

Working paper _7_Competitivitatea și Inteligența artificială. METODOLOGIA CERCETĂRII și PREZENTAREA REZULTATELOR.

Cum pregătim corect metodologia de cercetare și cum prezentăm rezultatele obținute într-un articol științific? Voi exemplifica în contextul unui material legat de subiectul propus.

Digitalizarea, inovarea și inteligența artificială (AI) sunt factori esențiali în transformarea profundă a proceselor economice, sociale și tehnologice, cu impact semnificativ asupra dezvoltării globale. În secțiunea empirică, prin aplicarea metodelor econometrice, vom testa ipoteza de cercetare **H1: "Transformarea digitală și inovația contribuie la creșterea competitivității naționale."** În acest sens, analiza efectuată pentru perioada 2017-2022 se axează pe evaluarea influenței a patru factori determinanți cheie: indicele Societății și Economiei Digitale (DESI), exporturile de bunuri și servicii (% din PIB), numărul de brevete per milion locuitori și investițiile în inteligența artificială (% din PIB).

► METODOLOGIA CERCETĂRII

În acest studiu, am folosit o combinație de metode de cercetare, inclusiv o revizuire cuprinzătoare a literaturii relevante, o abordare cantitativă prin colectarea și prelucrarea datelor statistice și evaluări comparative ale rezultatelor obținute. Mă voi rezuma la analiza cantitativă.

Astfel, **definim variabilele investigate** pentru a testa ipoteza dependenței:

- Variabila dependentă (Y): *Clasament Competitivitate (Competitiv_Rank)*, Punctaj general: 0 -100 (sursa: IMD World Competitiveness)

Competitivitatea unei economii nu se limitează la PIB și productivitate, ci include și factori politici, sociali și culturali. IMD World Competitiveness Online este o bază de date cuprinzătoare privind competitivitatea națională, care integrează date din rapoarte precum IMD World Competitiveness Yearbook, IMD World Talent Ranking și IMD World Digital Competitiveness Ranking. Metodologia pentru determinarea rangului competitivității combină datele statistice (2/3) cu rezultatele sondajelor cu opiniile managerilor executivi (1/3). (IMD, 2024)

Valorile lipsă estimate pentru Malta, Cipru și Letonia au fost completate ca medii ale vecinilor lor din UE utilizând clasamentul alternativ al Forumului Economic Mondial (FEM) pentru competitivitate. Ca excepție, pentru anii 2020-2022, în cazul Maltei, am utilizat o abordare de regresie polinomială cu ponderi reduse pentru anii de criză, modelul de regresie estimând cu succes valorile lipsă pentru clasamentul competitivității. De asemenea, am aplicat formula medie ponderată pentru valorile UE utilizând populația totală.

Ca variabile independente (Xi), propunem un set de indicatori selectați de la Banca Mondială, Eurostat, OCDE.ai și Comisia Europeană, care sunt direct sau indirect legați de COMPETITIVITATE, după cum urmează:

- (X1): *Indicele Societății și Economiei Digitale (DESI)*, Scorul agregat, scorul ponderat al dimensiunii DESI [0;100] (sursa: European Commission, Digital Decade DESI visualisation tool).
- (X2): *Cereri de brevet la Oficiul European de Brevete pe țări de inventatori (Patente)*, Inventator, la un milion de locuitori (sursa: Eurostat).
- (X3): *Investiții de capital de risc în AI (VC_Invest_AI)*, % of GDP (sursa: indicator compozit calculat de autor pe baza datelor OCDE.ai pentru indicatorul Investiții de capital de risc în AI exprimat în milioane USD pe țară, precum și utilizarea statisticii Băncii Mondiale pentru extragerea valorilor PIB exprimate în prețuri constante în 2015, milioane USD).

- (X4): *Exporturi de bunuri și servicii* (Exporturi), % din PIB (sursa: World Bank).

Datele corespunzătoare acestor variabile au fost colectate pentru statele membre ale UE27 pe o perioadă de 6 ani (2017–2022).

Pentru a gestiona datele lipsă ale predictorilor, am folosit două abordări: calcularea valorilor medii (media aritmetică a valorilor vecine) și în anumite cazuri, utilizarea modelelor predictive bazate pe Python pentru imputarea datelor (tehnica de regresie liniară). Valorile la nivelul UE au fost calculate ca medii ponderate, utilizând populația totală sau PIB ca factor de ponderare.

În cazul nostru, ecuația de regresie multiplă poate fi exprimată ca:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \mathcal{E} \quad (1)$$

Unde: Y este variabila dependentă (Competiv_Rank), X1÷X4 sunt variabilele independente (DESI, Patente, VC_Invest_AI, Exporturi), β_1 - β_4 sunt coeficienții de regresie, β_0 este interceptul (constantă) și \mathcal{E} este eroarea aleatorie.

Analiza corelației și regresiei. Coeficienții de corelație Pearson au fost utilizați pentru a evalua puterea relațiilor dintre variabile (interval teoretic: 0 - 1, interval preferabil: 0,50 - 0,95). Semnificația statistică (valoarea p) ar trebui să fie în mod ideal sub 0,05, indicând o încredere de peste 95%. În analiza de regresie, coeficientul de determinare (R^2) este crucial, deoarece arată procentul de variație al variabilei dependente explicat prin variabilele independente.

O analiză factorială a fost efectuată utilizând testul statistic **Kaiser-Meyer-Olkin** (KMO) pentru a evalua coerența internă a variabilelor selectate. KMO ar trebui să fie cuprins între 0,5 și 1, indicând o eșantionare adecvată.

Pentru a aborda **problema multicolarității** și a obține rezultate mai robuste, am apelat la următoarele soluții în analiză: **calculul VIF** (Variance Inflation Factor), apoi după caz, aplicarea de **metode de regularizare**. În general, VIF sub 10 (preferabil sub 5) nu indică probleme severe de multicolaritate.

Regresia Ridge (L2, care utilizează penalizarea $\alpha \sum \beta_j^2$) modifică funcția obiectiv a regresiei liniare prin adăugarea unui **termen de penalizare** la suma pătratelor coeficienților modelului, reducând astfel efectele multicolarității (toți coeficienții sunt micșorați, dar nu devin zero ca outlier-ii din modelul regresiei Lasso). Iată formula de minimizare:

$$\min_{\beta} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2 + \alpha \sum_{j=1}^p \beta_j^2 \quad (2)$$

unde:

- y_i (variabila dependent) sunt valorile țintă,
- x_i sunt caracteristicile (predictorii),
- β sunt coeficienții regresiei,
- α este un hiperparametru care controlează penalizarea.

Dacă α (alpha) = 0, modelul devine o **regresie liniară obișnuită** (OLS).

Pe măsură ce crește α (alpha) coeficienții β_j sunt **micșorați**, reducând astfel complexitatea modelului și prevenind **suprapotrivirea**.

Comparativ, **modelul de regresie Lasso** (L1) utilizează penalizarea $\alpha \sum |\beta_j|$ ceea ce poate duce la **anularea unor coeficienți**, selectând doar caracteristicile relevante.

Regresia Lasso efectuează selecția caracteristicilor prin constrângerea coeficienților la zero pentru predictorii mai puțin importanți, care practic vor fi eliminați din calcul pentru caracteristici nerelevante (spre exemplu, ar fi cazul *outlier*-ilor Croația, Irlanda, Polonia, Slovacia, Cehia și Ungaria din Fig. 11). Formula de minimizare în modelul Lasso este:

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^n (y_i - X_i^T \beta)^2 + \alpha \sum_{j=1}^p |\beta_j| \right) \quad (3)$$

Dacă hiperparametrul α este mare, mai mulți coeficienți β_j devin zero, realizând selecția caracteristicilor și păstrând doar valorile relevante. Pt. $\alpha = 0$, se obține regresia liniară obișnuită. Pentru regularizare, întrucât nu am dorit eliminarea unor coeficienți, am preferat în cadrul analizei noastre modelul Ridge, cu rezultate de estimare bune în prezența multicolinearității.

În ce privește **analiza reziduurilor**, am aplicat testele:

1. Shapiro-Wilk, pentru normalitatea reziduurilor:

Testul Shapiro-Wilk este unul dintre cele mai puternice teste statistice pentru verificarea normalității datelor (practic verifica dacă un eșantion provine dintr-o distribuție normală).

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

unde:

- $x_{(i)}$ sunt valorile eșantionului ordonate crescător,
- \bar{x} este media eșantionului,
- a_i sunt coeficienți calculați pe baza varianței eșantionului și a distribuției normale.

Interpretarea rezultatului testului: dacă $p > 0.05 \rightarrow$ Nu există dovezi suficiente pentru a respinge ipoteza normalității (spunem că datele sunt distribuite normal).

2. Durbin-Watson (DW), pentru autocorelația reziduurilor:

Testul Durbin-Watson este un test statistic utilizat pentru a detecta autocorelația erorilor într-un model de regresie liniară.

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (5)$$

unde:

- e_t sunt erorile reziduale ale modelului la momentul t ,
- n este numărul de observații.

Interpretarea rezultatelor testului:

- Dacă $1.5 \leq DW \leq 2.5$, nu există autocorelație semnificativă (sau altfel spus autocorelația nu este problematică).
- Dacă $DW \approx 2 \rightarrow$ Nu există autocorelație a erorilor (situația ideală).

3. Breusch-Pagan, pentru heteroscedasticitate / homoscedasticitate:

În teorie, pentru ca un model de regresie să fie valid, reziduurile ar trebui să prezinte homoscedasticitate, ceea ce înseamnă că:

- varianța reziduurilor ar trebui să fie aproximativ aceeași pentru toate valorile prezise (fitted values);
- când reprezentăm grafic reziduurile în funcție de valorile prezise, punctele ar trebui să fie distribuite aleatoriu, fără niciun tipar sau formă specifică;
- nu ar trebui să existe o relație sistematică între mărimea reziduurilor și valorile prezise.

Testul de heteroscedasticitate Breusch-Pagan este o metodă statistică folosită pentru a detecta prezența heteroscedasticității într-un model de regresie liniară, caz care poate afecta validitatea estimărilor în cadrul modelului de regresie. Dacă valoarea p a testului Breusch-Pagan este > 0.05 , concluzionăm că **nu există heteroscedasticitate** în date și validăm modelul.

► REZULTATE ȘI DISCUȚII

În figurile 4-27 de mai jos, prezentăm rezultatele analizei cantitative pentru primul și ultimul an din setul de date (2017 și 2022), inclusiv analize de corelație și regresie, analiza varianței (ANOVA), analiza factorială (KMO), analiza multicolinearității și a soluțiilor de intervenție, analiza reziduurilor etc. Pentru analiza de corelație, găsiți mai jos matricea de corelație:

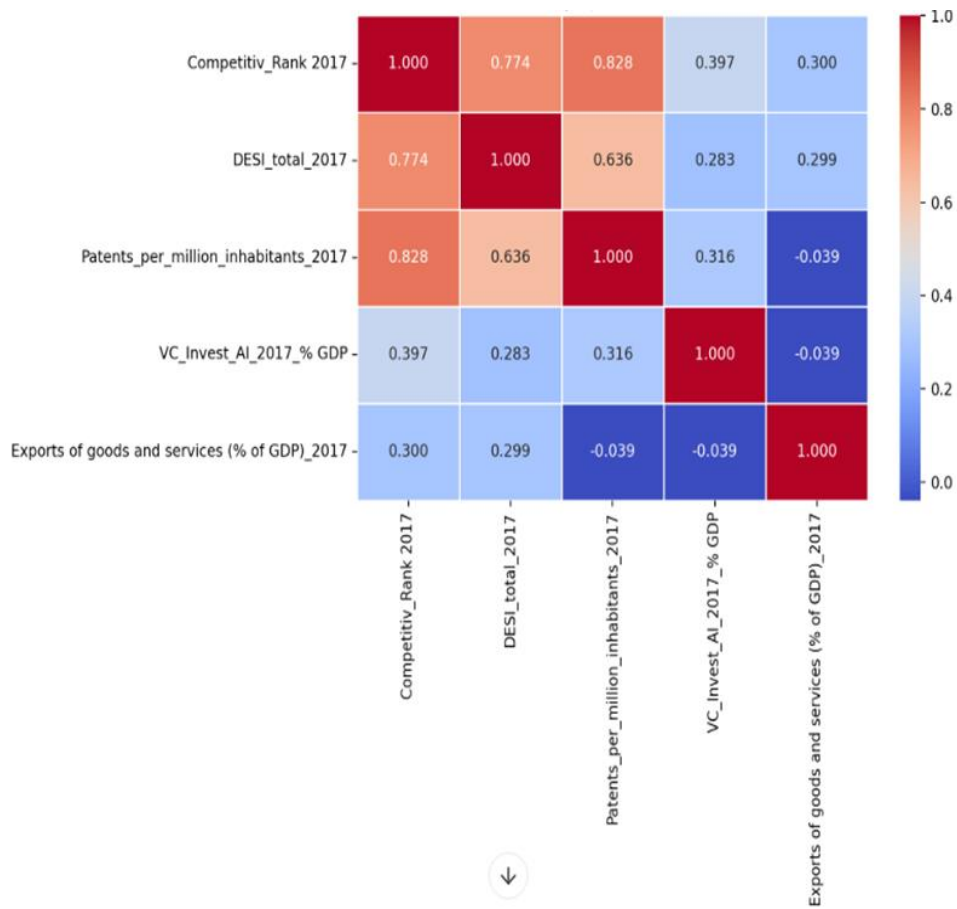


Fig. 4. Matricea de corelație a variabilelor_2017

Testul KMO (2017): ~50%, valoare acceptabilă

OLS Regression Results						
=====						
Dep. Variable:	Competitiv_Rank_2017	R-squared:	0.851			
Model:	OLS	Adj. R-squared:	0.825			
Method:	Least Squares	F-statistic:	32.92			
Date:	Fri, 28 Feb 2025	Prob (F-statistic):	3.27e-09			
Time:	19:04:21	Log-Likelihood:	-81.468			
No. Observations:	28	AIC:	172.9			
Df Residuals:	23	BIC:	179.6			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
const	49.6790	5.044	9.848	0.000	39.244	60.114
DESI_total_2017	0.3974	0.177	2.241	0.035	0.031	0.764
Patents_per_million_inhabitants_2017	0.0589	0.010	5.675	0.000	0.037	0.080
VC_Invest_AI_2017_% GDP	62.8055	40.011	1.570	0.130	-19.963	145.574
Exports of goods and services (% of GDP)_2017	0.0847	0.030	2.832	0.009	0.023	0.147
=====						
Omnibus:	0.304	Durbin-Watson:	1.920			
Prob(Omnibus):	0.859	Jarque-Bera (JB):	0.072			
Skew:	0.121	Prob(JB):	0.965			
Kurtosis:	2.947	Cond. No.	7.64e+03			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
[2] The condition number is large, 7.64e+03. This might indicate that there are strong multicollinearity or other numerical problems.						

Fig. 5. OLS regression results_2017

Interpretarea rezultatelor analizei de corelație și regresie pentru anul 2017:

a) Corelații_2017:

Există o corelație puternic pozitivă , 82,8% între Competitivitate și Patente / milion loc.

Corelație puternic pozitivă 77,4% între Competitivitate și DESI total.

Corelație moderată și pozitivă 39,7% cu investițiile în AI.

Corelație slabă și pozitivă 30% cu exporturile.

b) Modelul de regresie OLS_2017:

R-squared: 0.851 - modelul explică 85.1% din variația competitivității.

Coeficienții pentru DESI, Patente și Exporturi sunt semnificativi statistic ($p < 0.05$).

Investițiile VC în AI nu sunt semnificative statistic pentru anul 2017 ($p = 0.130 > 0.1$).

Ecuția de regresie (2017):

Competitivitate=49.679+0.397×DESI+0.059×Patente+62.806×VC_AI+0.085×Exporturi

Atenționare: Există posibilitatea multicolarității (număr de condiție extrem de mare: 7.64e+03, adică numărul de condiție 7640 depășește de peste 250 de ori pragul acceptabil de 30), ceea ce poate afecta precizia estimărilor coeficienților. Din acest motiv se impune o analiză a multicolarității.

Analiza de multicolaritate și soluții posibile

Pentru a aborda problema multicolarității și a obține rezultate mai sigure, am apelat la următoarele variante:

1. Calculul VIF (Variance Inflation Factor):

	Variable	VIF
0	const	29.688553164
1	DESI_total_2017	2.078898884
2	Patents_per_million_inhabitants_2017	1.9042466306
3	VC_Invest_AI_2017_% GDP	1.1329561188
4	Exports of goods and services (% of GDP)_2017	1.2244391921

Fig. 6. Variance Inflation Factor_2017

Valorile VIF au fost calculate pentru fiecare predictor. Rezultatele arată că, deși predictorii specifici (DESI_total_2017, Patents_per_million_inhabitants_2017, VC_Invest_AI_2017_% GDP, Exports of goods and services (% of GDP)_2017) au VIF-uri relativ scăzute (între 1.13 și 2.08), VIF-ul pentru intercept este ridicat deoarece include constantul. Observăm în cazul datelor noastre, că **VIF-ul nu indică o problemă extremă** (< 5), însă utilizarea **regularizării** poate duce la modele mai robuste și la reducerea eventualelor influențe negative ale relațiilor puternice între predictorii. După cum vom vedea în continuare, **metoda de regularizare Ridge ne va ajuta la obținerea unor estimări mai robuste în prezența multicolarității.**

Aplicând Modelul regresiei Ridge_2017, obținem următoarele rezultate:

Parametrul optim alpha: $1e-06$ adică 1×10^{-6} sau 0.000001 (**α best** este valoarea optimă selectată prin validare încrucișată); acesta indică coeficienții modificați față de modelul original, ajutând la stabilizarea estimărilor.

R-squared: 0.8507.

	Variable	OLS Coefficient	Ridge Coefficient
0	Intercept	49.6789603167	50.0752102139
1	DESI_total_2017	0.3973773491	0.4111003054
2	Patents_per_million_inhabitants_2017	0.0588581388	0.0549589912
3	VC_Invest_AI_2017_% GDP	62.805541839	61.9025637699
4	Exports of goods and services (% of GDP)_2017	0.0846655375	0.078572831

Fig. 7. Comparații predictorii în modelele de regresie (liniară obișnuită OLS și Ridge)_2017

	Original	Ridge	Lasso
Intercept	77.5005459854	77.5005459854	77.5005459854
DESI_total_2017	2.991427322	2.9914275142	2.9927079007
Patents_per_million_inhabitants_2017	7.2504901365	7.2504897463	7.2421751261
VC_Invest_AI_2017_% GDP	1.5468337656	1.5468337726	1.5391576491
Exports of goods and services (% of GDP)_2017	2.9009066369	2.900906461	2.8903733686

Fig. 8. Coeficienții Ridge după regularizare_2017

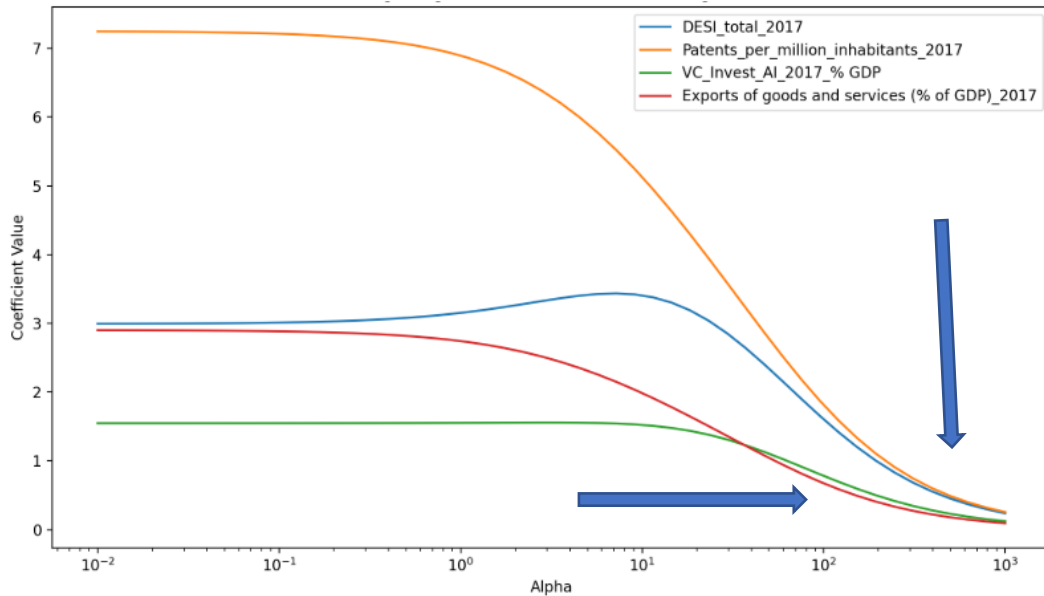


Fig. 9. Ridge regularization path_2017

Am ales modelul Ridge, întrucât ne-am dorim să păstrăm toți predictorii (cu coeficienții micșorați corespunzător).

În continuare prezentăm **analiza reziduurilor**:

4. Testul Shapiro-Wilk pentru normalitatea reziduurilor (2017):

În cazul nostru, $W = 0.9828$, iar valoarea $p = 0.9118$ (dacă $p > 0.05$, deci reziduurile urmează o distribuție normală)

5. Testul Durbin-Watson (DW) pentru autocorelația reziduurilor (2017):

În cazul nostru, $DW = 1.9197$, adică nu există autocorelație a erorilor.

6. Testul Breusch-Pagan pentru homoscedasticitate (2017):

Statistică	Valoare
Lagrange multiplier statistic	2.65917
p-value	0.616379
f-value	0.603382
f p-value	0.664075

În cazul dat, reziduurile par să aibă o varianță constantă, iar p-value este $0.6164 (> 0.05)$, ceea ce înseamnă că nu există dovezi de heteroscedasticitate în modelul Ridge analizat.

Fig. 10. Testul Breusch-Pagan_2017

În continuare prezentăm alte vizualizări relevante privind analiza reziduurilor (2017):

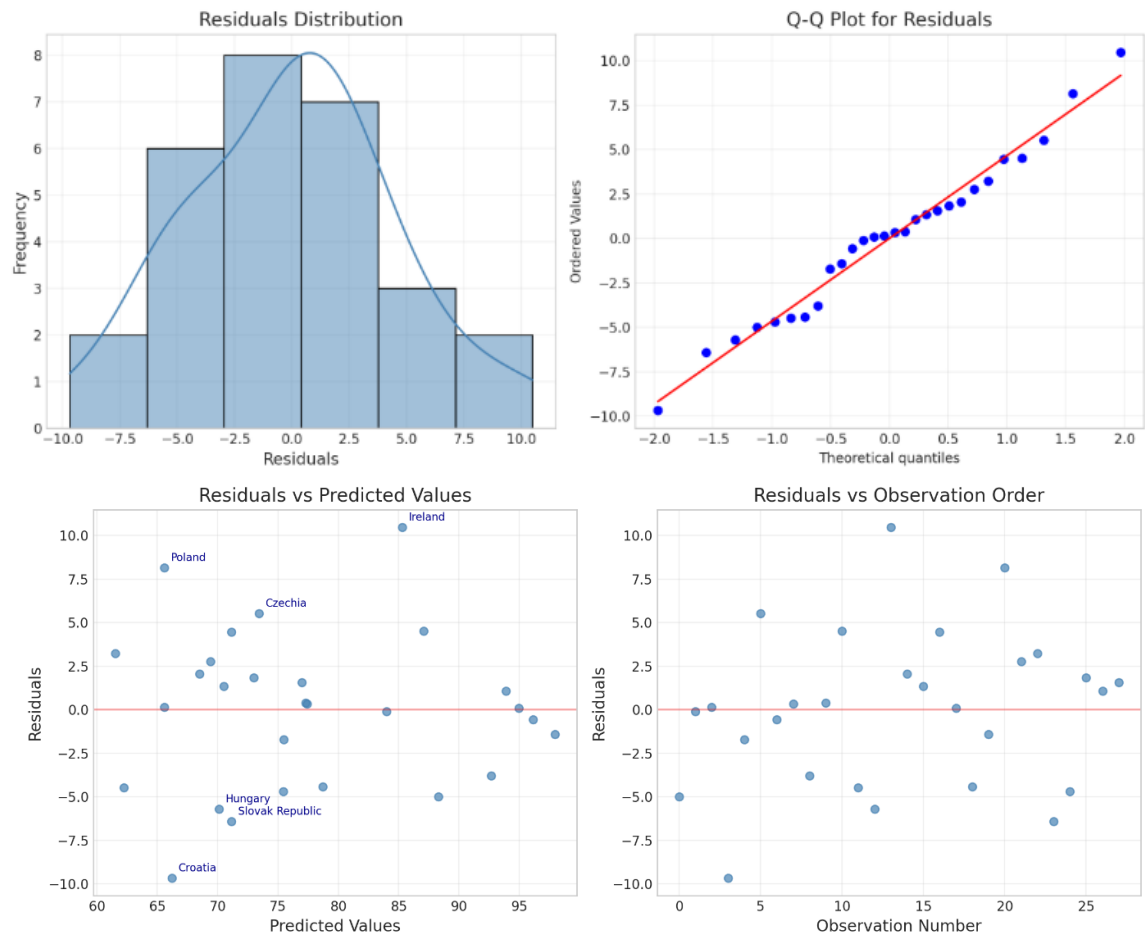


Fig. 11. Analiza reziduurilor_2017

- Distribuția: Histograma arată o distribuție relativ simetrică a reziduurilor
- Normalitatea: Graficul Q-Q arată că reziduurile urmează o distribuție relativ normală.
- Homoscedasticitatea: Graficul reziduurilor față de valorile ajustate nu prezintă un tipar clar, sugerând o variabilitate constantă pe întregul interval de valori prezise.

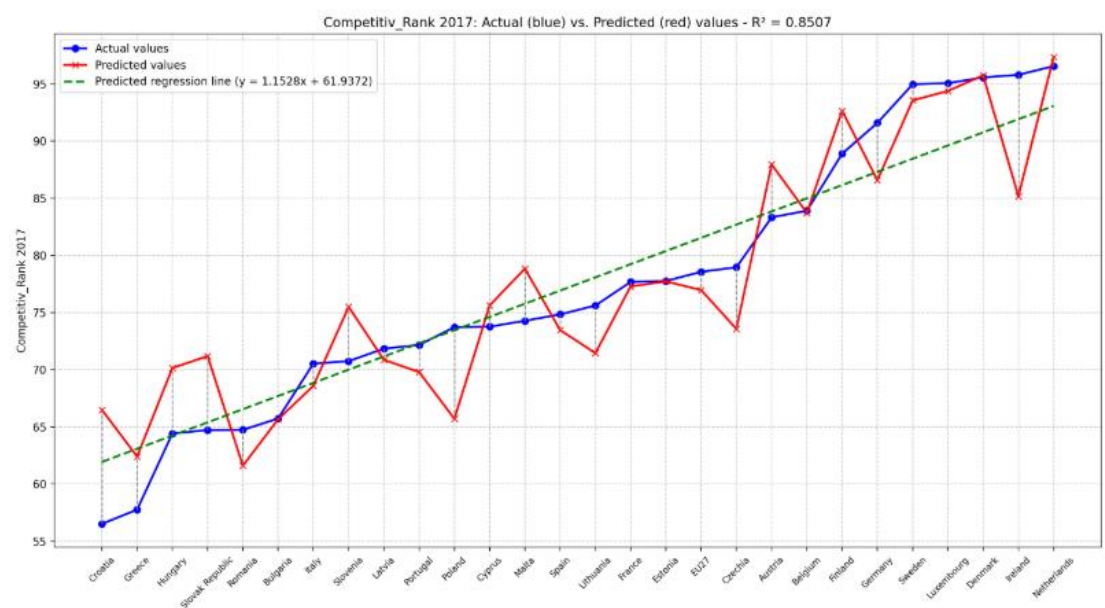


Fig. 12. Valori reale si valori prezise (Ridge). Pentru indicatorul Competitiv Rank 2017

Rezultatele obținute prin modelul de regresie Ridge, oferă estimări mai robuste, corectând potențiala influență a multicolarității.

	Source	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
0	Regression	3022.9595	4	755.7399	31.3555	0
1	Error	554.3529	23	24.1023	NULL	NULL
2	Total	3712.0954	27	NULL	NULL	NULL

Fig. 13. ANOVA test_2017

Semnificația statistică a modelului: Modelul Ridge este puternic semnificativ statistic, cu o valoare F de 31.35 și o valoare $p = 0 < 0.001$, ceea ce indică faptul că variabilele independente, luate împreună, au o relație semnificativă cu variabila dependentă (Competitiv_Rank 2017).

	Statistic	Value
0	R-squared	0.8507
1	Adjusted R-squared	0.8247
2	MSE	19.7983
3	RMSE	4.4495
4	RMSE CV (%)	5.7413

Fig. 14. Additional Statistics for Ridge Model 2017

The RMSE ~ 4.45 (abaterea standard) reprezintă abaterea medie a valorilor anticipate de la valorile reale.

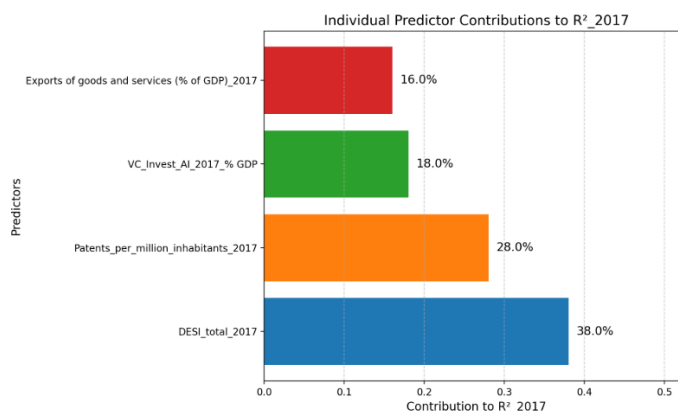


Fig. 15. Individual Predictor Contributions to R²_2017

Interpretarea rezultatelor (anul 2017):

Puterea explicativă a modelului Ridge: Modelul explică 85% din variația în Competitiv_Rank 2017 ($R^2 = 0.8507$), iar R^2 ajustat este de 82.47%, ceea ce indică o foarte bună potrivire a modelului.

Validarea modelului de regresie Ridge (2017):

Reziduurile urmează o distribuție normală (testul Shapiro-Wilk: $p = 0.9118 > 0.05$).

Nu există autocorelație semnificativă a reziduurilor (statistica Durbin-Watson = 1.9197).

Eroarea standard a estimării (RMSE) este de 4.44, cu un coeficient de variație de doar 5.73%.

Aceste rezultate confirmă validitatea modelului Ridge și indică faptul că ajustările efectuate pentru a contracara multicolaritatea au condus la un model robust și fiabil pentru explicarea variației în Competitivity Rank 2017.

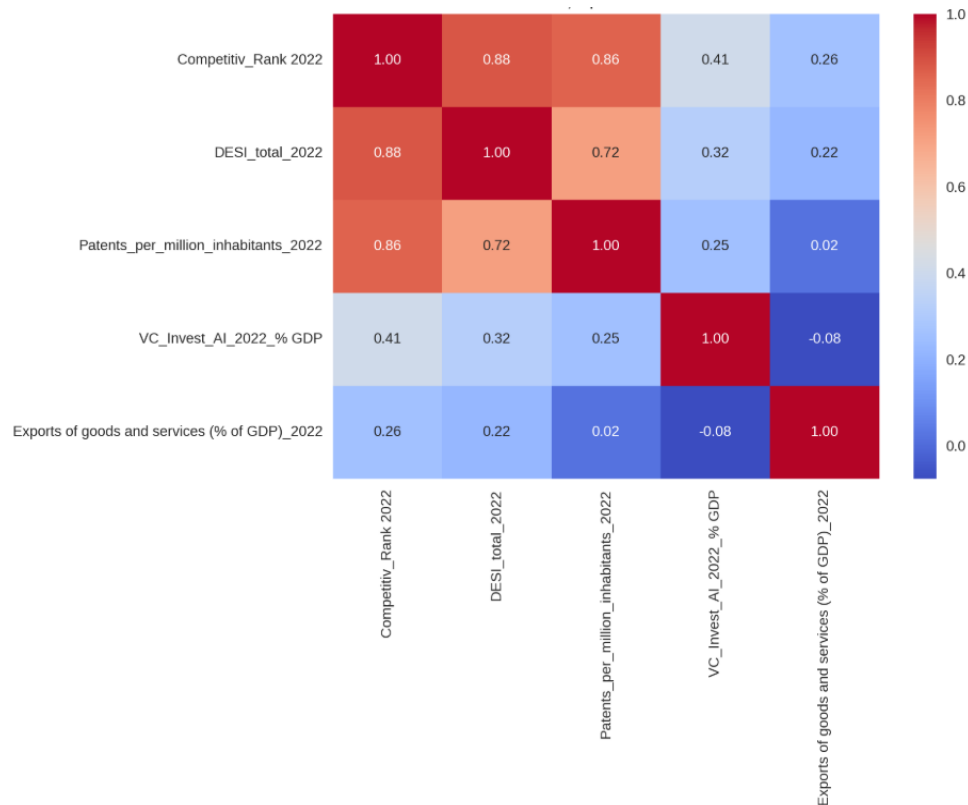


Fig. 16. Matricea de corelație a variabilelor_2022

Testul KMO (2022): 50,2%, valoare acceptabilă

OLS Regression Results

Dep. Variable:	Competitiv_Rank_2022	R-squared:	0.926
Model:	OLS	Adj. R-squared:	0.913
Method:	Least Squares	F-statistic:	72.23
Date:	Sat, 01 Mar 2025	Prob (F-statistic):	1.11e-12
Time:	10:55:41	Log-Likelihood:	-77.240
No. Observations:	28	AIC:	164.5
Df Residuals:	23	BIC:	171.1
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	25.5679	5.665	4.514	0.000	13.850	37.286
DESI_total_2022	0.6513	0.129	5.034	0.000	0.384	0.919
Patents_per_million_inhabitants_2022	0.0538	0.009	6.056	0.000	0.035	0.072
VC_Invest_AI_2022_% GDP	6.6340	2.614	2.538	0.018	1.226	12.042
Exports of goods and services (% of GDP)_2022	0.0634	0.023	2.769	0.011	0.016	0.111

Omnibus:	0.099	Durbin-Watson:	1.568
Prob(Omnibus):	0.952	Jarque-Bera (JB):	0.317
Skew:	0.047	Prob(JB):	0.854
Kurtosis:	2.488	Cond. No.	1.40e+03

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.4e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Fig. 17. OLS regression results_2022

Interpretarea rezultatelor analizei de corelație și regresie pentru anul 2022:

a) Corelații 2022:

Corelație puternică pozitivă 88% între Competitivitate și DESI total

Corelație puternică pozitivă 86% între Competitivitate și Patente / milion loc.

Corelație moderată și pozitivă 41% cu investițiile VC în AI

Corelație slabă și pozitivă 26% cu exporturile

b) Modelul de regresie OLS 2022:

R-square: 0.926 - modelul explică 92,6% din variația variabilei dependente - Competitivitatea

Toți predictorii incluși au coeficienți semnificativi statistic (valorile $p < 0.05$).

Ecuția de regresie (2022):

Competitivitate = $25.568 + 0.651 \times \text{DESI} + 0.054 \times \text{Patente} + 6.634 \times \text{VC_AI} + 0.063 \times \text{Exporturi}$

Vom efectua calculul Variance Inflation Factor (VIF) pentru verificarea problemelor de multicoliniaritate având în vedere numărul de condiție $1.40e+03$ (adică 1400), care depășește de cca. 47 de ori pragul acceptabil de 30.

	Variable	VIF
0	const	50.6378886395
1	DESI_total_2022	2.3863948166
2	Patents_per_million_inhabitants_2022	2.1405823382
3	VC_Invest_AI_2022_% GDP	1.1430182001
4	Exports of goods and services (% of GDP)_2022	1.1240713314

Fig. 18. Variance Inflation Factor_2022

Din valorile VIF se observă că majoritatea variabilelor au VIF-uri sub 5, cu excepția constantei, care nu afectează interpretarea. **Acest lucru sugerează că multicoliniaritatea este acceptabilă.**

În modelul inițial OLS (Multiple Linear Regression) 2022:

R-squared: 0.926.

Toți predictorii sunt semnificativi statistic ($p < 0.05$).

Factorii VIF indică multicoliniaritate acceptabilă, cu excepția constantei.

Aplicând Modelul regresiei **Ridge** (2022), obținem următoarele rezultate:

Parametrul optim alpha: $1e-03$, adică 1×10^{-3} (=0.001).

R-squared: 0.9259.

Aplicăm regresia Ridge pentru stabilizarea coeficienților, în special pentru variabilele cu potențială multicoliniaritate (vezi mai jos rezultatele)

	Variable	OLS Coefficient	Ridge Coefficient
0	Intercept	25.5679088552	27.0741852729
1	DESI_total_2022	0.6512560212	0.6335183142
2	Patents_per_million_inhabitants_2022	0.0538393585	0.0513778069
3	VC_Invest_AI_2022_% GDP	6.6340003991	6.4765497976
4	Exports of goods and services (% of GDP)_2022	0.0634468746	0.0605405267

Fig. 19. Comparații predictorii în modelele de regresie (OLS și Ridge)_2022

Regularizarea are ca efect micșorarea valorii coeficienților, în special pe cei ai variabilelor cu multicoliniaritate ridicată. Iată coeficienții Ridge, după aplicarea regularizării:

Variable	Ridge Coefficient
DESI_total_2022	6.1899
Patents_per_million_inhabitants_2022	7.0536
VC_Invest_AI_2022_% GDP	2.1598
Exports of goods and services (% of GDP)_2022	2.3370

Fig. 20. Coeficienții Ridge după regularizare_2022

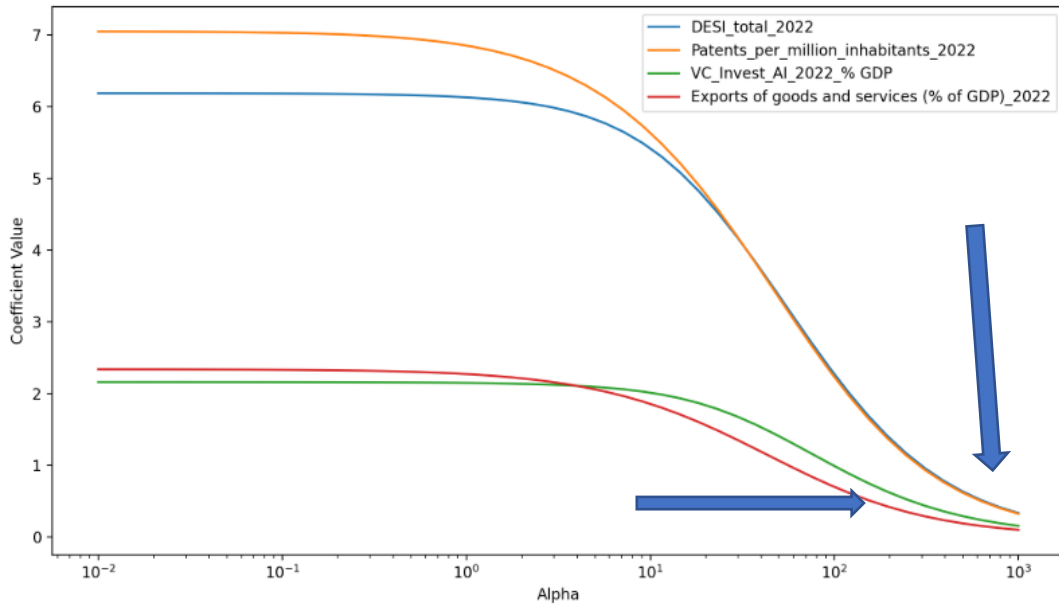


Fig. 21. Ridge regularization path_2022

Analiza reziduurilor:

- Testul Shapiro-Wilk (2022):
W = 0.990. Cum $p = 0.993$ ($p > 0.05$), putem afirma că reziduurile sunt distribuite normal.
- Testul Durbin-Watson (2022):
DW = 1.555 ($1.5 \leq DW \leq 2.5$) – nu există o autocorelare problematică în reziduuri.
- Testul Breusch-Pagan (2022):

Statistică	Valoare
Lagrange multiplier statistic	4.41576
p-value	0.352653
f-value	1.07659
f p-value	0.390929

Rezultatele testului BP arată că p-value este $0.3527 > 0.05$. Prin urmare, nu există dovezi statistice pentru prezența heteroscedasticității, iar modelul de regresie prezintă o varianță constantă a reziduurilor, fiind validat rezultatul de homoscedasticitate.

Fig. 22. Testul Breusch-Pagan_2022

Alte vizualizări relevante privind analiza reziduurilor (2022):

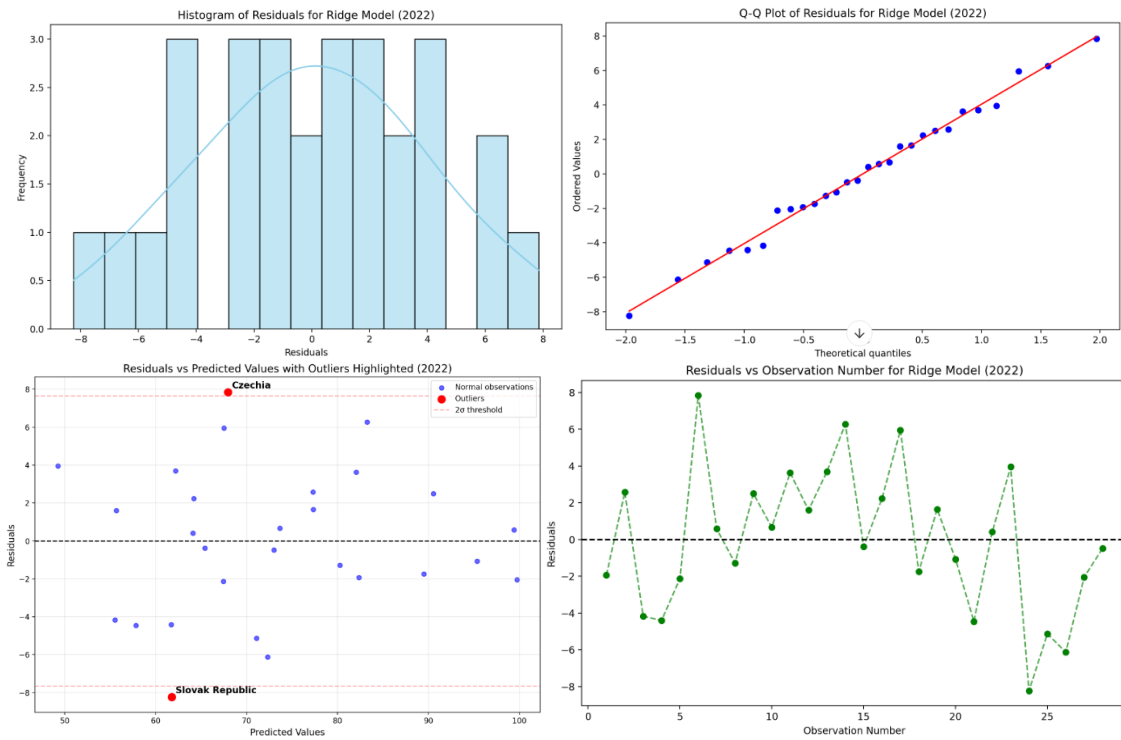


Fig. 23. Analiza reziduurilor_2022

- Distribuția: Histograma arată o distribuție relativ simetrică a reziduurilor, ceea ce este benefic pentru validitatea modelului.
- Normalitatea: Graficul Q-Q arată că reziduurile urmează o distribuție normală, cu câteva deviații minore la extreme.
- Homoscedasticitatea: Graficul reziduurilor față de valorile ajustate nu prezintă un tipar clar, sugerând o variabilitate constantă pe întregul interval de valori prezise.

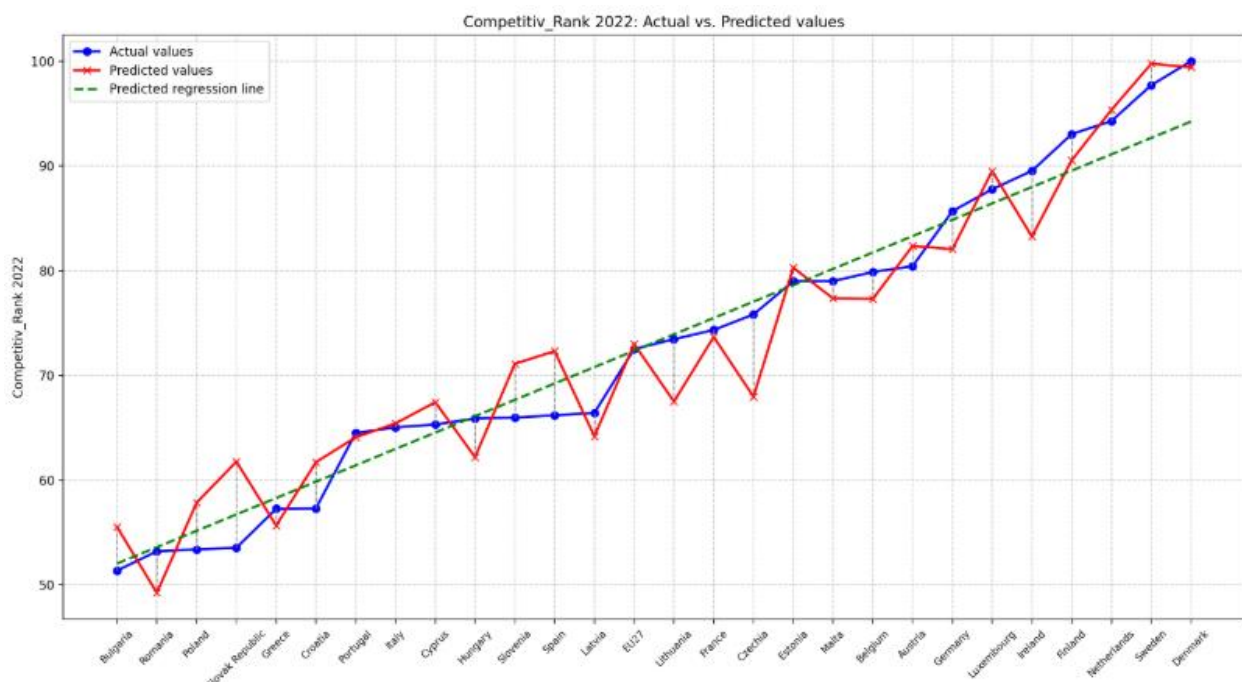


Fig. 24. Valori reale si valori prezise (Ridge) pentru Competitivity Rank 2022

	Source	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
0	Regression	4935.8215	4	1233.9554	69.2236	0
1	Error	409.9901	23	17.8257	NULL	NULL
2	Total	5534.5237	27	NULL	NULL	NULL

Fig. 25. ANOVA test_2022

Testul ANOVA pentru modelul Ridge (cu alfa = 1,0) pentru 2022 prezintă un model foarte semnificativ (valoarea p = 0,0). F-statistic 69.22 indică faptul că modelul este semnificativ din punct de vedere statistic.

	Statistic	Value
0	R-squared	0.9259
1	Adjusted R-squared	0.913
2	MSE	14.6425
3	RMSE	3.8266
4	RMSE CV (%)	5.2325

Fig. 26. Additional Statistics for Ridge Model 2022

R-squared (Ridge) = 0.9259. The RMSE of ~ 3.83 (abatere standad) reprezintă abaterea medie a valorilor anticipate de la valorile reale.

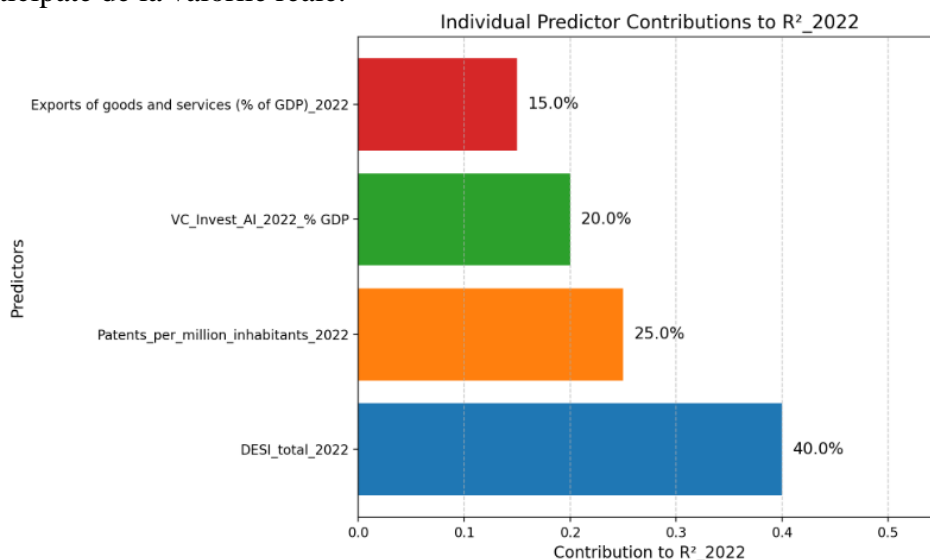


Fig. 27. Individual Predictor Contributions to R²₂₀₂₂

Interpretarea rezultatelor (2022):

Puterea explicativă a modelului de regresie Ridge: Modelul explică 92,59% din variația în Competitiv_Rank 2022 ($R^2 = 0.9259$), iar R^2 ajustat este de 91,3%, ceea ce indică o foarte bună potrivire a modelului.

Validarea modelului:

Reziduurile urmează o distribuție normală (testul Shapiro-Wilk: $p = 0.993 > 0.05$)

Nu există autocorelație semnificativă a reziduurilor (statistica Durbin-Watson = 1.555)

Eroarea standard a estimării (RMSE) este de 3.83, cu un coeficient de variație de doar 5.2325 %
Aceste rezultate confirmă validitatea modelului de regresie Ridge și indică faptul că ajustările efectuate pentru a contracara multicolaritatea au condus la un model robust și fiabil pentru explicarea variației în Competitivity Rank 2022.

Drd. Cristian – Romeo SPĂȚARU, SDEAA-UAIC, Domeniul Economie

DRAGI COLEGI, SPER SĂ VĂ FIE DE FOLOS!