



Model ekonometryczny PKB Indii w latach 1983-2020

Dane

y gdp lg	X1Patent applications, nonresidents	X2Population ages 05-09, female (% of female population)	X3Survival to age 65 male	X4domestic companies, t	X5Imports of goods and services (% of GDP)=x2	X6Exports of goods and services (% of GDP)	X7 Total reserves (includes gold, current US\$)	X8 x%(inflation, consumer prices (annual %)	X9Population, total	X10Population growth (annual %)	X11Life expectancy at birth, total (years)		
	261086227	1,76	3933	621	1,13	7,85	5,84	8,22E+09	11,87	7,46E+08	2,28	54,91	1983
	261086923	1,19	3871	634	1,30	7,73	6,28	8,54E+09	8,32	7,63E+08	2,29	55,27	1984
	261103713	1,49	3879	6311	1,09	7,65	5,25	9,49E+09	5,56	7,8E+08	2,27	55,62	1985
	261086349	1,10	3871	6340	1,91	7,02	5,2	1,05E+10	8,73	7,98E+08	2,26	55,98	1986
	261087918	1,13	3838	6413	1,08	6,98	5,6	1,15E+10	8,8	8,16E+08	2,24	56,33	1987
	2610381314	1,10	384	6309	1,10	7,46	6,04	9,19E+09	9,38	8,34E+08	2,21	56,69	1988
	261079827	1,00	3819	6339	1,07	8,15	7,02	8,05E+09	7,07	8,52E+08	2,19	57,18	1989
	261086123	1,07	3737	634	1,05	8,45	7,05	5,64E+09	8,97	8,7E+08	2,16	57,66	1990
	261087987	1,13	3737	637	1,08	8,49	8,49	7,62E+09	13,87	8,89E+08	2,12	58,15	1991
	261086477	1,18	3734	6839	1,70	9,59	8,84	9,54E+09	11,79	9,08E+08	2,1	58,63	1992
	261033387	1,11	3739	6940	1,30	9,82	9,83	1,47E+10	6,33	9,26E+08	2,07	59,12	1993
	261049823	1,11	3699	693	1,03	10,19	9,89	2,42E+10	10,25	9,45E+08	2,04	59,59	1994
	261010274	1,01	3669	6113	1,08	12,02	10,84	2,29E+10	10,22	9,64E+08	2,01	60,06	1995
	261081347	1,70	3633	69	1,09	11,54	10,39	2,49E+10	8,98	9,83E+08	1,97	60,53	1996
	261033313	1,13	3637	6860	1,05	11,93	10,69	2,84E+10	7,16	1E+09	1,94	61	1997
	261067219	1,71	3537	6340	1,70	12,68	11,02	3,06E+10	13,23	1,02E+09	1,91	61,47	1998
	261019473	1,03	3519	6409	1,70	13,36	11,45	3,6E+10	4,67	1,04E+09	1,87	61,88	1999
	261073764	1,13	3473	6440	1,80	13,9	13	4,11E+10	4,01	1,06E+09	1,84	62,28	2000
	261083363	1,11	3436	6339	1,70	13,43	12,56	4,91E+10	3,78	1,08E+09	1,82	62,69	2001
	2610731214	1,17	3397	6430	1,40	15,24	14,26	7,16E+10	4,3	1,1E+09	1,79	63,09	2002
	27112964	1,19	3336	6440	1,40	15,64	14,95	1,04E+11	3,81	1,12E+09	1,74	63,5	2003
	271083381	1,14	3311	6499	1,70	19,64	17,86	1,32E+11	3,77	1,14E+09	1,69	63,91	2004
	271083363	1,14	3279	6738	1,70	22,4	19,61	1,38E+11	4,25	1,15E+09	1,62	64,31	2005
	271083373	1,13	3238	6814	1,70	24,46	21,27	1,78E+11	5,8	1,17E+09	1,54	64,72	2006
	271071923	1,810	3201	6871	1,80	24,89	20,8	2,77E+11	6,37	1,19E+09	1,48	65,12	2007
	271010219	1,10	3144	6940	1,90	29,27	24,1	2,57E+11	8,35	1,21E+09	1,43	65,53	2008
	271089763	1,11	3134	6960	1,11	25,87	20,4	2,85E+11	10,88	1,22E+09	1,4	65,98	2009
	281473119	1,811	3081	704	1,00	26,85	22,4	3E+11	11,99	1,24E+09	1,39	66,43	2010
	28101023	1,14	304	714	1,10	31,08	24,54	2,99E+11	8,91	1,26E+09	1,37	66,87	2011
	281030349	1,140	3033	724	1,10	31,26	24,53	3E+11	9,48	1,27E+09	1,34	67,32	2012
	281083373	1,14	2943	734	1,10	28,41	25,43	2,98E+11	10,02	1,29E+09	1,31	67,77	2013
	281030311	1,10	2833	73	1,10	25,95	22,97	3,25E+11	6,67	1,31E+09	1,25	68,07	2014
	28108649	1,10	2840	7340	1,80	22,11	19,81	3,53E+11	4,91	1,32E+09	1,19	68,37	2015
	281030313	1,10	2733	7330	1,10	20,92	19,16	3,62E+11	4,95	1,34E+09	1,19	68,67	2016
	281081363	1,143	2748	7431	1,610	21,95	18,79	4,13E+11	3,33	1,35E+09	1,16	68,97	2017
	281030738	1,177	2733	7438	1,08	23,69	19,93	3,99E+11	3,94	1,37E+09	1,09	69,27	2018
	2810734417	1,177	2643	7438	1,10	21,27	18,69	4,63E+11	3,73	1,38E+09	1,03	69,5	2019
	281030319	1,143	2616	7337	1,10	19,1	18,71	5,9E+11	6,62	1,4E+09	0,96	69,73	2020

Opis danych

		Zlogarytmowane naturalnie w dolarach Amerykańskich
Y	Pkb	
X1	liczba zgłoszeń patentowych	w tys sztuk
X2	Damska populacja	%
X3	Męska populacja	%
X4	Krajowe firmy	tysiącach sztuk
X5	import	% GDP
X6	export	% GDP
X7	Rezerwy krajowe	US\$
X8	Inflacja	%
X9	Populacja	w szt
X10	Wzrost populacji	%
X11	Oczekiwana dlugosc zycia	lata

Źródło:

<https://www.kaggle.com/datasets/nejilee/indian-economy-from-1960-to-2020>

[India - Patent applications](#)

Eliminacja quasi-stałych.

$$V^* = 0,1$$

H0: zmienna jest quasi-stała

H1: zmienna nie jest quasi-stała

Odpowiedź: Odrzucamy x_{11} , ponieważ V_{11} jest mniejsze od V^* , stąd wiemy, że zmienna x_{11} jest quasi-stała.

Przyjmujemy H0. Dla zmiennych, $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$ odrzucamy H0 na rzecz hipotezy h1, stąd wiemy że te zmienne nie są quasi-stałe.

Dobór zmiennych

Zastosowana metoda doboru zmiennych: Regresja krokowa wstecz.

$H_0: \beta_4 = 0$

$H_1: \beta_4 \neq 0$

Brak podstaw do odrzucenia H_0 , β_4 jest istotne statystycznie

Przyjmujemy model z 4 zmiennymi objaśniającymi.

PODSUMOWANIE - WYJŚCIE								
Statystyki regresji								
Wielokrotność R	0,997148							
R kwadrat	0,994303							
Dopasowany R kwadrat	0,993613							
Błąd standardowy	0,070491							
Obserwacje	38							
ANALIZA WARIANCJI								
	df	SS	MS	F	Istotność F			
Regresja	4	28,62037	7,155092	1439,969	1,62E-36			
Resztkowy	33	0,163974	0,004969					
Razem	37	28,78434						
	Współczynniki	Błąd standardowy	t Stat	Wartość-p	Dolne 95%	Górne 95%	Dolne 95,0%	Górne 95,0%
Przecięcie	32,12811	0,48514	66,22437	1,12E-36	31,14109	33,11514	31,14109	33,11514
Zmienna X 1	0,019496	0,004569	4,267011	0,000157	0,010201	0,028792	0,010201	0,028792
Zmienna X 2	-0,15352	0,012623	-12,1619	9,72E-14	-0,1792	-0,12784	-0,1792	-0,12784
Zmienna X 3	0,029662	0,013832	2,144534	0,039447	0,001522	0,057803	0,001522	0,057803
Zmienna X 4	-0,03421	0,016136	-2,1202	0,041594	-0,06704	-0,00138	-0,06704	-0,00138

Wykrywanie zmiennych nietypowych, dźwigniowych i wpływowych

Wnioski: brak obserwacji wpływowych, 2 obserwacje dźwigniowe ale nie wpływowe.

	reszty	leverage(dźwigniowy)	influence(wpływowy)	DFFITS
	u	$h \leq h_c = 1$	$u^*h/(1-h)$	
Kopiuuj (Ctrl+C)				
1960	-0,099864	0,116	-0,013084	-0,556
1961	-0,13313	0,104	-0,015429	-0,713
1962	-0,10285	0,156	-0,018985	-0,699
1963	-0,023752	0,129	-0,0035128	-0,137
1964	0,084705	0,109	0,010333	0,449
1965	0,11929	0,093	0,012187	0,588
1966	0,09631	0,073	0,0075821	0,405
1967	0,13417	0,068	0,0098138	0,559
1968	-0,014399	0,109	-0,0017547	-0,074
1969	-0,00257	0,058	-0,0001577	-0,009
1970	-0,051753	0,105	-0,0060737	-0,264
1971	0,038152	0,076	0,0031588	0,160
1972	0,02497	0,050	0,0013072	0,082
1973	0,024701	0,043	0,0011059	0,075
1974	-0,00098045	0,041	-4,1734e-05	-0,003
1975	-0,030619	0,062	-0,002032	-0,114
1976	0,065892	0,320*	0,031025	0,781
1977	-0,014787	0,114	-0,0019059	-0,079
1978	-0,073602	0,077	-0,0061601	-0,315
1979	-0,080941	0,113	-0,010292	-0,438
1980	0,025298	0,141	0,0041414	0,155
1981	0,014592	0,127	0,0021197	0,083
1982	-0,045706	0,078	-0,0038925	-0,195
1983	-0,038693	0,111	-0,004855	-0,204
1984	0,022721	0,146	0,0038824	0,142
1985	-0,094546	0,175	-0,020004	-0,692
1986	-0,0034985	0,160	-0,00066641	-0,023
1987	0,1163	0,104	0,013462	0,613
1988	0,035907	0,212	0,0096858	0,295
1989	-0,057936	0,232	-0,017491	-0,514
1990	0,036178	0,204	0,0092806	0,288
1991	0,072158	0,083	0,0065598	0,323
1992	-0,010388	0,144	-0,0017485	-0,064
1993	0,035159	0,150	0,0061998	0,225
1994	0,072013	0,182	0,016	0,535
1995	-0,029308	0,201	-0,00735	-0,230
1996	-0,027267	0,216	-0,0074974	-0,226
1997	-0,081925	0,320*	-0,038482	-0,981

''*'' oznacza dźwigniową obserwację (leverage point), $h(i) > 2(k+1)/n$

Kryterium Cross-validation = 0,21923

Badanie współliniowości

H0	Brak współliniowości modelu
H1	Występuje współliniowość modelu
$VIF_k > 10$	Odrzucamy H0 występuje współliniowość modelu
Wnioski: Występuje inflacja która zakłóca model	
VIF _k	27,88234

W wyniku testu wyszło że występuje współliniowość modelu zmienne objaśniające są skorelowane, co wpływa na jakość i wiarygodność wyników. Estymatory kmlk są nieobciążone ale mają dużą wariancję ich wartości mogą odbiegać od wartości prawdziwych w małych próbach. Trudno jednoznacznie określić wpływ poszczególnych zmiennych na zmienne objaśniana.

W związku z współliniowością należałoby połączyć silnie skorelowane zmienne

Badanie istotności parametrów

Model 1: Estymacja KMNK, wykorzystane obserwacje 1960-1997 (N = 38)

Zmienna zależna (Y): ygdplg

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	32,1283	0,485217	66,21	1,13e-036	***
X1Patentapplicat~	0,0194930	0,00456965	4,266	0,0002	***
X2Populationages~	-0,153526	0,0126250	-12,16	9,75e-014	***
X5Importsofgoods~	0,0296619	0,0138344	2,144	0,0395	**
X6Exportsofgoods~	-0,0342069	0,0161388	-2,120	0,0417	**

Metoda weryfikacji istotności testem t studenta.

$H_0: \beta_i = 0$

$H_1: \beta_i \neq 0$

We wszystkich odrzucamy H_0
, parametry są istotne.

Zgodność modelu z danymi empirycznymi

Współczynnik korelacji wielorakiej	0,997147431
Jak silnie wszystkie zmienne objaśniające wpływają razem na zmienna objaśniana - występuje bardzo silnie korelacja dodatnia. Bardzo wysoki związek liniowy między zmienną zależną a zmiennymi niezależnymi. Model regresji jest w stanie dosyć dokładnie przewidywać wartości zmiennej zależnej.	
odchylenie standardowy reszt	0,070491134
Wartości obserwowane różnią się średnio o 0,07049 od wartości wyliczonych przez model.	
współczynnik determinacji	0,994303
99% zmienności zmiennej objaśnianej zostało wyjaśnione przez model.	

Badanie losowości składnika losowego epsylon

H0 :

$$y1=b0+b1*x1+b2*x2+b3*x3+b4*x4$$

H1:

$$y1!=b0+b1*x1+b2*x2+b3*x3+b4*x4$$

Wykorzystujemy test serii

S= 17
n2 21 -ujemne
n1 17 -dodatnie

s1

ES 19,78947
sigma s 3,005647
Z -0,92808
t* 2

|z| < t*

Brak podstaw do odrzucenia h0

Postac modelu została dobrana poprawnie

Składnik epsylon jest losowy, składnik losowy zachowuje się zgodnie z założeniami kmnk. Model jest poprawnie wyspecyfikowany.

Badanie normalności składnika losowego

H0 składnik losowy ma rozkład normalny					
H1 składnik losowy nie ma rozkład normalnego					
Test Jarque Berry					
A	0,103748919				
K	2,371398765				
M3	2,94084E-05				
M4	4,41559E-05				
S	0,06568953				
n=38					
JB	0,693808538				
CHI^2*	5,991464547				
JB<CHI^2*					
brak podstaw do odrzucenia h0 składnik losowy ma rozkład normalny					
Wniosek: Składnik losowy modelu ma rozkład normalny, jest spełnione założenie kmnk, składnik losowy nie wykazuje systematycznych wzorców co sugeruje ze model jest dobrze dopasowany. Estymatory pozostają nieobciążone.					

Badanie autokorelacji

test Durбина-Watsona

H0: Brak autokorelacji

H1: występuje autokorelacja

```
Stat. Durбина-Watsona = 1,10446  
  
H1: dodatnia autokorelacja składnika losowego  
wartość p = 0,000100438  
H1: ujemna autokorelacja składnika losowego  
wartość p = 0,99982
```

D=1,10446

dL=1,26

dU=1,72

d < dL odrzucamy H0 występuje autokorelacja pierwszego rzędu.

Stwierdziwszy autokorelację składnika losowego musimy zastosować uogólnioną metodę najmniejszych kwadratów aby estymatory nie były obciążone.

```
Model 1: Estymacja Cochrane-Orcutt, wykorzystane obserwacje 1961-1997 (N = 37)  
Zmienna zależna (Y): ygdplg  
rho = 0,446584
```

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	32,3463	0,542089	59,67	2,21e-034	***
X1Patentapplicat~	0,0179524	0,00517070	3,472	0,0015	***
X2Populationages~	-0,158343	0,0141933	-11,16	1,46e-012	***
X5Importsofgoods~	0,0337685	0,0122893	2,748	0,0098	***
X6Exportsofgoods~	-0,0408274	0,0150158	-2,719	0,0105	**

Podstawowe statystyki dla danych quasi-różnicowanych (rho):

Suma kwadratów reszt	0,124366	Błąd standardowy reszt	0,062341
Wsp. determ. R-kwadrat	0,995466	Skorygowany R-kwadrat	0,994899
F(4, 32)	547,3231	Wartość p dla testu F	5,77e-29
Autokorel.reszt - rho1	0,007686	Stat. Durбина-Watsona	1,877074

Podstawowe statystyki dla oryginalnych danych:

Średn. arytm. zm. zależnej	27,28983	Odch. stand. zm. zależnej	0,872840
----------------------------	----------	---------------------------	----------

$$\hat{y} = 32,34 + 0,0179x_1 - 0,158x_2 + 0,033x_5 - 0,040x_6$$

Badanie homoskedastyczności

TEST WHITE

gretl: test LM - mnożnika Lagrange'a (heteroskedastyczność)

Test White'a na heteroskedastyczność reszt (zmiennosc wariacji resztowej)
Estymacja KMNK, wykorzystane obserwacje 1960-1997 (N = 38)
Zmienna zależna (Y): uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	0,910336	1,45628	0,6251	0,5381	
X1Patentapplicat~	-0,0153412	0,0234689	-0,6537	0,5198	
X2Populationages~	-0,0537513	0,0774552	-0,6940	0,4947	
X5Importsofgoods~	-0,0915030	0,0632371	-1,447	0,1614	
X6Exportsofgoods~	0,122289	0,0755619	1,618	0,1192	
sq_X1Patentappli~	-3,46527e-05	0,000109849	-0,3155	0,7553	
X2_X3	0,000353926	0,000636707	0,5559	0,5837	
X2_X4	0,000469898	0,000575675	0,8163	0,4227	
X2_X5	-0,000279981	0,000633750	-0,4418	0,6628	
sq_X2Populationa~	0,000777904	0,00103496	0,7516	0,4599	
X3_X4	0,00233564	0,00168862	1,383	0,1799	
X3_X5	-0,00301227	0,00194875	-1,546	0,1358	
sq_X5Importsofgo~	-0,00227172	0,00117834	-1,928	0,0663	*
X4_X5	0,00571162	0,00287585	1,986	0,0591	*
sq_X6Exportsofgo~	-0,00387833	0,00181361	-2,138	0,0433	**

Wsp. determ. R-kwadrat = 0,519353

Statystyka testu: $TR^2 = 19,735414$,
z wartością p = $P(\text{Chi-kwadrat}(14) > 19,735414) = 0,138706$

H_0 występuje homoskedastyczność

H_1 występuje heteroskedastyczność

p value > 0,05

przyjmujemy H_0 występuje homoskedastyczność

Wniosek W modelu występuje homoskedastyczność więc wszystkie wariancje składników losowych są stałe dla wszystkich obserwacji niezależnie od wartości zmiennej objaśniającej.

Jedno z kluczowych założeń kmnk zostało spełnione więc estymatory kmnk są efektywne czyli mają najmniejszą możliwą wariancję są liniowe nieobciążone i zgodne.

Wnioski

1.

Na podstawie przeprowadzonych testów można sformułować następujące wnioski:

Test normalności - nie odrzucono hipotezy o normalnym rozkładzie składnika losowego co oznacza że składnik losowy ma rozkład zbliżony do normalnego.

Test homoskedastyczność - Brak podstaw do odrzucenia H_0 o stałej wariancji reszt więc wariancja reszt jest stała co spełnia założenie kmnk.

Test serii - nie odrzucono H_0 o losowości składnik losowego epsilon co oznacza że składnik ma charakter przypadkowy a model został poprawnie wyspecyfikowany.

Współliniowość - stwierdzono występowanie współliniowości między zmiennymi objaśniającymi. Powoduje to wzrost wariancji estymatorów, niestabilności współczynników i utrudnia jednoznaczna interpretację wpływu poszczególnych zmiennych. Autokorelacja reszt- wykryto autokorelację reszt co oznacza że składnik losowy nie jest w pełni niezależny. Narusza jedno z podstawowych założeń kmnk i może prowadzić do błędnych wniosków z testów istotności w związku z tym przeprowadziliśmy umnk aby zapobiec autokorelacji reszt.

2.

$R^2 = 0,99$ oznacza że model wyjaśnia 99% zm. zależnej formalnie dopasowanie jest bardzo wysokie

3.

$$\begin{aligned} \hat{y} &= 32,12 + 0,0194x_1 - 0,153x_2 \\ &+ 0,029x_5 - 0,034x_6 \end{aligned} \quad \text{model kmnk}$$
$$\begin{aligned} \hat{y} &= 32,34 + 0,0179x_1 - \\ &0,158x_2 + 0,033x_5 - 0,040x_6 \end{aligned} \quad \text{model umnk}$$

Ze względu na współliniowość interpretacja poszczególnych parametrów może być nie poprawna $\hat{y} = 32,12 + 0,0194x_1 - 0,153x_2 + 0,029x_5 - 0,034x_6$

1. Wyraz wolny – $8,903E^{13}$ - oznacza przewidywana wartość PKB gdy wszystkie zmienne objaśniające są równe 0 po . Ekonomicznie: punkt startowy pkb wartość wszystkich zmiennych objaśniających jest równa 0, w praktyce jest to bardziej wartość teoretyczna niż realistyczna.

2. $B_1 = 1,019$ (wnioski o patenty) każdy dodatkowy tysiąc patentów zwiększa pkb o 1,019 jednostki przy założeniu stałości pozostałych zmiennych objaśniających. Ekonomicznie : większa liczba patentów powoduje większą innowacyjność oraz wzrost produktywności co powoduje wzrost pkb. Wartość dodatnia jest zgodna z intuicją ekonomiczną.

3. Parametr $B_2 = 0,858$ wzrost x_2 o jedna jednostkę powoduje wzrost PKB o 0,858 jednostek przy założeniu stałości reszty zmiennych. Ekonomiczny: Wzrost populacji kobiet powoduje wzrost pkb. Wzrost liczby pracowników wiąże się ze wzrostem pkb.

4. $B_3 = 1,029$ (import krajowy) wzrost udziału importu o 1 % zwiększa pkb o 1,029 jednostki przy założeniu że reszta zmiennych jest stała. Ekonomicznie : większy import może świadczyć o większym wroście aktywności gospodarczej lepszych możliwościach produkcyjnych i konsumpcyjnych co pozytywnie wpływa na pkb.

5. $B_4 = 0,966$ (export krajowy) Wzrost udziału o 1 % zwiększa pkb o 0,966 jednostki przy założeniu że reszta zmiennych modelu jest stała . Ekonomicznie: Export świadczy o wroście aktywności gospodarczej . Interpretowane wartości zostały odlogarytmowane.

4. Ocena końcowa - Podsumowując : model spełnia część kluczowych założeń (normalność, skedastyczność,

Wnioski

2.

$R^2 = 0,99$ oznacza że model wyjaśnia 99% zmiennej zależnej zmiennymi objaśniającymi. Formalnie dopasowanie jest bardzo wysokie.

4.

Ocena końcowa - Podsumowując : model spełnia część kluczowych założeń kmnk (normalność , homoskedastyczność , losowaść składnika losowego) , ale narusza brak współliniowości i autokorelacji co oznacza że estymatory kmnk nie są efektywne. Proces estymacji nie zakończył się pełnym sukcesem model formalnie dopasowuje się dobrze do danych ale z punktu widzenia wiarygodności interpretacyjnej i statystycznej nie można go uznać za wypełni poprawny, w związku z autokorelacją i współliniowością przeprowadziliśmy proces estymacji umnk który jest odporny na autokorelację. Dzięki umnk estymatory pozostaną nieobciążone i efektywne nawet przy naruszeniu tych założeń . Wnioski do modelu umnk . Estymatory są nieobciążone i efektywne w przeciwieństwie do kmnk w obecności autokorelacji estymatory mają minimalną wariancję spośród liniowych estymatorów nieobciążonych. Dzięki temu współczynniki regresji można interpretować wiarygodnie.