Model de analiză macroeconomică bazat pe funcția de regresie

Prof.univ dr. Constantin ANGHELACHE
Universitatea „Artifex" /Academia de Studii Economice - Bucureşti
Prof. univ. dr. Mario G.R. PAGLIACCI
Universita degli Studi di Perugia - Italia
Asist. univ. drd. Ligia PRODAN
Universitatea Creştină „Dimitrie Cantemir"- Bucureşti

Abstract

Funcţia de regresie *stă la baza efectuării a numeroase analize micro sau macroeconomice. După studiul logic al variabilelor pentru a fi analizate, se continuă cu reprezentarea grafică a seriilor de date şi interpretarea primară, se prezintă fundamentarea modelului econometric utilizat.*

Cuvinte cheie: *Regresie simplă, corelaţie, Produs Intern Brut, consum final*

***

Modelul liniar de regresie presupune identificarea variabilelor pentru definirea modelului şi precizarea variabilei reziduale; contextul în care este utilizat modelul de regresie. Pentru analiza datelor seriilor cronologice (de timp) se utilizează o funcţie temporală care, în esenţă, este tot de regresie, cu o variabilă de timp (*t*). Scopul folosirii modelului de regresie este de a obţine parametrii ce corespund setului de variabile formulat, prin analiza dependenţei dintre variabile, în cazul în care seriile de date sunt înregistrate la nivelul unităţilor statistice pentru o perioadă sau un moment, precum şi pentru evidenţierea dependenţei dintre variabile într-un anumit orizont de timp. Este cunoscut că în analiza teoretică, dependenţa dintre variabile este stochastică. Considerarea variabilei reziduale în cadrul unui astfel de model este necesară. Ceilalţi factori care influenţează variabila rezultativă sunt grupaţi în variabila reziduală.

Modelele unifactoriale neliniare sunt liniarizate prin transformări aplicate variabilelor modelului de regresie. Un model de forma $y_i = a \cdot x_i^b$ se transformă într-un model liniar prin logaritmarea celor doi termeni ai egalităţii, rezultând funcţia liniară $\log y_i = \log a + b \cdot \log x_i$.

Modelul se recomandă atunci când punctele $(\log x_i, \log y_i)_{i=1,n}$ sunt situate, ca nor de puncte, în jurul unei drepte. Sunt cazuri când pentru estimarea parametrilor se folosesc alte tehnici de estimare, care, neputând fi liniarizat prin transformări elementare, estimarea parametrilor se face prin metode numerice.

Modelul liniar de regresie se bazează pe seriile de date pentru cele două caracteristici, fiind reprezentate prin vectorii x (variabila factorială) şi y (variabila rezultativă).

Este necesară definitivarea metodelor folosite pentru estimarea celor doi parametri; precizarea metodelor utilizate pentru testarea proprietăţilor estimatorilor modelului de regresie şi stabilirea cadrului privind folosirea modelului de regresie în efectuarea de previziuni.

În definirea funcţiei de regresie liniară sunt considerate, cel mai frecvent, patru ipoteze:

- seriile de date nu sunt afectate de erori de înregistrare.
- pentru fiecare valoare fixată a caracteristicii factoriale, variabila reziduală este de medie zero, respectiv:
  $E[\varepsilon_i | X = x_i] = 0$, pentru orice *i*,
- lipsa corelării reziduurilor exprimă faptul că între termenii reziduali nu se manifestă fenomenul de covarianţă
- ipoteza necorelării variabilei reziduale cu cea independentă, ceea ce presupune că $\mathrm{cov}(X, \varepsilon_j) = 0$, pentru orice *j*, arătând o creştere a valorilor variabilei factoriale care nu conduce automat la un spor al valorilor variabilei reziduale.

Se poate defini modelul liniar de regresie prin funcţia:

$y_i = b + a \cdot x_i + \varepsilon_i$, $i = 1,..,$n

Ipoteze formulate asupra variabilei reziduale:

$$\begin{cases} E(\varepsilon_i) = 0 \\ \mathrm{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, i \neq j \\ \sigma_\varepsilon^2, i = j \end{cases} ; \\ \varepsilon_i \to N(0, \sigma_\varepsilon^2) \end{cases}$$

$y_i = b + a \cdot x_i + \varepsilon_i$, $i = 1,..,$n

Ipotezele sunt formulate asupra variabilei:

$$\begin{cases} E(y_i | X = x_i) = b + a \cdot x_i \\ \operatorname{cov}(yi, yj) = \begin{cases} 0, i \neq j \\ \sigma_\varepsilon^2, i = j \end{cases} \\ y_i \to N(b + a \cdot x_i \cdot \sigma_\varepsilon^2) \end{cases}.$$

Când între cele două variabile există o dependenţă liniară, folosind serii de date (y_i,x_i), $i = 1, n$, valorile variabilei rezultative se estimează prin relaţia: $\hat{y}_i = \hat{b} + \hat{a}x_i$, seria reziduurilor fiind estimată din egalitatea: $e_i = y_i - \hat{y}_i = y_i - (\hat{b} + \hat{a}x_i)$.

Determinarea parametrilor modelului liniar se face, de regulă, prin utilizarea **Metodei celor mai mici pătrate** sau a verosimilităţii maxime.

Utilizând Metoda celor mai mici pătrate, valorile caracteristicii rezultative sunt estimate pe baza relaţiei:

$\hat{y}_i = \hat{b} + \hat{a}x_i$, unde $\hat{a}$ şi $\hat{b}$ sunt estimatorii parametrilor dreptei de regresie.

Valorile reale ale caracteristicii rezultative sunt egale cu estimaţia obţinută cu ajutorul modelului de regresie, corectată cu eroarea reziduală:

$y_i = \hat{y}_i + e_i$

În estimarea parametrilor se are în vedere condiţia ca suma pătratelor diferenţelor dintre valoarea reală şi cea estimată să fie minimă, obţinând egalitatea:

$$\min_{\hat{a},\hat{b}} \phi(\hat{a},\hat{b}) = \min_{\hat{a},\hat{b}} \sum_i^n e_i^2 = \min_{\hat{a},\hat{b}} \sum_i^n (y\hat{i} - \hat{b} - \hat{a}x_i)^2 .$$

Condiţiile de optim ale funcţiei conduc la un sistem de două ecuaţii:

$$\begin{cases} \dfrac{\partial(\hat{a},\hat{b})}{\partial(\hat{b})} = -\sum_i 2(y\hat{i} - \hat{b} - \hat{a}x_i) = 0 \\ \dfrac{\partial(\hat{a},\hat{b})}{\partial(\hat{a})} = -2\sum_i (y\hat{i} - \hat{b} - \hat{a}x_i) \cdot x_i = 0 \end{cases}$$

Ecuaţiile sunt stabilite aplicând metoda momentelor.

- Prima ecuaţie rezultă din condiţia $E(\varepsilon_i) = 0$, definind egalitatea:

$\frac{1}{n}\sum_i e_i = 0$ sau $\sum_i e_i = 0$;

- A doua ecuaţie a sistemului se poate stabili plecând de la ipoteza de necorelare a seriilor valorilor variabilei factoriale cu cea a valorilor variabilei reziduale ($\operatorname{cov}(X,\varepsilon)=0$), satisfăcând egalitatea:

$$\frac{1}{n}\sum_i x_i e_i = 0 \cdot$$

În vederea determinării celor doi estimatori se rezolvă sistemul liniar de ecuaţii rezultat:

$$\begin{cases} n\hat{b} + \hat{a}\left(\sum_i x_i\right) = \sum_i y_i \\ \left(\sum_i x_i\right)\cdot \hat{b} + \hat{a}\left(\sum_i x_i\right) = \sum_i y_i \end{cases} .$$

Testarea dacă soluţia sistemului îndeplineşte condiţiile de ordinul al doilea se face prin determinarea derivatelor de ordinul al doilea ale funcţiei[1].

Matricea rezultată are două proprietăţi: este pozitiv definită, iar determinantul matricei este pozitiv, respectiv:

$$\Delta X = 4n\sum_i x_i^2 - 4\left(\sum_i xi\right)^2 = 4n\left[\sum_i (x_i - \overline{x})^2\right] > 0$$

Relaţiile de calcul ale celor doi estimatori, $\hat{a}$ şi $\hat{b}$, rezultă din rezolvarea sistemul liniar de ecuaţii.

Coeficientului pantei dreptei de regresie se obţine din relaţia:

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} .$$

Utilizarea Metodei celor mai mici pătrate are şi unele inconveniente:

- nu oferă rezultate acceptabile dacă nu sunt satisfăcute ipotezele formulate;
- notând prin $\hat{a}^n, \hat{b}^n$ estimatorii determinaţi pe baza seriei (x_i, y_i), $i=\overline{1,n}$ iar prin $\hat{a}^{n+1}, \hat{b}^{n+1}$ pe cei evaluaţi pentru seria de valori (x_i, y_i), $i=\overline{1,n+1}$, rezultă că între cele două perechi de estimatori nu există o relaţie simplă de recurenţă;

1. Bardsen, G., Nymagen, R., Jansen, E. (2005) – *„The Econometrics of Macroeconomic Modelling”,* Oxford University Press

- estimatorii sunt distorsionaţi dacă seriile de date prezintă schimbări majore.

Utilizarea Metodei verosimilităţii maxime în estimarea parametrilor are la bază specificarea funcţiei de repartiţie reziduale.

Variabila reziduală are proprietatea:

$$\varepsilon_i \in N(0,\sigma_\varepsilon) \Leftrightarrow f(\varepsilon_i) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-\frac{\varepsilon_i^2}{2\sigma_\varepsilon^2}}.$$

Rezultă $yi \in N(\tilde{b},\tilde{a}x_i,\tilde{\sigma}_\varepsilon)$. Modelul de regresie devine specificat când sunt determinaţi parametrii $\tilde{a},\tilde{b}$ şi $\tilde{\sigma}_{\varepsilon,\varepsilon}$.

Pentru modelul liniar de regresie, funcţia de verosimilitate este dată de relaţia:

$$\ell(\tilde{a},\tilde{b},\tilde{\sigma}_\varepsilon^2) = \prod_{i=1}^{n} f(y_i / x_i).$$

Determinarea formei estimatorilor se face utilizând condiţiile de maximum pentru logaritmul funcţiei de verosimilitate.

$$L(a,b,\sigma_\varepsilon^2) = \ln l(a,b,\sigma_\varepsilon^2) = -\frac{n}{2}\left[\ln(2\pi) - \ln\sigma_\varepsilon^2\right] - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}(y_i - b - ax_i)^2$$

Pe baza proprietăţii funcţiei logaritm, se obţine:

$$\max_{\tilde{a},\tilde{b},\tilde{\sigma}_\varepsilon^2} \ell(\tilde{a},\tilde{b},\tilde{\sigma}_\varepsilon^2) \Leftrightarrow \max_{\tilde{a},\tilde{b},\tilde{\sigma}_\varepsilon^2} L(\tilde{a},\tilde{b},\tilde{\sigma}_\varepsilon^2).$$

Metoda verosimilităţii maxime se obţine acelaşi set de estimatori pentru parametrii modelului ca în cazul aplicării Metodei celor mai mici pătrate.

În cazul utilizării Metodei verosimilităţii maxime se obţine direct şi estimatorul dispersiei variabilei reziduale.

Scopul regresiei simple este de a evidenţia relaţia dintre o variabilă dependentă explicată (endogenă, rezultativă) şi o variabilă independentă (explicativă, factorială, exogenă, predictori).

Exemplu:

Pentru a putea construi un Model liniar de regresie a fost definit consumul final drept variabilă independentă, în timp ce valoarea Produsului Intern Brut a fost considerată a fi o variabilă dependentă.

Pentru a determina parametrii modelului liniar de regresie s-a considerat o serie de date privind la evoluţia celor doi indicatori macroeconomici de rezultate.

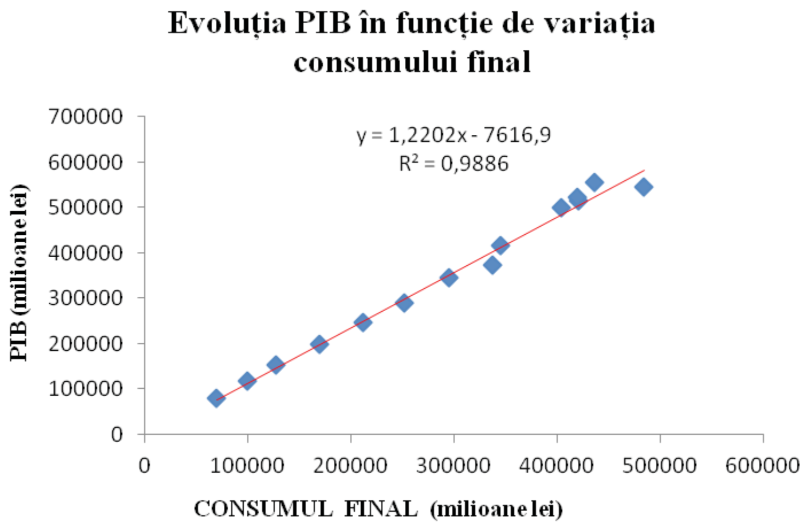
Evoluţia Produsului Intern Brut şi a consumului final al României între anii 1998-2011

milioane lei

Anul	PIB Y_X	CONSUMUL FINAL X
1998	373798,2	337468,6
1999	545730,2	484361,5
2000	80377,3	69253,3
2001	116768,7	99473,7
2002	152017,0	127118,8
2003	197427,6	168818,7
2004	247368,0	211054,6
2005	288954,6	251038,1
2006	344650,6	294867,6
2007	416006,8	344937,0
2008	514700,0	420917,5
2009	501139,4	404275,5
2010	523693,3	419801,2
2011	556708,4	436485,0

Sursa: Anuarul Statistic al României- Produsul intern brut, pe categorii de utilizări / I.N.S.

Pentru a identifica tipologia funcţiei de regresie s-a realizat reprezentarea grafică a perechilor de puncte ce cuprind valorile Produsului Intern Brut şi cele ale consumului final corespondente.



Se poate aprecia că între Produsul Intern Brut şi consumul final, există o legătură directă şi de formă liniară, respectiv $Y_x = a + bX + \varepsilon$.

Y_x - este variabila dependentă (explicată, endogenă, rezultativă),

a - ***Y*** intercept (termenul constant),

b - panta dreptei de regresie,

X - vectorul variabilei independente (explicative, exogene),

ε - o variabilă, interpretată ca eroare (perturbare, eroare de măsurare).

Este rezonabil să presupunem că media variabilei Y depinde de X printr-o relaţie liniară. Din calculele efectuate utilizând funcţia modelului de regresie liniară obţinem parametrii

a = -7616,882095 şi b = 1,220180278. Funcţia de regresie devine: $\overline{Y_X}$ = -7616,882095 + 1,220180278 *X*.

Pe baza datelor prezentate, cu ajutorul programului Excel/Data Analysis, au fost obţinute rezultatele:

Estimarea modelului de regresie în Excel

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.9942589							
R Square	0.9885508							
Adjusted R Square	0.9875967							
Standard Error	18784.648							
Observations	14							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regression	1	365605382319.68	365605382319.68	1036.1114	0.00			
Residual	12	4234356045	352863003.7					
Total	13	369839738364.30						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-7616.882	12109.51335	-0.628999851	0.5411382	-34001.2	18767.4809	-34001.25	18767.4809
CF	1.2201803	0.037907119	32.18868441	5.106E-13	1.137588	1.30277279	1.1375878	1.30277279

Multiple R este coeficientul de corelaţie multiplă, în acest caz, de corelaţie simplă între x şi y. Între valoarea Produsului Intern Brut şi cea a consumului final înregistrat în ţara noastră în anii 1998 - 2011 există o legătură directă şi foarte puternică, concluzie formulată pe baza valorii lui Multiple R(0,9942).

R Square, R^2 este coeficientul de determinaţie, care arată validitatea modelului ales, pentru explicarea variaţiei lui y; Multiple R se obţine din R Square: $r=\sqrt{R^2}$, iar în acest exemplu $\mathbf{R^2 = 0,9885}$ este o valoare apropiată de 1, indicând că modelul este bine ales, consumul final, x, explică variaţia Produsului Intern Brut, y, într-o proporţie de 98,85%.

Adjusted R Square este un coeficient de determinaţie corectat cu grade de libertate având aceeaşi semnificaţie ca şi R^2.

Standard Error este eroarea standard care arată cu cât se abat în medie valorile observate y_i de la valorile teoretice aflate pe dreapta de regresie, $\hat{y}_i$ (în acest caz cu ±18784,648). Valoare ridicată la puterea a 2-a reprezintă dispersia reziduurilor.

Observations este n, numărul de observări, n=14.

ANOVA reprezintă tabelul de analiză a variantei. Pentru varianta datorată factorului x, Regression, varianta reziduală, datorată celorlalţi factori neînregistraţi, Residual şi varianta totală, datorată tuturor factorilor, Total, se specifică:

- df (degrees freedom), gradele de libertate: k – numărul de variabile explicative x (regresie simplă k=1), n-k-1 pentru reziduuri (14-1-1=12 grade de libertate) şi n-1 pentru total variaţie (14-1=13); Suma df pentru Regression şi Residual este egală cu df pe Total: k + (n – k – 1) = n – 1.

- SS (prescurtarea de la Sum Square) - suma pătratelor abaterilor, numite variante:

- Regression: $SS^2_{y/x} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$,
- Residual: $SS^2_{y/z} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$,
- Total: $SS^2_y = \sum_{i=1}^{n}(y_i - \bar{y})^2$.

Între variante există relaţia: Total= Regression + Residual, adică $SS^2_y = SS^2_{y/x} + SS^2_{y/z}$.

MS (prescurtarea de la Modified Sum) numite sume modificate, de fapt, dispersii modificate:

- Regression: $MS^2_{y/x} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{k}$, dispersia datorată modelului de regresie ales,

- Residual: $MS^2_{y/z} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$, dispersia reziduurilor.

F, testul Fisher de semnificaţie globală a regresiei, reprezintă raportul dintre cele două dispersii corectate cu gradele de libertate $F = \frac{MS^2_{y/x}}{MS^2_{y/z}}$.

Intercept este denumirea pentru termenul liber (constant) al modelului.

Coefficients – conţine valorile estimate ale coeficienţilor a şi b. Din valorile prezentate rezultă modelul estimat în exemplu:

PIB = -7616,882095 + 1,220180278 CF.

Valabilitatea modelului de regresie este confirmată de valorile testelor F –statistic (1036,1114- valoare mult superioară nivelului tabelat considerat a fi reper în analizele de valabilitate a modelelor econometrice), precum şi de gradul de risc nul (reflectat prin valoarea testului Significance F).

Lower 95%, Upper 95% – limitele inferioară şi superioară ale intervalului de încredere pentru parametrul respectiv. Limitele la pragul 0,05 sunt calculate automat, indiferent de iniţializarea procedurii Regression. Se poate interpreta că parametrii modelului liniar sunt cuprinşi în intervalele următoare:

$-34001.24512 < a < 18767.48093$;

$1.137587762 < b < 1.302772794$.

RESIDUAL OUTPUT			
Observation	*Predicted PIB*	*Residuals*	*Standard Residuals*
1	404155.648	-30357.448	-1.682066841
2	583391.4675	-37661.26752	-2.086762013
3	76884.62874	3492.671264	0.193524387
4	113758.9648	3009.735197	0.166765526
5	147490.9706	4526.0294	0.250781422
6	198372.3662	-944.7661657	-0.052348269
7	249907.7784	-2539.77836	-0.140725827
8	298694.8565	-9740.256497	-0.539694986
9	352174.748	-7524.147982	-0.416903286
10	413268.4424	2738.357618	0.151728846
11	505978.35	8721.650022	0.483255322
12	485672.1098	15467.29021	0.8570225
13	504616.2627	19077.03727	1.057033905
14	524973.5065	31734.89355	1.758389313

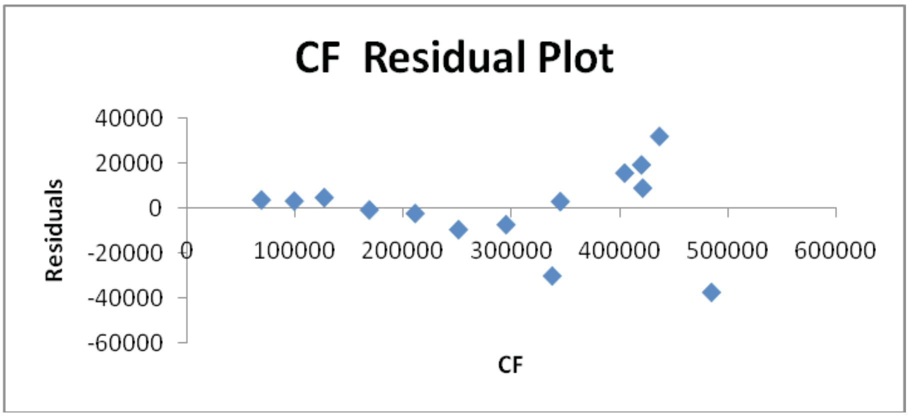
Predicted y - valoarea y prognozată pentru observaţia respectivă; se obţine înlocuind valorile X ale observaţiei în $\overline{Y_X} = -7616{,}882095 + 1{,}220180278X$, modelul estimat. Rezultă că suma valorilor ajustate $\overline{Y_X}$ este egală cu suma valorilor empirice Y_X, ceea ce permite să afirmăm că estimarea parametrilor ecuaţiei de regresie este corectă.

Residuals – valoarea erorii de predicţie (diferenţa dintre valoarea observată şi valoarea prognozată).

Standard Reziduals – valoarea standardizată a erorii (obţinută prin împărţirea reziduului la abaterea standard a reziduurilor).

Analiza calităţii modelului de analiză ales (prezentare grafică):

Diagrama variabilă independentă vs. reziduuri

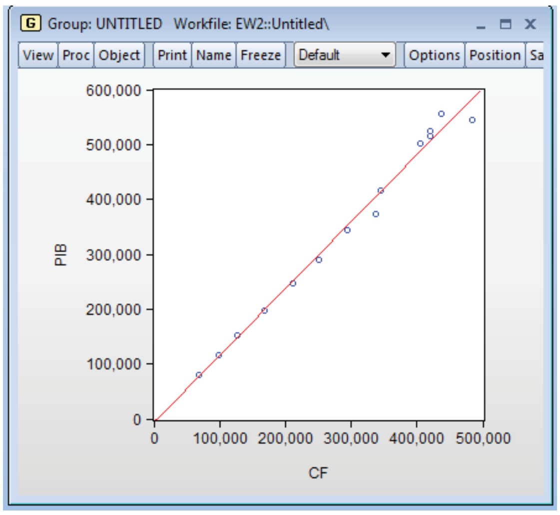


După forma norului de puncte, rezultă că nu există corelaţie între variabila independenta *x* şi reziduuri, modelul fiind bine ales.

Pentru a analiza corelația dintre evoluția Produsului Intern Brut și cea a consumului final s-a supus cercetării seria de date ce cuprinde valorile dintre anii 1998 – 2011, prelucrate cu ajutorul pachetului informatic Eviews7.

Identificarea tipului de model econometric care reflectă evoluția fenomenului studiat. S-a generat graficul perechilor de puncte PIB – CF.

Diagrama consum final vs. P.I.B.

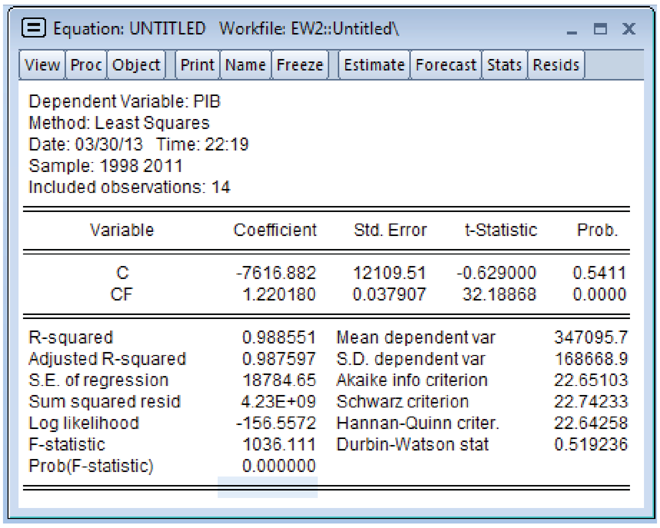


Perechile de puncte din grafic urmăresc traiectoria unei drepte, fiind posibilă analiza fenomenului cercetat cu ajutorul modelului de regresie liniară simplă.

În cadrul cercetării s-a utilizat Produsul Intern Brut drept variabilă dependentă, iar consumul final variabilă independentă. În cadrul modelului, s-a introdus termenul liber C.

Rezultatele obținute cu ajutorul programului Eviews:

Caracteristicile modelului de regresie


Equation: UNTITLED Workfile: EW2::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PIB
Method: Least Squares
Date: 03/30/13 Time: 22:19
Sample: 1998 2011
Included observations: 14

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7616.882	12109.51	-0.629000	0.5411
CF	1.220180	0.037907	32.18868	0.0000

R-squared	0.988551	Mean dependent var	347095.7
Adjusted R-squared	0.987597	S.D. dependent var	168668.9
S.E. of regression	18784.65	Akaike info criterion	22.65103
Sum squared resid	4.23E+09	Schwarz criterion	22.74233
Log likelihood	-156.5572	Hannan-Quinn criter.	22.64258
F-statistic	1036.111	Durbin-Watson stat	0.519236
Prob(F-statistic)	0.000000		

Rezultă că modelul de regresie liniară simplă ce reflectă corelația dintre valorile Produsului Intern Brut și cele ale consumului final se prezintă astfel:

PIB = -7616,8 + 1,22 CF.

Concluzie

Consumul final constituie un factor extrem de important pentru evoluția PIB. Pentru o creștere cu o unitate monetară a consumului final se va obține o majorare cu 1,22 unități monetare a PIB.

Se remarcă faptul că valoarea termenului liber C este una foarte ridicată, ceea ce ne permite să afirmăm faptul că factorii ce nu au fost considerați la construcția modelului prezintă o influență destul de ridicată asupra evoluției Produsului Intern Brut.Valoarea negativă a termenului liber relevă faptul că variabilele ce nu au fost incluse în modelul econometric au un efect negativ asupra evoluției PIB.

Bibliografie selectivă:

- Anghelache C., (2008) – „Tratat de statistică teoretică şi economică”, Editura Economica, Bucureşti
- Anghelache, C (coordonator) şi alţii (2012) – „Modele statistico– econometrice de analiză economică – utilizarea modelelor în studiul economiei României”, Revista Română de Statistică (Supliment), ISBN 1018—046X
- Anghelache, C. şi alţii (2012) – „*Econometrie*”, Editura Artifex, Bucureşti
- Anghelache, C. şi alţii (2012) – „*Elemente de econometrie teoretică şi aplicată*”, Editura Artifex, Bucureşti
- Anghelache, C., Mitruţ, C. (coordonatori), Bugudui, E., Deatcu, C. (2009) – „*Econometrie: studii teoretice şi practice*”, Editura Artifex, Bucureşti
- Bardsen, G., Nymagen, R., Jansen, E. (2005) – *„ The Econometrics of Macroeconomic Modelling”,* Oxford University Press
- Turdean, M.S., Prodan L., (2012) – „Statistică pentru afaceri”, Editura ProUniversitaria, Bucuresti, ISBN 978-606-647-312-5
- Voineagu, V., Ţiţan, E. şi colectiv (2007) – “*Teorie şi practică econometrică*”, Editura Meteor Press
- ***Anuarul Statistic al României, Ediţiile 2008 - 2012, Institutul Naţional de Statistică