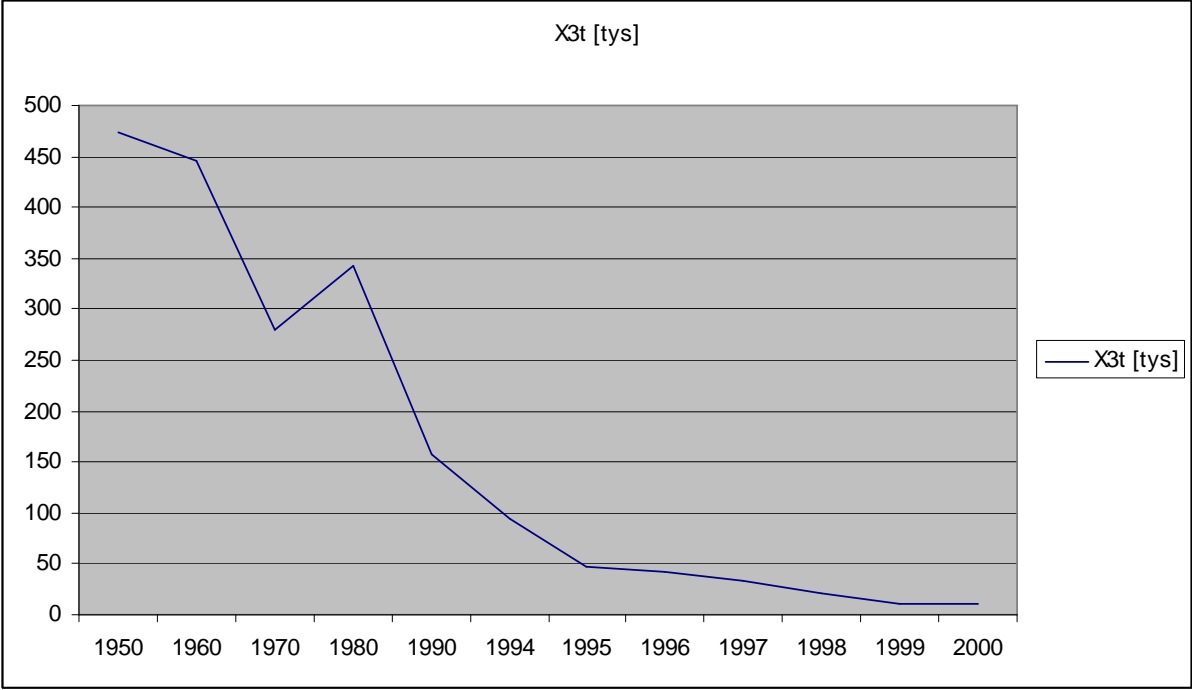
Dane do modelu obejmują 13 obserwacji w latach 1946 – 2000, przy czym rok 1946 w dalszej części pracy zostanie pominięty, gdyż z przyczyn historycznych wprowadza zakłócenia do obserwacji (brane pod uwagę jest 12 obserwacji).

t	Y_t	X_{1t}	X_{2t}	X_{3t}	X_{4t}
1946	1835,9 ¹	8,0	15,6	380,7	281,9
1950	60,9	9,2	15,8	474,4	267,1
1960	28,0	14,4	15,4	445,3	244,2
1970	14,1	17,1	16,5	279,2	280,3
1980	22,7	21,0	14,7	342,6	307,4
1990	18,4	23,6	14,6	157,4	255,4
1994	25,9	23,9	14,7	94,9	207,7
1995	26,3	23,9	14,7	47,0	207,1
1996	21,3	23,9	14,8	42,7	203,6
1997	20,2	23,9	14,8	32,5	204,9
1998	22,2	23,9	14,8	20,3	209,4
1999	21,5	23,9	14,7	10,6	219,4
2000	27,0	23,9	14,7	10,3	211,2

1.2 Graficzna prezentacja danych







2. Dobór zmiennych objaśniających metodą Hellwiga

Współczynniki korelacji liniowej dla obserwacji z próby 1-12
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,5760 dla $n = 12$

Wynik z programu Gretl:

Yt	X1t	X2t	X3t	X4t	
1,0000	-0,7090	0,2480	0,5225	0,1456	Yt
	1,0000	-0,7670	-0,9061	-0,5734	X1t
		1,0000	0,6277	0,4979	X2t
			1,0000	0,7807	X3t
				1,0000	X4t

$R_0 =$

	X _{1t}	X _{2t}	X _{3t}	X _{4t}
Y	-0,70905	0,24802	0,5225	0,14561

$R =$

	X _{1t}	X _{2t}	X _{3t}	X _{4t}
X _{1t}	1	-0,767	0,9061	-0,5734
X _{2t}	-0,76697	1	0,6277	0,49786
X _{3t}	-0,90613	0,62765	1	0,78071
X _{4t}	-0,57343	0,49786	0,7807	1

gdzie:

$R_0 = [r_j]$ – macierz współczynników korelacji liniowej pomiędzy j . zmienną objaśniającą a zmienną objaśnianą, $j = 1, 2, \dots, k$,

$R = [r_{ij}]$ – macierz współczynników korelacji liniowej pomiędzy i . a j . zmienną objaśniającą, $i = 1, 2, \dots, k, j = 1, 2, \dots, k$,

k – liczba zmiennych objaśniających w modelu.

Ilość możliwych pozdbiorów ze zbioru zmiennych objaśniających $\{X_1, X_2, X_3, X_4\}$ (bez zbioru pustego) wynosi:

$$S = 2^k - 1 = 2^4 - 1 = 16 - 1 = 15$$

Te możliwe podzbiory to:

	h	1	2	3	4	H - suma
C1	x1	0,50275				0,5027
C2	x2		0,06151			0,0615
C3	x3			0,27296		0,2730
C4	x4				0,02120	0,0212
C5	x1;x2	0,28453	0,03481			0,3193
C6	x1;x3	0,26375		0,14320		0,4070
C7	x1;x4	0,31953			0,01347	0,3330
C8	x2;x3		0,03779	0,16770		0,2055
C9	x2;x4		0,04107		0,01415	0,0552
C10	x3;x4			0,15329	0,01191	0,1652
C11	x1;x2;x3	0,18808	0,02569	0,10773		0,3215
C12	x1;x2;x4	0,21481	0,02716		0,01024	0,2522
C13	x1;x3;x4	0,20276		0,10159	0,00901	0,3134
C14	x2;x3;x4		0,02894	0,11334	0,00930	0,1516
C15	x1;x2;x3;x4	0,15486	0,02127	0,08235	0,00743	0,2659

max: **0,5027**

Z czego do modelu wybieramy podzbiór o największej pojemności integralnej tj. 0,5027 – C1 – czyli tylko zmienna X1

$$Y_t = a_0 + a_1 * X_{1t} + \zeta$$

3. Oszacowanie parametrów strukturalnych

Klasyczna metoda najmniejszych kwadratów polega na takim wyznaczeniu parametrów modelu, ażeby suma kwadratów między zaobserwowanymi wartościami zmiennej objaśnianej a odpowiednimi wartościami modelowymi była najmniejsza.

Wyniki z Gretl:

Estymacja KMNK z wykorzystaniem 12 obserwacji 1-12
Zmienna zależna: Yt

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
const	61,581	11,558	5,3280	0,00033	***
X1t	-1,70417	0,53595	-3,1797	0,00982	***

Średnia arytmetyczna zmiennej zależnej = 25,7083

Odchylenie standardowe zmiennej zależnej = 11,7658

Suma kwadratów reszt = 757,198

Błąd standardowy reszt = 8,70172
 Wsp. determinacji $R^2 = 0,502749$
 Skorygowany $R^2 = 0,453024$
 Stopnie swobody = 10
 Logarytm wiarygodności = -41,8956
 Kryterium informacyjne Akaike'a = 87,7912
 Kryterium bayesowskie Schwarza = 88,761
 Kryterium infor. Hannana-Quinna = 87,4321

Czyli po podstawieniu mode przyjmuje postać:

$$Y_t = 61,581 - 1,70417 * X_{1t}$$

3.1 Interpretacja oszacowanych parametrów

$a_1 = -1,70417$ – wzrost ludności miasta o 1 milion mieszkańców spowoduje spadek emigracji zagranicznej o 1704 osoby.

3.2 Interpretacja błędów szacunku parametrów

$S_{a0} = 11,558$ – szacując a_0 na poziomie 61,581 mylimy się średnio o (+-) 11,558

$S_{a1} = 0,53595$ – szacując a_1 na poziomie -1,70417 mylimy się średnio o (+-) 0,53595

4. Współliniowość zmiennych objaśniających

Dla k modeli:

$$X_{1t} = \alpha_{1,0} + \alpha_{1,2}X_{2t} + \alpha_{1,3}X_{3t} + \dots + \alpha_{1,k}X_{kt} + \epsilon_{1t},$$

$$X_{2t} = \alpha_{2,0} + \alpha_{2,1}X_{1t} + \alpha_{2,3}X_{3t} + \dots + \alpha_{2,k}X_{kt} + \epsilon_{2t},$$

\vdots

$$X_{kt} = \alpha_{k,0} + \alpha_{k,1}X_{1t} + \alpha_{k,2}X_{2t} + \dots + \alpha_{k,k-1}X_{(k-1)t} + \epsilon_{kt},$$

obliczamy współczynnik determinacji R_j^2 oraz czynnik inflacji wariancji estymatora α_j :

$$CIW_j = \frac{1}{1 - R_j^2}. \quad (22)$$

Jeżeli:

$R_j^2 = 0$ oraz $CIW_j = 1$ – brak współliniowości zmiennych,

$R_j^2 > 0$ oraz $CIW_j > 1$ – przybliżona współliniowość zmiennych,

$CIW_j > 10$ – współliniowość zmiennych trwale zakłócająca jakość modelu.

Czyli:

- dla $X1t$

Estymacja KMNK z wykorzystaniem 12 obserwacji 1-12
Zmienna zależna: $X1t$

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
Const	55,6801	14,896	3,7379	0,00572	***
X2t	-2,7706	0,977138	-2,8354	0,02197	**
X3t	-0,0270816	0,0045267	-5,9826	0,00033	***
X4t	0,0485163	0,0202076	2,4009	0,04312	**

Wsp. determinacji $R^2 = 0,933695$

$X1t = 55,6801 - 2,7706 X2t - 0,0270816 X3t + 0,0485163 X4t$
 $R^2 = 0,933695 \rightarrow CIW = 15,08182$

- dla $X2t$

Estymacja KMNK z wykorzystaniem 12 obserwacji 1-12
Zmienna zależna: $X2t$

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
const	17,3578	1,4602	11,8873	<0,00001	***
X3t	-0,00388055	0,00233279	-1,6635	0,13479	
X4t	0,00894354	0,0059898	1,4931	0,17375	
X1t	-0,180912	0,0638043	-2,8354	0,02197	**

Wsp. determinacji $R^2 = 0,6978$

$X2t = 17,3578 - 0,180912 X1t - 0,00388055 X3t + 0,00894354 X4t$
 $R^2 = 0,6978 \rightarrow CIW = 3,309067$

- dla $X3t$

Estymacja KMNK z wykorzystaniem 12 obserwacji 1-12
Zmienna zależna: $X3t$

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
const	1316,3	680,12	1,9354	0,08898	*
X4t	2,02978	0,517641	3,9212	0,00441	***
X1t	-30,1798	5,04457	-5,9826	0,00033	***
X2t	-66,228	39,8128	-1,6635	0,13479	

Wsp. determinacji $R^2 = 0,942534$

$X3t = 1316,3 - 30,1798 X1t - 66,228 X2t + 2,02978 X4t$
 $R^2 = 0,942534 \rightarrow CIW = 17,40159$

- dla X4t

Estymacja KMNK z wykorzystaniem 12 obserwacji 1-12
Zmienna zależna: X4t

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
const	-365,684	302,839	-1,2075	0,26172	
X1t	8,63188	3,59527	2,4009	0,04312	**
X2t	24,3688	16,3206	1,4931	0,17375	
X3t	0,324059	0,0826426	3,9212	0,00441	***

Wsp. determinacji $R^2 = 0,7731$

$X4t = -365,684 + 8,63188 X1t + 24,3688 X2t + 0,324059 X3t$
 $R^2 = 0,7731 \rightarrow CIW = 4,407228$

Inny sposób:

Ocena współliniowości VIF - czynnika powiększania wariancji

Minimalna możliwa wartość = 1.0

Wartości > 10.0 mogą wskazywać na problem współliniowości-rozdęcia wariancji

- 2) X1t 15,082
- 3) X2t 3,309
- 4) X3t 17,402
- 5) X4t 4,407

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

Własności macierzy $X'X$:

1-norm = 1291569,6

Wyznacznik = 8,7065065e+011

Wskaźnik uwarunkowania macierzy CN = 2,6024123e-009

Podsumowanie:

X1 i X3 wykazują bardzo dużą współliniowość jednak nie ma to znaczenia, gdyż do modelu wybraliśmy tylko X1.

5. Liniowość modelu

H_0 : oszacowany model jest liniowy,

H_1 : oszacowany model nie jest liniowy.

poziom istotności: $\alpha = 0,05$,

liczba serii: 12

liczba reszt ujemnych: $n_1 = 5$

liczba reszt dodatnich: $n_2 = 7$

poziom krytyczny odczytany z tablic $r^* = 3$

Przy 5% poziomie istotności $r > r^*$, więc nie ma podstaw do odrzucenia H_0 o liniowości modelu.

6. Współczynnik determinacji

Współczynnik determinacji R-kwadrat, który określa w jakim stopniu zmienność zmiennej zależnej została wyjaśniona zmiennością zmiennej niezależnej.

R-kwadrat przyjmuje wartości z przedziału o zera do jeden. Im jest bliżej jedynki tym lepiej została wyjaśniona zmienność zmiennej zależnej przez model regresji i tym samym model lepiej wyjaśnia opisywaną zależność.

Model 9: Estymacja KMNK z wykorzystaniem 12 obserwacji 1-12
Zmienna zależna: Y_t

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
const	446,843	79,3872	5,6286	0,00079	***
X1t	-6,17607	1,13696	-5,4321	0,00097	***
X2t	-19,5848	4,44937	-4,4017	0,00315	***
X3t	-0,0920403	0,0340584	-2,7024	0,03053	**
X4t	0,0765815	0,0852387	0,8984	0,39880	

Średnia arytmetyczna zmiennej zależnej = 25,7083

Odchylenie standardowe zmiennej zależnej = 11,7658

Suma kwadratów reszt = 158,161

Błąd standardowy reszt = 4,75335

Wsp. determinacji $R^2 = 0,896136$

Skorygowany $R^2 = 0,836785$

Statystyka F (4, 7) = 15,099 (wartość p = 0,00149)

Logarytm wiarygodności = -32,4995

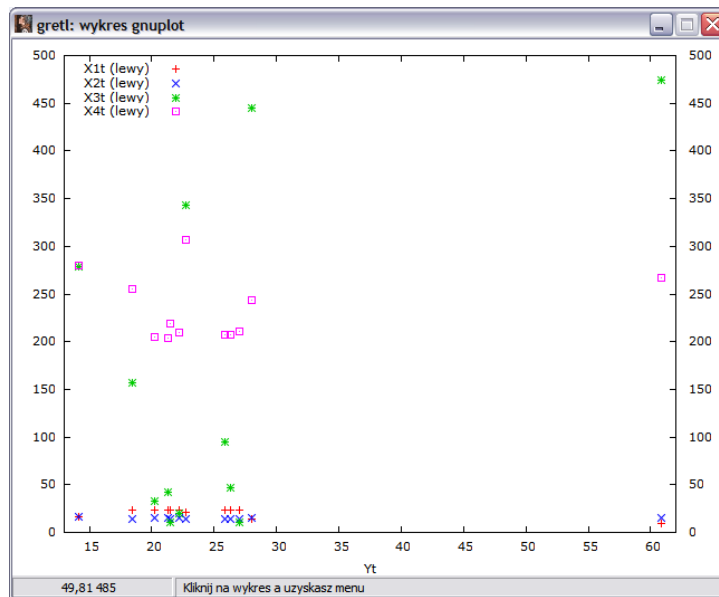
Kryterium informacyjne Akaike'a = 74,999

Kryterium bayesowskie Schwarz = 77,4235

Kryterium infor. Hannana-Quinna = 74,1013

Wartość współczynnika determinacji jest bliska jeden co świadczy o bardzo dobrym dopasowaniu modelu regresji do danych. Zmienność zamówień w ponad 89% została wyjaśniona zmiennością temperatury.

Wykres rozrzutu:



7. Istotność zmiennych objaśniających – test t-Studenta

$H_0 : \alpha_j = 0$ – j . zmienna jest nieistotna w modelu, $j = 0, 1, 2, \dots, k$,

$H_1 : \alpha_j \neq 0$ – j . zmienna jest istotna w modelu, $j = 0, 1, 2, \dots, k$.

Zmienna losowa:

$$t_{\hat{\alpha}_j} = \left| \frac{\hat{\alpha}_j}{S_{\hat{\alpha}_j}} \right|, \quad (23)$$

ma rozkład t-Studenta z $n - (k + 1)$ stopniami swobody.

Statystyki t-Studenta oraz krytyczne (nominalne) poziomy istotności dla poszczególnych zmiennych wynoszą odpowiednio:

Estymacja KMNK z wykorzystaniem 12 obserwacji 1-12

Zmienna zależna: Y_t

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
Const	446,843	79,3872	5,6286	0,00079	***
X1t	-6,17607	1,13696	-5,4321	0,00097	***
X2t	-19,5848	4,44937	-4,4017	0,00315	***
X3t	-0,0920403	0,0340584	-2,7024	0,03053	**
X4t	0,0765815	0,0852387	0,8984	0,39880	

Statystyka t-Studenta odczytana z tablic dla 5% poziomu istotności i $df=12-(4+1)=7$ stopni swobody wynosi:

$$t_{\alpha=0,05;df=7} = 1,89458$$

i jest większa od wartości statystyk t-Studenta dla poszczególnych zmiennych, co prowadzi do wniosku o braku statystycznej istotności zmiennych.

8. Prognozy

Wprowadźmy oznaczenia dla okresu prognozowanego τ :

$$\tau = n + 1, n + 2, \dots, n + s = T.$$

Zatem długość okresu prognozy wynosi: $T - n = s$.

8.1 Prognoza zmiennej objaśnianej na podstawie modelu trendu wielomianowego

Postać trendu wielomianowego dla zmiennej Y_t :

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_q t^q + \epsilon_t,$$

Rząd q ustalamy na podstawie otrzymanego optymalnego modelu:

Zatem oszacowany model przyjmuje postać:

8.2 Ocena ex ante prognozy – średni błąd prognozy

8.3 Prognoza przedziałowa

8.4 Prognoza zmiennych objaśniających na podstawie modeli trendu liniowego

9. Ocena prognozy