

Rzeszów 14.06.2023

Ekonometria - Projekt

Złotek Sebastian
Inżynieria i Analiza Danych

Spis treści

1. Czym jest model ekonometryczny.....	2
2. Wykorzystane dane.....	2
3. Dobór zmiennych metodą quasi-stałych	3
4. Metoda Hellwiga.....	4
5. Metoda KMNK i jej implementacja	10
6. Weryfikacja modelu.....	11
7. Test symetrii składnika losowego.....	13
8. Istotność statystyczna F	14
9. Test Jaeque-Bera.....	15
10. Badanie autokorelacji składników losowych	16
11. Badanie koincydencji	17
12. Test Shapiro-Wilka	18
13. VIF	19
14. Test losowości reszt (składnika losowego)-test serii.....	20
15. Test Goldfelda – Quandta.....	22
16. Test t-Studenta	23
17. Podsumowanie	23

1. Czym jest model ekonometryczny

Model ekonometryczny to matematyczne opisanie zależności między zmiennymi ekonomicznymi, które pozwala na prognozowanie ich wartości na podstawie danych historycznych. Modele te zawierają zarówno zmienne objaśniające jak i zmienną zależną, którą chcemy przewidzieć oraz ewentualne zmienne dodatkowe. Modele te często zawierają również założenia dotyczące charakteru tych zależności, np. liniowość. Modele te są używane w wielu dziedzinach, w tym w ekonomii, finansach i marketingu.

2. Wykorzystane dane

Dane pochodzą ze strony <https://stat.gov.pl/wskazniki-makroekonomiczne/> i są to:

- Przychody operacyjne OTWARTYCH FUNDUSZY EMERYTALNYCH,
- Koszty operacyjne OFE,
- Zrealizowany zysk/strata z inwestycji OFE,
- Niezrealizowany zysk/strata z wyceny OFE,
- Wynik finansowy OFE,
- Aktywa ogółem OFE,
- Aktywa netto OFE,
- Przychody ogółem powszechnych towarzystw emerytalnych (PTE),
- Koszty ogółem PTE,
- Aktywa ogółem PTE.

Rok	Przychody operacyjne OTWARTYCH FUNDUSZY EMERYTALNYCH	Koszty operacyjne OFE	Zrealizowany zysk/strata z inwestycji OFE	Niezrealizowany zysk/strata z wyceny OFE	Wynik finansowy OFE	Aktywa ogółem OFE	Aktywa netto OFE	Przychody ogółem powszechnych towarzystw emerytalnych (PTE)	Koszty ogółem PTE	Aktywa ogółem PTE
	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
2011	9 197,4	1 228,5	1 776,4	-20 758,5	-11 013,3	227 451,0	224 720,1	1 810,4	1 048,7	3 601,4
2012	11 346,8	1 312,0	2 507,5	24 884,3	37 426,6	272 299,9	269 596,5	1 653,6	770,0	3 991,4
2013	11 403,8	1 574,9	1 383,5	8 709,5	19 921,9	300 817,3	299 272,4	1 783,3	1 291,5	3 904,9
2014	5 480,7	1 165,7	4 486,0	-8 948,5	-147,6	149 407,7	149 054,6	2 153,9	805,2	3 882,0
2015	4 032,4	974,7	-9 902,3	-11 827,2	-6 844,7	140 895,4	140 496,1	1 026,9	429,3	2 787,3
2017	3 766,5	969,8	2 441,9	23 165,2	28 403,8	179 956,3	179 529,9	1 022,0	447,5	2 827,4
2018	4 310,6	968,2	1 041,3	-21 394,9	-17 011,2	157 511,3	157 333,9	1 092,4	393,0	2 929,1
2019	4 803,8	945,6	209,4	-2 648,5	1 419,2	154 968,0	154 816,1	1 026,0	410,6	2 859,0
2020	1 803,4	796,3	-12,8	-4 484,5	-3 490,3	148 824,2	148 604,1	950,9	518,4	2 750,1
2021	4 384,1	1 013,4	1 485,4	38 642,8	43 498,9	188 225,0	187 985,1	1 048,3	496,4	2 705,7

3. Dobór zmiennych metodą quasi-stałych

Usunięcie zmiennych z modelu ekonometrycznego za pomocą metody quasi-stałych polega na przekształceniu trudnych do bezpośredniego pomiaru zmiennych w łatwiejsze do interpretacji i włączeniu ich jako zmiennych objaśniających w modelu. W tym procesie oryginalne zmienne są uwzględnione jako składniki tworzące nowe zmienne.

Po zbudowaniu modelu z uwzględnieniem przekształconych zmiennych, można usunąć oryginalne zmienne, ponieważ są już uwzględnione w nowych zmiennych. To usunięcie może być konieczne ze względu na założenia dotyczące charakteru zmiennych, takie jak normalność rozkładu, które są niezgodne z oryginalnymi zmiennymi, ale są spełnione przez przekształcone zmienne.

W rezultacie, usunięcie zmiennych z modelu ekonometrycznego za pomocą metody quasi-stałych pozwala na uzyskanie modelu, który jest bardziej zgodny z założeniami dotyczącymi charakteru zmiennych i może przynieść bardziej dokładne prognozy.

Zmienność, a właściwie jej poziom wyrażamy:

$$V_i = \frac{S_i}{\bar{X}_i} \cdot 100\%$$

Przy czym:

S_i – odchylenie standardowe

\bar{X}_i – średnia z próby

odch. Stand		215,5550159	3686,4745	19484,627	20295,1	53437,9	52657,36	420,868383	293,553294	518,5873
średnia		1094,91	541,63	2533,97	9216,33	275588	263764,436	1356,77)	661,06	3223,83
Vj		19,7%	681%	769%	220%	19%	19,96%	18,3%	44%	16%

Zmienne $\geq 20\%$ zostają, gdyż jest ona zmienną objaśniającą.

W celu uproszczenia przyjmę nowe oznaczenia dla zmiennych :					
Y	Przychody operacyjne OTWARTYCH FUNDUSZY EMERYTALNYCH				
X1	Zrealizowany zysk/strata z inwestycji OFE	zamiast	X2		
X2	Niezrealizowany zysk/strata z wyceny OFE	zamiast	X3		
X3	Wynik finansowy OFE	zamiast	X4		
X4	Aktywa ogółem PTE	zamiast	X8		
rok	y	X1	X2	X3	X4
2012	9197,4	1776,4	-20758,5	-11013,3	1048,7
2013	11346,8	2507,5	24884,3	37426,6	770
2014	11403,8	1383,5	8709,5	19921,9	1291,5
2015	5480,7	4486	-8948,5	-147,6	805,2
2016	4032,4	-9902,3	-11827,2	-6844,7	429,3
2017	3766,5	2441,9	23165,2	28403,8	447,5
2018	4310,6	1041,3	-21394,9	-17011,2	393
2019	4803,8	209,4	-2648,5	1419,2	410,6
2020	1803,4	-12,8	-4484,5	-3490,3	518,4
2021	4384,1	1485,4	38642,8	43498,9	496,4
postać modelu:					
$Y = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_3 + \alpha_3 x_4 + \alpha_4 x_8 + \varepsilon$					

Nadaje nowe oznaczenia zmiennym i tworze postać modelu

4. Metoda Hellwiga

Metoda Hellwiga (zwana również metodą Generalized SID) jest metodą wyboru zmiennych w modelach ekonometrycznych. Polega ona na określeniu wpływu każdej zmiennej na predykcję modelu, a następnie na porównaniu tych wpływów i wybraniu tylko najważniejszych zmiennych.

Metoda Hellwiga opiera się na testach statystycznych i polega na wyborze zmiennych za pomocą kryterium minimalizacji sumy kwadratów błędów. Zmienna, która najbardziej zwiększa sumę kwadratów błędów, gdy jest usunięta z modelu, jest uważana za najmniej istotną i jest usuwana z modelu.

Metoda ta jest używana w wielu dziedzinach, w tym w ekonometrii, finansach, marketingu i badaniach jakości, aby wybrać najważniejsze zmienne i uzyskać bardziej dokładne prognozy. W porównaniu z innymi metodami wyboru zmiennych, metoda Hellwiga jest uważana za szybszą i bardziej wydajną, a także za bardziej skuteczną w wyborze zmiennych o niskiej korelacji.



Zmienne X1 i Y są	slabo skorelowane i jest to zjawisko negatywne dla modelu
Zmienne X2 i Y są	slabo skorelowane i jest to zjawisko negatywne dla modelu
Zmienne X3 i Y są	slabo skorelowane i jest to zjawisko negatywne dla modelu
Zmienne X4 i Y są	silnie skorelowane i jest to zjawisko pozytywne dla modelu

Gdy zachodzi konieczność ograniczenia liczby elementów zbioru X pierwotnie wybranych zmiennych objaśniających posłużymy się metodą optymalnego wyboru predyktant- **metodą Hellwiga**.

Każdą z "kandydatek" możemy uważać za źródło wiedzy o zmiennej Y, zasadne jest więc traktowanie jej jako **nośnika informacji** o zmiennej objaśnianej.

Rozpatrujemy wszystkie niepuste kombinacje m-zmiennych ze zbioru X. Ich liczba wynosi: $2^m - 1$

Zbiór elementów tworzący S-tą kombinację oznaczamy przez C_S .

Aby stosować metodę Hellwiga, najpierw wyznacz współczynniki korelacji między zmiennymi X i między X i Y:

$$h_{sj} = \frac{r_j^2}{\sum_{i \in C_s} |r_{ij}|}$$

Indywidualną pojemność informacyjną nośnika informacji oznaczamy:

a następnie obliczamy integralną pojemność informacyjną S-tej kombinacji: $H_S = \sum_{j \in C_s} h_{sj}$

Za najlepszą (optymalną) kombinację nośników informacji uznajemy ten podzbiór kandydatek na zmienne objaśniające, dla którego pojemność integralna jest największa.

X1	X2	X3	X4	Y
1 776,40	-20 758,50	-11 013,30	1 048,70	9 197,40
2 507,50	24 884,30	37 426,60	770,00	11 346,80
1 383,50	8 709,50	19 921,90	1 291,50	11 403,80
4 486,00	-8 948,50	-147,60	805,20	5 480,70
-9 902,30	-11 827,20	-6 844,70	429,30	4 032,40
2 441,90	23 165,20	28 403,80	447,50	3 766,50
1 041,30	-21 394,90	-17 011,20	393,00	4 310,60
209,40	-2 648,50	1 419,20	410,60	4 803,80
-12,80	-4 484,50	-3 490,30	518,40	1 803,40
1 485,40	38 642,80	43 498,90	496,40	4 384,10

nr.kombin.	S	Kombinacja	Indywidualne pojemności nośników informacji h	Integralna pojemność nośników informacji H
1	{X1}		h11= 0,08185	H1= 0,08184542
2	{X2}		h22= 0,01841	H2= 0,01840618
3	{X3}		h33= 0,0855	H3= 0,08549916
4	{X4}		h44= 0,67776	H4= 0,67776359
5	{X1,X2}		h51= 0,06458 h52= 0,01452	H5= 0,07910546
6	{X1,X3}		h61= 0,06224 h63= 0,06502	H6= 0,12726223
7	{X1,X4}		h71= 0,06108 h74= 0,50578	H7= 0,56685975
8	{X2,X3}		h82= 0,00927 h83= 0,04306	H8= 0,05233297
9	{X2,X4}		h92= 0,01759 h94= 0,64758	H9= 0,66516909
Na tą kombinację wskazuje metoda Hellwiga.	10	{X3,X4}	h103= 0,07839 h104= 0,62137	H10= 0,6997582
	11	{X1,X2,X3}	h111= 0,05173 h112= 0,00817 h113= 0,03717	H11= 0,09706354
	12	{X1,X2,X4}	h121= 0,05092 h122= 0,01401 h124= 0,48878	H12= 0,55371097
	13	{X1,X3,X4}	h131= 0,04945 h133= 0,06082 h134= 0,4737	H13= 0,58397799
	14	{X2,X3,X4}	h142= 0,00906 h143= 0,04118 h144= 0,59591	H14= 0,64614868
	15	{X1,X2,X3,X4}	h151= 0,04258 h152= 0,008 h153= 0,03576 h154= 0,45876	H15= 0,54509574

PODSUMOWANIE - WYJŚCIE						
Statystyki regresji						
Wielokrotny	0,9440608					
R kwadrat	0,8912508					
Dopasowa	0,8601796					
Błąd stand.	1256,4856					
Obserwacj	10					
ANALIZA WARIANCJI						
	df	SS	MS	F	Istotność F	
Regresja	2	90570558	45285279	28,68415	0,000424	
Resztkowy	7	11051292	1578756			
Razem	9	101621850				
Współczynniki i standardy błędów						
	Współczynniki	Standardy błędów	t Stat	Wartość p	Dolne 95%	Górne 95%
Przecięcie	281,90644	861,29882	0,327304	0,753009	-1754,742	2318,55
X2	-0,864531	0,1200449	-7,201727	0,000177	-1,148392	-0,5807
X3	0,8638729	0,1152512	7,495568	0,000138	0,591347	1,1364
SKŁADNIKI RESZTOWE - WYJŚCIE						
Obserwacje z wyjątkowo wysokimi wartościami resztowymi						
1	8714,173	483,22697				
2	11100,495	246,30516				
3	9962,2676	1441,5324				
4	7890,6509	-2409,9509				
5	4593,9316	-561,53163				
6	4792,1566	-1025,6566				
7	4082,9367	227,66325				
8	3797,6242	1006,1758				
9	1143,7182	659,68182				
10	4451,5462	-67,446238				

Optymalnym zbiorem zmiennych objaśniających jest kombinacja {X3 X4} która ma największą integralną pojemność nośników informacji.

Zatem dane do modelu prezentują się następująco:

rok	y	x3	x4
2012	9 197,40	-11 013,30	1 048,70
2013	11 346,80	37 426,60	770,00
2014	11 403,80	19 921,90	1 291,50
2015	5 480,70	-147,60	805,20
2016	4 032,40	-6 844,70	429,30
2017	3 766,50	28 403,80	447,50
2018	4 310,60	-17 011,20	393,00
2019	4 803,80	1 419,20	410,60
2020	1 803,40	-3 490,30	518,40
2021	4 384,10	43 498,90	496,40

6	4792,1566	-1025,6566
7	4082,9367	227,66325
8	3797,6242	1006,1758
9	1143,7182	659,68182
10	4451,5462	-67,446238

PODSUMOWANIE - WYJŚCIE

Statystyki regresji	
Wielokrotność R	0,851789768
R kwadrat	0,72554581
Dopasowany R kwadrat	0,647130327
Błąd standardowy	1996,086979
Obserwacje	10

W naszym przypadku R^2 jest wartością bardzo wysoką, bo jest to ponad 72%. Oznacza to, że model ma wysoki stopień

W naszym przypadku R^2 jest dość niską, bo jest to ponad 72%. Oznacza to, że modelem wyjaśnienia zmienności zmiennej objaśnianej ale niestety występuje w nim jedna zmienna statystycznie nieistotna (X_3).

Ponieważ model z jedną zmienną byłby zbyt mało reprezentatywny zbadałem inne modele oparte o moje zmienne kóre pozostały po doborze z udziałem współczynnika zmienności.

ANALIZA WARIANCJI					
	df	SS	MS	F	Istotność F
Regresja	2	73731307,5	49687817	7,7712895	0,01725
Resztkowy	7	27890542,58	6393767,4		
Razem	9	101621850,1			

	Wsółczynniki	Błqđ standardowy	t Stat	Wartość-p	Dołne 95%	Górne 95%	Dołne 95,0%	Górne 95,0%
Przecięcie	-1,7920438	1563,340001	-0,020336	0,9843429	-3728,5	3664,92	-3728,50372	3664,91964
x3	0,03447207	0,031230923	1,1039446	0,3061146	-0,3794	0,10833	-0,03937219	0,1083266
x4	0,87385067	2,159173984	4,040358	0,049306	3,61821	13,8295	3,618207749	13,8294939



SKŁADNIKI RESZTOWE - WYJŚCIE

Observacja			Przewidywane y			Składniki resztowe		
	1	8737,202327		460,1976728				
	2	7975,937618		3370,862382				
	3	11921,91257		-518,1125724				
	4	6987,56368		-1506,86368				
	5	3477,370908		555,0290917				
	6	4851,414831		-1084,914831				
	7	2810,182601		1500,417399				
	8	3599,151094		1204,648906				
	9	4370,316347		-2566,916347				
	10	5798,44802		-1414,34802				

Model opraty o zmienne X1,X2,X3 X4:

PODSUMOWANIE - WYJŚCIE

Statystyki regresji		
Wielokrotność R	0,955351648	
R kwadrat	0,912696771	
Dopasowany R kwadrat	0,842854188	
Błąd standardowy	1332,05973	
Observacje	10	

ANALIZA WARIANCJI

	df	SS	MS	F	Istotność F
Regresja	4	92749934,46	23187484	13,067913	0,00739
Resztkowy	5	8871915,624	1774383,1		
Razem	9	101621850,1			

	Współczynniki	Błąd standardowy	t Stat	Wartość-p	Dolne 95%	Górne 95%	Dolne 95,0%	Górne 95,0%
Przecięcie	-382,086863	1094,255792	-0,3491751	0,7411821	-3195	2430,79	-3194,96092	2430,7872
X1	-0,06210244	0,127514277	-0,4870234	0,6468408	-0,3899	0,26568	-0,38988832	0,26568344
X2	-0,693533247	0,214099169	-3,2393084	0,0229669	-1,2439	-0,1432	-1,24389268	-0,1431738
X3	0,702158525	0,206308874	3,4034335	0,0191794	0,17182	1,23249	0,17182468	1,23249237
X4	2,654435564	2,474124692	1,0728787	0,332354	-3,7055	9,01438	-3,70550443	9,01437556

W modelu powyżej występują zmienne statystycznie nieistotne a zatem najpierw odrzucam X1:

PODSUMOWANIE - WYJŚCIE

Statystyki regresji		
Wielokrotność R	0,953181645	
R kwadrat	0,908555249	
Dopasowany R kwadrat	0,862832873	
Błąd standardowy	1244,506916	
Observacje	10	

ANALIZA WARIANCJI

	df	SS	MS	F	Istotność F
Regresja	3	92329065,3	30776355	19,87113	0,00161
Resztkowy	6	9292784,784	1548797,5		
Razem	9	101621850,1			

	Współczynniki	Błąd standardowy	t Stat	Wartość-p	Dolne 95%	Górne 95%	Dolne 95,0%	Górne 95,0%
Przecięcie	-223,9306092	976,2765698	-0,2293721	0,8261987	-2612,8	2164,93	-2612,79332	2164,932
X2	-0,693135949	0,200025537	-3,4652373	0,0133814	-1,1826	-0,2037	-1,18258081	-0,203691
X3	0,698546472	0,192624142	3,6264741	0,0110115	0,22721	1,16988	0,22721218	1,1698807
X4	2,413141773	2,264686489	1,0655522	0,327615	-3,1283	7,95463	-3,12834644	7,9546299

Jedna ze zmiennych objaśniających jest nieistotna (X4) i ją odrzucam:

PODSUMOWANIE - WYJŚCIE

Statystyki regresji		
Wielokrotność R	0,944060819	
R kwadrat	0,89125083	
Dopasowany R kwadrat	0,860179638	
Błąd standardowy	1256,485568	
Observacje	10	

ANALIZA WARIANCJI

	df	SS	MS	F	Istotność F
Regresja	2	90570558,2	45285279	28,684154	0,00042
Resztkowy	7	11051291,88	1578756		
Razem	9	101621850,1			

ANALIZA WARIANCJI									
	df	SS	MS	F	Istotność F				
Regresja	2	90570558,2	45285279	28,684154	0,00042				
Resztkowy	7	11051291,88	1578756						
Razem	9	101621850,1							
	Współczynniki	Błąd standardowy	t Stat	Wartość-p	Dolne 95%	Górne 95%	Dolne 95,0%	Górne 95,0%	
Przecięcie	281,906444	861,2988204	0,3273039	0,7530086	-1754,7	2318,55	-1754,74163	2318,55452	
X2	-0,864530603	0,120044902	-7,2017269	0,0001772	-1,1484	-0,5807	-1,14839169	-0,5806695	
X3	0,863872948	0,115251172	7,4955675	0,0001378	0,59135	1,1364	0,59134723	1,13639867	
SKŁADNIKI RESZTOWE - WYJŚCIE									
Obserwacja	Przewidywane Y	Składniki resztowe							
1	8714,173029	483,2269708							
2	11100,49484	246,3051616							
3	9962,267642	1441,532358							
4	7890,650899	-2409,950899							
5	4593,931625	-561,5316251							
6	4792,156561	-1025,656561							
7	4082,936749	227,6632509							
8	3797,624234	1006,175766							
9	1143,718183	659,6818172							
10	4451,546238	-67,44623829							
Model oszacowany w oparciu o zmienne X2 X3 ma obie zmienne statystycznie istotne i bardzo dobre dopasowanie wynoszące 89,1%									
Równanie tego modelu prezentuje się następująco:									
$\hat{y}_t = 281,9 - 0,865X_{2t} + 0,864X_{3t}$									

Do modelu oszacowanego zostają zmienne X2 i X3, czyli

- Niezrealizowany zysk/strata z wyceny OFE,
- Wynik finansowy OFE.

A jego postać wygląda następująco:

$$\hat{y}_t = 281,9 - 0,865X_{2t} + 0,864X_{3t}$$

5. Metoda KMNK i jej implementacja

Metoda k-minimalnych niezależnych kombinacji (KMNK) jest metodą ekonometryczną do wyboru zmiennych objaśniających w modelu regresji liniowej. Polega na iteracyjnym usuwaniu najmniej istotnych zmiennych i dobieraniu k-minimalnych kombinacji, które najlepiej objaśniają zmienną zależną. Celem jest znalezienie najlepszego zestawu zmiennych, które zapewniają najlepszą jakość prognoz.

Szacowanie modelu KMNK na macierzach											
$\epsilon =$	1	-20 758,50	-11 013,30	$y =$	9 197,40	XT	1	1	1	1	1
	1	24 884,30	37 426,60		11 346,80		-20 758,50	24 884,30	8 709,50	-8 948,50	-11 827,20
	1	8 709,50	19 921,90		11 403,80		23 165,20	-21 394,90	-2 648,50	-4 484,50	38 642,80
	1	-8 948,50	-147,60		5 480,70		-11 013,30	37 426,60	19 921,90	-147,60	-6 844,70
	1	-11 827,20	-6 844,70		4 032,40		28 403,80	-17 011,20	1 419,20	-3 490,30	43 498,90
	1	23 165,20	28 403,80		3 766,50						
	1	-21 394,90	-17 011,20		4 310,60						
	1	-2 648,50	1 419,20		4 803,80						
	1	-4 484,50	-3 490,30		1 803,40						
	1	38 642,80	43 498,90		4 384,10						
$KTX =$	10	25339,7	92163,3	$XTX^{-1} =$	0,469886205	5,64629E-05	-5,5658E-05	XTY	60529,5	$a =$	281,90644
	25339,7	3860716932	4130483976		5,64629E-05	9,12793E-09	-8,6361E-09		2,38E+08		-0,8645306
	92163,3	4130483976	4968304372		-5,5658E-05	-8,6361E-09	8,41348E-09		7,47E+08		0,8638729

Parametr a_k wyznaczony w procesie regresji linowej określa, jak zmiana jednej jednostki zmiennej objaśniającej x_k wpłynie na zmienną objaśnianą Y . Oznacza to, że gdy wartość x_k wzrośnie o 1 jednostkę, to wartość Y zmieni się o tyle samo, ile wynosi oszacowany parametr a_k , przy założeniu, że wartości pozostałych zmiennych pozostaną bez zmian.

6. Weryfikacja modelu

Sprawdzenie poprawności modelu ekonometrycznego to krok, w którym weryfikujemy, czy nasz model jest odpowiednio dopasowany do danych, którymi się zajmujemy. Ta czynność jest bardzo ważna, ponieważ pozwala ocenić, jak skuteczny i trafny jest nasz model w prognozowaniu zachowań rynków czy gospodarek.

WERYFIKACJA MODELU								
t	x	y	x ²	x ³	\hat{y}	$e=(y-\hat{y})$		
1	1	9197,4	#####	-11 013,30	8714,17	483,227	9887565,803	7082108,01
2	1	11346,8	24 884,30	37 426,60	11100,5	246,305	28024847,82	25477708,9
3	1	11403,8	8 709,50	19 921,90	9962,27	1441,53	28631595,72	15282764,4
4	1	5480,7	-8 948,50	-147,60	7890,65	-2409,95	327470,0625	3377144,59
5	1	4032,4	#####	-6 844,70	4593,93	-561,532	4082622,303	2128734,62
6	1	3766,5	23 165,20	28 403,80	4792,16	-1025,66	5227853,603	1589600,09
7	1	4310,6	#####	-17 011,20	4082,94	227,663	3035783,523	3880952,21
8	1	4803,8	-2 648,50	1 419,20	3797,62	1006,18	1560375,723	5086494,31
9	1	1803,4	-4 484,50	-3 490,30	1143,72	659,682	18058675,2	24100557
10	1	4384,1	38 642,80	43 498,90	4451,55	-67,4462	2785060,323	2564494,01
						0,00	101621850,1	90570558,2
a =	281,906					SSE	SST	SSR
	-0,86453							
	0,86387							
wyznaczono								
modelu								
20,8%								

Wyrzistość modelu to stopień, w jakim model jest dobrze zdefiniowany i łatwy do zrozumienia. Oznacza to, że model jest prosty i logiczny, a jego parametry są łatwo interpretowalne. Model jest uważany za wyraźny, jeśli jest zbudowany na podstawie jasno określonych założeń i hipotez oraz jest zgodny z danymi i zachowaniem rynków czy gospodarek. Wysoka wyrzistość modelu pozwala na łatwiejsze wyjaśnienie i interpretację jego wyników oraz zwiększa jego użyteczność w prognozowaniu i podejmowaniu decyzji.

Wyrzistość w naszym projekcie wynosi 20,8%

- SSE, czyli suma kwadratów błędów – 11051291,9
- SST, czyli całkowita suma kwadratów – 101621850,1
- SSR, czyli regresyjna suma kwadratów – 90570558,2

Współczynnik determinacji modelu

Statystyki regresji	0
Wielokrotność R	0,944060819
R kwadrat	0,89125083
Dopasowany R kwadrat	0,860179638
Błąd standardowy	1256,485568
Obserwacje	10
współczynnik indeterminacji	0,10874917

Model ekonometryczny w projekcie ma wartość R kwadrat równą 0,8912, czyli 89,12% co oznacza naprawdę dobre dopasowanie modelu

Współczynnik zbieżności

Współczynnik zbieżności w ekonometrii to miernik, który określa szybkość, z jaką algorytm regresji linowej dochodzi do optymalnego rozwiązania. Współczynnik zbieżności odpowiada za monitorowanie postępu algorytmu i umożliwia jego zatrzymanie w momencie, gdy wyznaczone parametry osiągną stabilne i akceptowalne wartości. Im większy jest współczynnik zbieżności, tym szybciej algorytm dochodzi do optymalnego rozwiązania i tym mniejsze jest ryzyko, że algorytm będzie wprowadzać niepotrzebne i błędne zmiany do modelu. Wysoki współczynnik zbieżności to wyznacznik dobrej jakości modelu i jego skuteczności w prognozowaniu i analizie zjawisk ekonomicznych.

Oblicza się go za pomocą wzoru:

$$1 - R^2$$

Zatem w naszym projekcie wynosi 0,1088 czyli podobieństwo między skupieniami jest naprawdę duże.

8. Istotność statystyczna F

Jest to miar, która służy do oceny znaczenia zbioru zmiennych objaśniających w modelu regresji. Test istotności F polega na porównaniu wariancji zmiennych objaśniających ze składnikiem losowym w modelu. Jeśli wartość testu jest wystarczająco duża, oznacza to, że zmienna objaśniająca jest istotna statystycznie i ma istotny wpływ na zmienną objaśnianą. Przeciwnie, jeśli wartość testu jest niska, oznacza to, że zmienna objaśniająca nie jest istotna statystycznie i nie ma znaczącego wpływu na zmienną objaśnianą. Test istotności F jest często stosowany w ekonometrii i statystyce, aby ocenić znaczenie zmiennych objaśniających w modelu i zoptymalizować jego strukturę.

PODSUMOWANIE - WYJŚCIE						
Wielokrotność R	0,9440608					
R kwadrat	0,8912508					
Dopasowany R kwadrat	0,8601796					
Błąd standardowy	1256,4856					
Obserwacje	10					
k	2					
n	10					
R ² /k	0,4456254					
1-R ² /n-k-1	0,0155356					
Femp=	28,684154					
α =	0,05					
s1 = k =	2					
s2 = n-k-1 =	7					
Aby uzyskać $F_{0,05,2,6}$ korzystamy z polecenia ROZKŁ.F.ODWR		$F_{emp} = \frac{R^2/k}{1-R^2/(n-k-1)}$		$F_{0,05,2,7} =$		4,73741

ANALIZA WARIANCJI					
	df	SS	MS	F	Istotność F
Regresja	2	9,1E+07	4,5E+07	28,6842	0,0004241
Resztkowy	7	1,1E+07	1578756		
Razem	9	1E+08			

Femp wynosi 28,644154

$F_{0,05,2,7}$ wynosi 4,73741

9. Test Jaeque-Bera

Test statystyczny, który służy do oceny, czy dana próbka danych pochodzi z rozkładu normalnego. Test opiera się na porównaniu dwóch estymatorów wariancji: estymowanej z danych i teoretycznej dla rozkładu normalnego. Jeśli wartość testu jest duża, oznacza to, że dana próbka danych nie pochodzi z rozkładu normalnego. W przeciwnym razie, jeśli wartość testu jest mała, oznacza to, że dana próbka danych pochodzi z rozkładu normalnego. Test Jaeque-Bera jest często stosowany w ekonometrii i statystyce, aby ocenić normalność rozkładu danych i zastosować odpowiednie metody analizy statystycznej.

	$e=(y-y^A)$	e^2	e^3	e^4
	483,227	233508,31	112837511	5,453E+10
	246,305	60666,233	14942406	3,68E+09
	1441,53	2078015,5	2,996E+09	4,318E+12
	-2410	5807863,3	-1,4E+10	3,373E+13
	-561,53	315317,77	-1,77E+08	9,943E+10
	-1025,7	1051971,4	-1,08E+09	1,107E+12
	227,663	51830,556	11799913	2,686E+09
	1006,18	1012389,7	1,019E+09	1,025E+12
	659,682	435180,1	287080399	1,894E+11
	-67,446	4548,9951	-306812,6	20693356
suma		11051292	-1,08E+10	4,053E+13
suma e^4^2		1,643E+27		
n		10		
kurtoza		1,5275701		
skośność		-1,1036323		
w		-4,007E-45		
k		4,095E-48		
JB		3,75		
rozkład χ^2		5,9914645		

Nie ma podstaw do odrzucenie hipotezy zerowej, więc składniki losowe modelu mają rozkład normalny).

W naszym przypadku statystyka jest mniejsza lub równa wartości krytycznej odczytanej z tablic dla 2 stopni swobody i poziomu istotności 0,05 wynoszącej 5,99, nie mamy podstaw do odrzucenia hipotezy zerowej – odchylenia losowe w modelu mają rozkład normalny.

10. Badanie autokorelacji składników losowych

Badanie autokorelacji składników losowych to proces, w którym sprawdza się, czy wartości składników losowych z czasem są skorelowane. Autokorelacja polega na tym, że wartość składnika losowego w danym momencie jest skorelowana z wartością tego samego składnika losowego w poprzednim okresie. W ekonometrii autokorelacja może wystąpić w modelach z czasem i może mieć wpływ na jakość prognoz. Dlatego ważne jest, aby sprawdzić autokorelację składników losowych i ewentualnie zastosować odpowiednie korekty, takie jak model ARIMA, aby uniknąć błędów w prognozowaniu. W tym celu można użyć narzędzi takich jak współczynnik autokorelacji i współczynnik dwustronnej autokorelacji.

e=(y-y^)	e-1	et-e(t-1)							
483,2269708			n	k = 1		k = 2		k = 3	
246,3051616	483,227	-236,92		<i>d_l</i>	<i>d_g</i>	<i>d_l</i>	<i>d_g</i>	<i>d_l</i>	<i>d_g</i>
1441,532358	246,3052	1195,23							
-2409,950899	1441,532	-3851,5							
-561,5316251	-2409,951	1848,42							
-1025,656561	-561,5316	-464,12	6	0,61	1,40				
227,6632509	-1025,657	1253,32	7	0,70	1,36	0,47	1,90		
1006,175766	227,6633	778,513	8	0,76	1,33	0,56	1,78	0,37	2,29
659,6818172	1006,176	-346,49	9	0,82	1,32	0,63	1,70	0,45	2,13
-67,44623829	659,6818	-727,13	10	0,88	1,32	0,70	1,64	0,53	2,02
			11	0,93	1,32	0,76	1,60	0,59	1,93
			12	0,97	1,33	0,81	1,58	0,66	1,86
			13	1,01	1,34	0,86	1,56	0,71	1,82
			14	1,04	1,35	0,91	1,55	0,77	1,78
			15	1,08	1,36	0,95	1,54	0,81	1,75
			16	1,11	1,37	0,98	1,54	0,86	1,73
			17	1,13	1,38	1,02	1,54	0,90	1,71
			18	1,16	1,39	1,05	1,54	0,93	1,70
suma kwadratów różnic	22776355								
suma kwadratów reszt	11051292								
n	10								
k	2								
DI	0,7								
Dg	1,64								
4-Dg	2,36								
4-DI	3,3								
Stat. DW	2,060968								
Ponieważ statystka zawiera się w obszarze Dg<DW<4-Dg stwierdzamy brak autokorelacji składnika losowego co jest zjawiskiem pozytywnym.									

Ponieważ statystyka zawiera się w obszarze $Dg < DW < 4-Dg$ stwierdzamy brak autokorelacji składnika losowego co jest zjawiskiem pozytywnym

Test Durбина-Watsona jest narzędziem używanym do wykrywania autokorelacji w składnikach losowych modelu ekonometrycznego. Autokorelacja oznacza zależność między wartościami składnika losowego w kolejnych okresach czasowych. Jeśli autokorelacja występuje, prognozy oparte na modelu mogą być niedokładne.

Test Durбина-Watsona jest oparty na obliczeniu współczynnika autokorelacji dwustronnej, który jest interpretowany jako wartość bezwzględna różnicy między wartościami składnika losowego w kolejnych okresach czasowych. Współczynnik ten jest porównywany z kryterium krytycznym, aby stwierdzić, czy istnieje autokorelacja. Współczynnik Durбина-Watsona zazwyczaj zawiera się w przedziale (0, 4), gdzie wartość bliska 0 oznacza silną autokorelację, a wartość bliska 4 oznacza brak autokorelacji.

Jeśli wynik testu wskazuje na istnienie autokorelacji, należy zastosować odpowiednie korekty, takie jak model ARIMA, aby uniknąć błędów w prognozowaniu.

11. Badanie koincydencji

Badanie koincydencji jest narzędziem używanym w ekonometrii do oceny wzajemnej zależności między dwoma zmiennymi. Polega ono na porównaniu rozkładów dwóch zmiennych i określeniu, czy są one ze sobą powiązane. W badaniu koincydencji stosuje się wskaźniki korelacji, takie jak współczynnik korelacji Pearsona lub współczynnik korelacji Spearmana, aby zmierzyć stopień powiązania między dwoma zmiennymi.

Badanie koincydencji jest często stosowane w ekonometrii w celu identyfikacji wzajemnych zależności między zmiennymi, co jest kluczowe w modelowaniu i prognozowaniu zjawisk ekonomicznych. Może być również używane do identyfikacji zmiennych, które mają istotny wpływ na inne zmienne, co pozwala na ulepszanie modeli i prognoz.

rok	y	x2	x3						
2011	9197,4	-20759	-11013						
2012	11346,8	24884,3	37426,6						
2013	11403,8	8709,5	19921,9						
2014	5480,7	-8948,5	-147,6						
2015	4032,4	-11827	-6844,7						
2016	3766,5	23165,2	28403,8						
2017	4310,6	-21395	-17011						
2018	4803,8	-2648,5	1419,2						
2019	1803,4	-4484,5	-3490,3						
2020	4384,1	38642,8	43498,9						

	y		Współczynnik
y	1	Przecięcie	281,906
x2	0,13567	x3	-0,86453
x3	0,2924	x5	0,86387

13. VIF

Test VIF (Variance Inflation Factor) jest jednym ze sposobów oceny kolizji (czyli wzajemnej zależności) między zmiennymi objaśniającymi w modelu regresji. VIF mierzy, jak bardzo zmienna objaśniająca zwiększa wariancję innych zmiennych objaśniających w modelu, co może prowadzić do problemów zinterpretowania wyników i oceny istotności statystycznej.

VIF jest obliczany dla każdej zmiennej objaśniającej w modelu regresji, a wartość VIF jest bezpośrednio proporcjonalna do poziomu kolizji między tą zmienną a pozostałymi zmiennymi w modelu. Zasadniczo, im większa wartość VIF, tym większa jest kolizja między zmienną objaśniającą a pozostałymi zmiennymi.

y	X2	X3
9197,4	-20759	-11013
11346,8	24884,3	37426,6
11403,8	8709,5	19921,9
5480,7	-8948,5	-147,6
4032,4	-11827	-6844,7
3766,5	23165,2	28403,8
4310,6	-21395	-17011
4803,8	-2648,5	1419,2
1803,4	-4484,5	-3490,3
4384,1	38642,8	43498,9

PODSUMOWANIE - WYJŚCIE					
Statystyki regresji		0			
Wielokrotność R		0,94406			
R kwadrat		0,89125			
Dopasowany R kwadrat		0,86018			
Błąd standardowy		1256,49			
Obserwacje		10			

ANALIZA WARIANCJI					
	df	SS	MS	F	Istotność F
Regresja	2	9,1E+07	4,5E+07	28,6842	0,00042
Resztkowy	7	1,1E+07	1578756		
Razem	9	1E+08			

	Współcz.	Błąd stan.	t Stat	Wartość	Dołne 95%	Górne 95%	Dołne 95%	Górne 95%
Przecięcie	281,906	861,299	0,3273	0,75301	-1,754,7	2318,55	-1,754,7	2318,55
X2	-0,8645	0,12004	-7,2017	0,00018	-1,1484	-0,5807	-1,1484	-0,5807
X3	0,86387	0,11525	7,49557	0,00014	0,59135	1,1364	0,59135	1,1364

X2		X3	
Średnia	2533,97	Średnia	9216,33
Błąd standardowy	6494,88	Błąd stan.	6765,02
Mediana	-3566,5	Mediana	635,8
Tryb	#N/D	Tryb	#N/D
Odchylenie standardowe	20538,6	Odchylenie	21392,9
Wariancja próbki	4,2E+08	Wariancja	4,6E+08
Kurtoza	-0,8836	Kurtoza	-1,3091
Skośność	0,57652	Skośność	0,52805
Zakres	60037,7	Zakres	60510,1
Minimum	-21395	Minimum	-17011
Maksimum	38642,8	Maksimum	43498,9
Suma	25339,7	Suma	92163,3
Licznik	10	Licznik	10

Dla VIF_x3 - występuje bardzo nieznaczna współliniowość predyktorów, więc warto rozważyć zmianę modelu

Dla VIF_x5 - występuje bardzo nieznaczna współliniowość predyktorów, więc warto rozważyć zmianę modelu

14. Test losowości reszt (składnika losowego)-test serii

Test losowości reszt jest stosowany do sprawdzenia, czy składnik losowy modelu ekonometrycznego jest losowy. W testach tego typu bada się zgodność rozkładu reszt z rozkładem normalnym (lub jakimkolwiek innym z góry założonym rozkładem). W przypadku, gdy rozkład reszt nie jest losowy, sugeruje to, że model nie jest odpowiedni dla danych, lub że występują jakieś braki w modelu, takie jak brak uwzględnienia ważnej zmiennej objaśniającej.

Test losowości reszt (składnika losowego)-test serii

Hipotezy dla testu serii mają postać:

H_0 : reszty wykazują losowość,

H_1 : reszty nie wykazują losowości.

e	serie	reszta +	reszta -	liczba serii
483,23	A	0	1	1
246,31	A	0	1	
1441,53	A	0	1	
-2409,95	B	1	0	1
-561,53	B	1	0	
-1025,66	B	1	0	
227,66	A	0	1	1
1006,18	A	0	1	
659,68	A	0	1	
-67,45	B	1	0	1
Suma=		4	6	4

Liczba reszt dodatnich $n_+ =$ 4

Liczba reszt ujemnych $n_- =$ 6

Liczba serii $s =$ 4

Przyjmę poziom istotności testu :

$\alpha = 0,05$ stąd $\gamma = \alpha/2 = 0,05/2 = 0,025$

Z tablicy niżej odczytuję wartość krytyczną $s^*(\alpha/2; n_+; n_-) = s^*(0,025; 4; 6) =$ 2

Ponieważ nie spełniona jest nierówność $s \leq s^*(\alpha/2; n_+; n_-)$, brak podstaw do odrzucenia hipotezy H_0

Wartość $\gamma = 1 - \alpha/2 = 0,975$

Z tablicy niżej odczytujemy wartość krytyczną $s^*(1-\alpha/2; n_+; n_-) = s^*(0,975; 4; 6) =$ 8

Spełniona jest nierówność $s^*1 < s < s^*2$, zatem brak podstaw do odrzucenia hipotezy H_0 .

Z testu wynika, że reszty są losowe.



Rozkład warunkowy liczby serii:

$\omega=0,025$	n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
2	2																														
3	2		3																												
4	2		2	2																											
5	2		2	2	2																										
6	2		2	2	2	3																									
7	2		2	2	2	3	3																								
8	2		2	2	2	3	3	4																							
9	2		2	2	2	3	3	4	4																						
10	2		2	2	2	3	3	4	5	5																					
11	2		2	2	2	3	3	4	5	5	6																				
12	2		2	2	2	3	3	4	5	6	6	7																			
13	2		2	2	2	3	3	4	5	6	6	7	7																		
14	2		2	2	2	3	3	4	5	6	7	7	8	8																	
15	2		2	2	2	3	3	4	5	6	7	8	8	9	9																
16	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10															
17	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11														
18	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12											
19	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13									
20	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14							
21	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15					
22	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16			
23	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	
24	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18
25	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18
26	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18
27	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18
28	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18
29	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18
30	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18

$\omega=0,075$	n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
2	2																														
3	2		3																												
4	2		2	2																											
5	2		2	2	2																										
6	2		2	2	2	3																									
7	2		2	2	2	3	3																								
8	2		2	2	2	3	3	4																							
9	2		2	2	2	3	3	4	4																						
10	2		2	2	2	3	3	4	5	5																					
11	2		2	2	2	3	3	4	5	6	6																				
12	2		2	2	2	3	3	4	5	6	7	7																			
13	2		2	2	2	3	3	4	5	6	7	8	8																		
14	2		2	2	2	3	3	4	5	6	7	8	9	9																	
15	2		2	2	2	3	3	4	5	6	7	8	9	9	10																
16	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10															
17	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	10														
18	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
19	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
20	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
21	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
22	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
23	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
24	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
25	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
26	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
27	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
28	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
29	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17
30	2		2	2	2	3	3	4	5	6	7	8	9	9	10	10	11	11	11	12	12	13	13	14	14	15	15	16	16	17	17

16. Test t-Studenta

Test t-Studenta jest używany do oceny, czy średnia populacji jest różna od określonej wartości hipotezy. Polega on na porównaniu estymatora średniej z wartością hipotezy i obliczeniu statystyki t , która jest funkcją odchylenia standardowego estymatora. Na podstawie wartości statystyki t i liczby stopni swobody można określić p-wartość, która odzwierciedla prawdopodobieństwo uzyskania takiego wyniku przez przypadek. Jeśli p-wartość jest mniejsza niż poziom istotności, wyklucza się hipotezę zerową i stwierdza, że istnieje różnica między średnią populacji a wartością hipotezy.

	Współczynniki Błąd standardowy		t Stat	Wartość-p
X2	-0,8645306	0,120044902	-7,2017269	0,00017719
X3	0,86387295	0,115251172	7,49556755	0,000137817
		t kryt=	2,36462425	

Statystyki testowe t_1 i t_2 są większe od wartości krytycznej (względem wartości bezwzględnej). Zmienne objaśniające x_1 i x_2 w sposób istotny wpływają na zmienną objaśnianą.

17. Podsumowanie

Projekt pomógł nam poszerzyć wiedzę z ekonometrii, rozkładów, istotności statystycznej i wszelakich testów, co uświadomiło nam jak ważna jest zarówno statystyka jak i znajomość m.in. programu MS Excel w pracy np. analityka. Ponadto można powiedzieć, że nasz model przeszedł wszystkie testy pomyślnie, co udowadnia, że dobrane dane były niejako odpowiednie.