ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI
FACULTATEA DE CIBERNETICĂ, STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ

Emigrația temporară, anul 2016, Romania
Proiect Econimetrie

Profesor coordonator: Conf. univ. dr. Mihaela Covrig

Proiect realizat de Petruța-Florentina IVAN
Grupa: 1055 Seria: F
Data: 28/05/2019

Cuprins

1. Introducere. Motivarea alegerii temei 2
2. Descrierea datelorce vor fi utilizate în modelul econometric 3
3. Modelul econometric unifactorial 6
   3.1. Definirea modelului econometric unifactorial 6
   3.2. Coeficienții modelului de regresie liniară simplă și interpretarea lor 6
4. Testarea validității modelului de regresie 8
5. Testarea semnificaţiei parametrului pantă β_1 10
6. Homoscedasticitatea erorilor aleatoare 11
   6.1. Testul White 11
   6.2. Testul Glejser 13
7. Corectarea heteroscedasticităţii erorilor aleatoare 15
   7.1. MCMMP-Ponderată 15
8. Non-autocorelarea erorilor aleatoare 16
   8.1. Testul Durbin-Waston 16
   8.2. Testul BREUSCH-GODFREY 17
9. Normalitatea erorilor aleatoare 19
   9.1. Testul Jarque-Bera (JB) privind distribuţia normală a erorilor 19
10. Previziune 19
11. Modelul econometric multifactorial 21
12. Testarea validitatii modelului de regresie multifactoriala 24
13. Testarea validitatii modelului de regresie multifactoriala 24
14. Remedii pentru Multicoliniaritate: Transformarea variabilelor 26
    14.1. Testul Jarque-Bera (JB) privind distribuţia normală a erorilor 27
15. Concluzii 28
16. Bibliografie 29

Introducere

„A doua componentă a dinamicii demografice este migrația internațională […] Imigrările au aceeaşi semnificaţie cu naşterile, emigrările reprezintă pierderi…”
(Vladimir Trebici, 1991)

Potrivit datelor Statisticii, România continuă să fie o ţară de emigrare, fenomenul migraţiei externe continuând să fie a doua cauză principală a reducerii populaţiei ţării, după scăderea generată de sporul natural negativ (în 2016 atingând valoarea de -54316).

Românii au plecat din România, masiv, în fluxuri intense sau mai puţin intense, încă din anii 1990, însă exilul începuse ceva mai devreme, încă sub regimul Ceauşescu, când oamenii nu mai sperau că ceva se va schimba vreodată semnificativ în bine în România. Senzaţia aceasta a rămas și determină mulţi români să ia drumul emigrării.

În 2016, o treime dintre cei 210.000 de români care au plecat în străinătate au fost tineri cu vârsta cuprinsă între 20 și 30 de ani, arată datele Statisticii. Migraţia în rândul tinerilor cu vârstă de muncă a continuat astfel să se accentueze, în ciuda faptului că salariile au crescut.

”Soldul migraţiei internaţionale în anul 2016 a fost negativ, numărul emigranţilor depăşind numărul imigranţilor cu peste 76.000 persoane. În cursul anului 2016, bărbaţii au emigrat într-o proporţie mai mare decât femeile (55,4%). Şi în rândul imigranţilor bărbaţii au fost majoritari (53,6%)”, se arată în datele Statisticii.

Emigranți temporari sunt persoanele care emigrează în străinatate pentru o perioadă de cel puțin 12 luni. Emigrație înseamnă acțiunea prin care o persoană care a avut anterior reședința obișnuită pe teritoriul României, încetează să mai aibă reședinta obișnuită pe teritoriul acesteia pentru o perioadă care este sau se așteaptă să fie de cel puțin 12 luni. Reședința obișnuită reprezintă locul în care o persoană își petrece în mod obișnuit perioada zilnică de odihnă, fără a ține seama de absențele temporare pentru recreere, vacanțe, vizite la prieteni și rude, afaceri, tratamente medicale sau pelerinaje religioase.

Emigrația temporară este influențată de mai mulți factori atât economici cât și sociali. În proiect se urmărește analiza modelului economic pe tema emigrației temporare pe județe, în anul 2016, în raport cu câștigul salarial nominal net și somajul înregistrat, doi factori economici esențiali în economia unei țări.

Descrierea datelor ce vor fi utilizate în modelul econometric

Variabilele sunt:

Y-variabila care arată valoarea emigrațiilor temporare pentru fiecare județ, exprimată în număr de persoane (variabila dependentă sau variabila explicată sau variabila endogenă sau variabila răspuns).

X_1-variabila care arată valoarea câștigului salarial nominal net pentru fiecare județ, exprimată în Lei (variabila independentă sau variabila explicativă sau variabila exogena sau variabila factor).

X_2- variabila care arată valoarea somajului înregistrat pentru fiecare județ, exprimată în numar de persoane (variabila independentă sau variabila explicativă sau variabila exogena sau variabila factor).

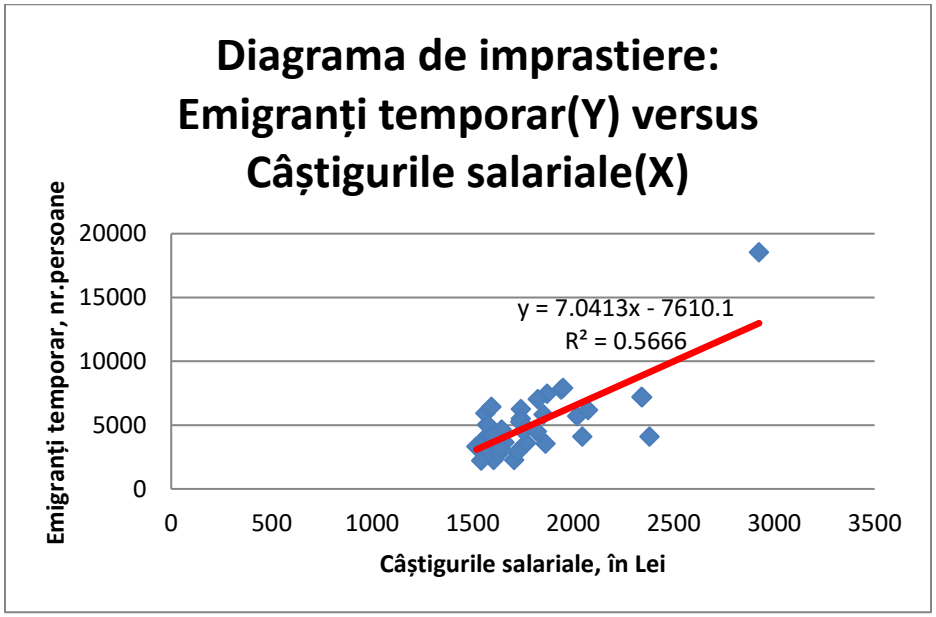


Figura1: Emigranți temporar versus Câștigul salarial nominal net-diagrama de împrăștiere și dreapta de regresie estimată.

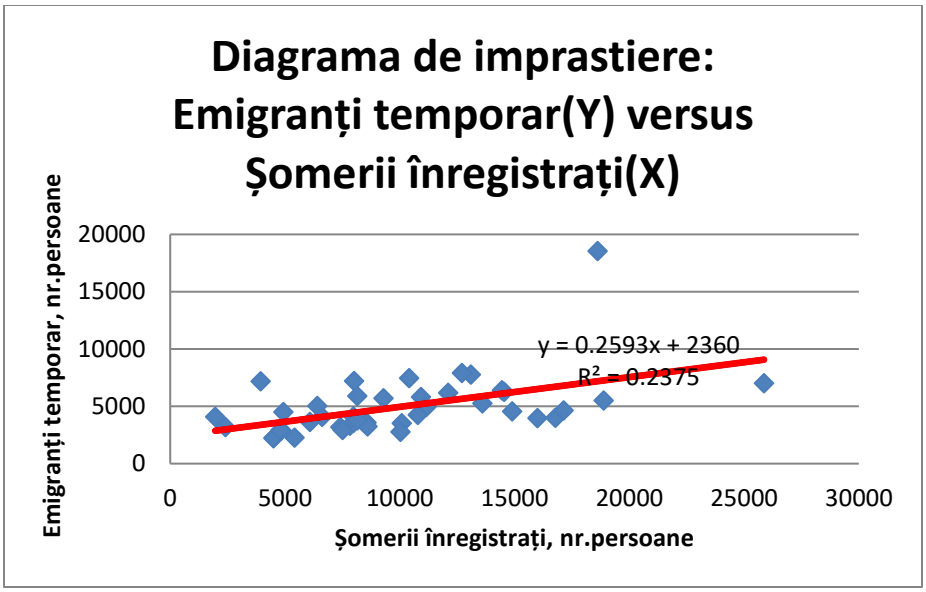


Figura2: Emigranți temporar versus Șomerii înregistrați-diagrama de împrăștiere și dreapta de regresie estimată.

În urma efectuării diagramelor de împrăștiere se observă faptul că emigrația temporară este mai influențată de câștigul salarial nominal net.

Sunt selectate cele 42 de județe ale României și datele necesare pentru cele 3 variabile din anul 2016:

Nr.crt	JUDETE	Emigranti temporar	Castigul Salarial	Somerii inregistrati
1	Bihor	5912	1563	8155
2	Bistrita-Nasaud	2924	1587	4857
3	Cluj	7197	2340	8020
4	Maramures	5044	1572	6409
5	Satu Mare	3632	1659	6095
6	Salaj	2261	1604	5427
7	Alba	3578	1767	8566
8	Brasov	5700	2020	9306
9	Covasna	2214	1542	4497
10	Harghita	3313	1519	7829
11	Mures	5823	1853	10924
12	Sibiu	4093	2046	6617
13	Bacau	6257	1739	14560
14	Botosani	4110	1571	7994
15	Iasi	7766	1940	13104
16	Neamt	4960	1575	11171
17	Suceava	6414	1593	14453
18	Vaslui	4001	1627	16791
19	Braila	3255	1587	8609
20	Buzau	4638	1644	17144
21	Constanta	7453	1871	10421
22	Galati	5504	1740	18889
23	Tulcea	2261	1706	4517
24	Vrancea	3547	1556	7993
25	Arges	6168	2074	12115
26	Calarasi	3203	1624	7412
27	Dambovita	5246	1736	13610
28	Giurgiu	2917	1724	4809
29	Ialomita	2945	1579	7514
30	Prahova	7899	1950	12718
31	Teleorman	3961	1591	15999
32	Ilfov	4089	2380	1962
33	Municipiul Bucu	18536	2925	18619
34	Dolj	7027	1824	25869
35	Gorj	3531	1862	10095
36	Mehedinti	2770	1630	10034
37	Olt	4554	1753	14897
38	Valcea	3778	1601	8153
39	Arad	4492	1819	4937
40	Caras-Severin	3184	1637	2413
41	Hunedoara	4249	1600	10789
42	Timis	7172	2343	3944

Modelul econometric unifactorial

Y-variabila care arată valoarea emigrațiilor temporare pentru fiecare județ, exprimată în număr βde persoane (variabila dependentă sau variabila explicată sau variabila endogenă sau variabila răspuns).

X-variabila care arată valoarea câștigului salarial nominal net pentru fiecare județ, exprimată în Lei (variabila independentă sau variabila explicativă sau variabila exogena sau variabila factor).

Definirea modelului econometric:

$$Y = f(X) + \varepsilon,$$

Unde $f(x) = \beta_0 + \beta_1 \cdot x,$

Adică, $Y = \beta_0 + \beta_1 \cdot x + \varepsilon \Leftrightarrow y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \Leftrightarrow$

$$\text{Emigranți}_i = \beta_0 + \text{CâștigulS}_i + \varepsilon_i$$

Pentru fiecare dintre cele n=42 județe selectate s-au înregistrat valorile celor două variabile, X şi Y, obţinându-se astfel seria bivariată de date statistice.

Pe baza eșantionului n=42 s-au determinat estimatorii $\hat{\beta}_0$ și $\hat{\beta}_1$ ai parametrilor β_0-intercept, respectiv β_1-parametrul pantă al modelului de regresie prin metoda celor mai mici patrate (MCMMP).

$$\hat{\beta}_0 = -7610.0793 \qquad \hat{\beta}_1 = 7.041274$$

Modelul de regresie estimat: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i + \hat{\varepsilon}_i$

Interpretarea valorilor coeficienţilor :

$\hat{\beta}_0$, intercept, arată nivelul mediu al variabilei dependente la nivelul eșantionului, dacă nivelul variabilei independente ar fi egal cu 0 unități; $\hat{\beta}_0$ nu are întotdeauna interpretare economică; dacă numărul de emigranți temorari este 0 , atunci câștigul salarial nominal net ar avea o valoare negativă de -7610.0793, situație puțin probabilă.

$\hat{\beta}_1$, panta dreptei de regresie, arată cu câte unități se modifică, în medie, la nivelul eșantionului, nivelul variabilei dependente, dacă nivelul variabilei independente (explicative) crește cu o unitate; astfel, valoarea câștigului salarial crește, în medie, cu 7.041274 lei dacă numarul de persoane emigrante temporar crește cu o unitate.

Group: UNTITLED Workfile: UNTITLED::Untitled\

obs	EMIGRANTI...	CASTIGUL_...
1	5912	1563
2	2924	1587
3	7197	2340
4	5044	1572
5	3632	1659
6	2261	1604
7	3578	1767
8	5700	2020
9	2214	1542
10	3313	1519
11	5823	1853
12	4093	2046
13	6257	1739
14	4110	1571
15	7766	1940
16	4960	1575
17	6414	1593
18		

Workfile: UNTITLED

Range: 1 42 -- 42 obs
Sample: 1 42 -- 42 obs

c
castigul_salarial
emigranti_temporar
judete
nr_crt
regresie1
regresie2
resid
somerii_inregistrati

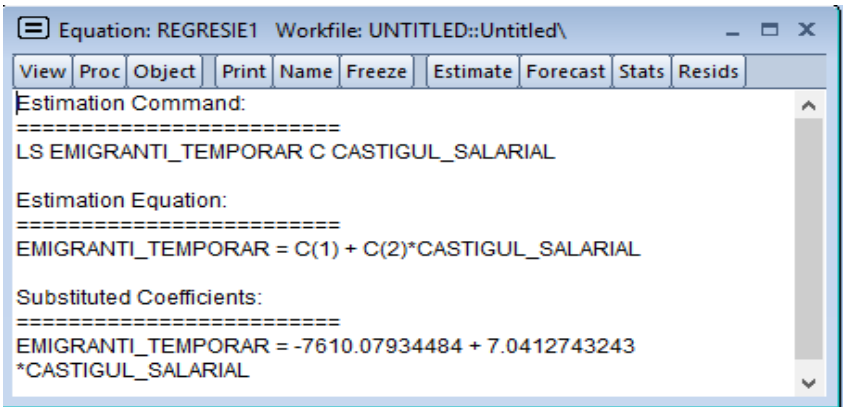
Equation: REGRESIE1 Workfile: UNTITLED::Untitled\

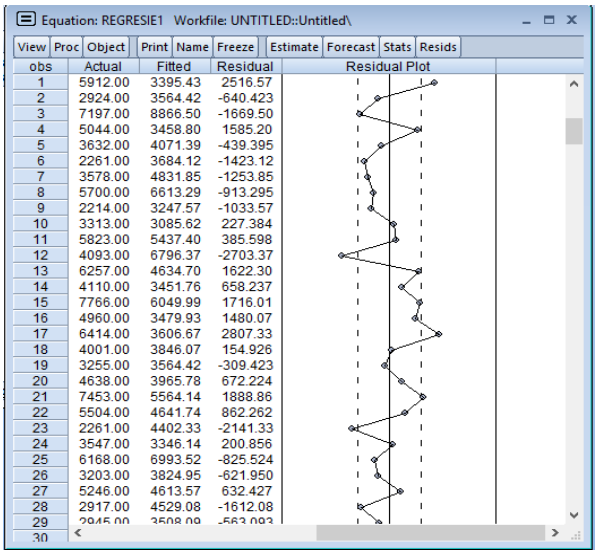
Dependent Variable: EMIGRANTI_TEMPORAR
Method: Least Squares
Date: 05/24/19 Time: 17:55
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7610.079	1757.452	-4.330178	0.0001
CASTIGUL_SALARIAL	7.041274	0.973646	7.231864	0.0000

R-squared	0.566630	Mean dependent var	4942.333
Adjusted R-squared	0.555796	S.D. dependent var	2679.773
S.E. of regression	1786.033	Akaike info criterion	17.85983
Sum squared resid	1.28E+08	Schwarz criterion	17.94258
Log likelihood	-373.0564	Hannan-Quinn criter.	17.89016
F-statistic	52.29986	Durbin-Watson stat	2.209005
Prob(F-statistic)	0.000000		

Untitled New Page





obs	Actual	Fitted	Residual
1	5912.00	3395.43	2516.57
2	2924.00	3564.42	-640.423
3	7197.00	8866.50	-1669.50
4	5044.00	3458.80	1585.20
5	3632.00	4071.39	-439.395
6	2261.00	3684.12	-1423.12
7	3578.00	4831.85	-1253.85
8	5700.00	6613.29	-913.295
9	2214.00	3247.57	-1033.57
10	3313.00	3085.62	227.384
11	5823.00	5437.40	385.598
12	4093.00	6796.37	-2703.37
13	6257.00	4634.70	1622.30
14	4110.00	3451.76	658.237
15	7766.00	6049.99	1716.01
16	4960.00	3479.93	1480.07
17	6414.00	3606.67	2807.33
18	4001.00	3846.07	154.926
19	3255.00	3564.42	-309.423
20	4638.00	3965.78	672.224
21	7453.00	5564.14	1888.86
22	5504.00	4641.74	862.262
23	2261.00	4402.33	-2141.33
24	3547.00	3346.14	200.856
25	6168.00	6993.52	-825.524
26	3203.00	3824.95	-621.950
27	5246.00	4613.57	632.427
28	2917.00	4529.08	-1612.08
29	2945.00	3508.09	-563.093
30			

Output-ul modelului de regresie Eviews.

Testarea validității modelului de regresie

Ipotezele statistice:

H_0: modelul de regresie $Y = \beta_0 + \beta_1 \cdot x + \varepsilon$ nu este valid sau nu este semnificativ statistic (toți parametrii variabilelor explicative din modelul de regresie nu sunt semnificativi statistic, în cazul nostru $\beta_1 = 0$),

cu alternativa

H_1: modelul de regresie $Y = \beta_0 + \beta_1 \cdot x + \varepsilon$ este valid sau semnificativ statistic (pentru cel puțin o variabilă explicativă din model, parametrul corespunzător este semnificativ statistic, adică semnificativ diferit de 0; în cazul nostru $\beta_1 \neq 0$).

Testul statistic este:

$$F = \frac{MSE}{MSR} = \frac{\frac{SSE}{k}}{\frac{SSR}{n-k-1}} \sim Fisher_{(k=1,n-k-1=40)}$$

unde k este numărul de variabile explicative din modelul de regresie (în cazul nostru, k=1 deoarece avem un model de regresie liniară unifactorială sau șimplă, cu o șingură variabilă explicativă).

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.752748227
R Square	0.566629893
Adjusted R Square	0.55579564
Standard Error	1786.032965
Observations	42

ANOVA					
	df	SS	MS	F	Significance F
Regressio	1	166832041.2	166832041	52.29986	8.93578E-09
Residual	40	127596550.1	3189913.75		

	Coefficients	Standard Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.(	Upper 95.(
Intercept	-7610.079345	1757.45171	-4.3301784	9.712E-05	-11162	-4058.14	-11162	-4058.14
Castigul S	7.041274324	0.97364582	7.23186418	8.936E-09	5.073463	9.009086	5.073463	9.009086

Output-ul estimării modelului de regresie în Excel.

$F_{critic} = F_{(0,05;1,40)} = 4.084745733$

$Regiunea_{critică}: F > F_{critic} = 4.084745733$

Din tabelul ANOVA : $F_{calculat} = 52.29986$

Decizia:

Cum $F_{calculat} = 52.29986 > F_{critic} = 4.084745733$, atunci respingem ipoteza nulă H_0 la nivelul de seminificație $\propto = 0.05$ și acceptăm ipoteza alternativă H_1 concluzionând că datele de selecție sunt în favoarea ipotezei alternative, adică modelul de regresie este valid statistic.

Pentru $Significance\ F = 8.93578E - 09 = 8.93578 \cdot 10^{-9}$, deci ȘignificanceF $<\propto = 0.05$, prin urmare respingem H_0 și acceptăm H_1 modelul de regresie fiind valid statistic.

Coeficientul de determinație:

R-squared = 0.566630, arată că la nivelul eșantionului, 56,67% (adică 100•R-squared%) din variația totală a variabilei dependente (emigranți temporar pe județ) este explicată de variația totală a variabilei independente (câștigul salarial nominal net pe fiecare județ). Restul de 43,33% din variația totală a variabilei dependente este explicată de alți factori, numiți factori reziduali.

Testarea semnificaţiei parametrului pantă $\boldsymbol{\beta_1}$

Ipotezele testate:

H_0:$\beta_1 = 0$ (parametrul este nesemnificativ statistic sau parametrul nu este semnificativ diferit de 0)

H_1:$\beta_1 \neq 0$ (parametrul este semnificativ statistic sau parametrul este semnificativ diferit de 0)

Statistica testului este:

$$t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim Student_{(n-k-1=40)}$$

Din tabelul coefficients din Excel:

$se(\hat{\beta}_1)$ =0.97364582

În ipoteza că H_0 este adevarată, $\beta_1 = 0$, și pe baza datelor de selecție, se obține valoarea calculată a statisticii testului:

$t_{calculat}(\hat{\beta}_1)$ =7.23186418

$Regiunea\ _{critică}$: $t < -t_{critic(\frac{\propto}{2}, n-k-1)}$ $și\ t > t_{critic(\frac{\propto}{2}, n-k-1)}$

$t_{critic} = t_{(0.05,40)} = 2.0210753$

Decizia:

Cum $t_{calculat} = 7.23186418 > t_{critic} = 2.0210753$, atunci respingem ipoteza nulă H_0, $\beta_1 = 0$, și acceptăm ipoteza alternativă H_1, $\beta_1 \neq 0$, concluzionând că parametrul β_1 este seminificativ statistic, seminificativ diferit de 0.

Pentru $P - value = 8.936E - 09 = 8.93 \cdot 10^{-9} < \propto = 0.05$, prin urmare respingem ipoteza nulă H_0, $\beta_1 = 0$, și acceptăm ipoteza alternativă H_1, $\beta_1 \neq 0$, concluzionând că parametrul β_1 este seminificativ statistic, seminificativ diferit de 0.

Homoscedasticitatea erorilor aleatoare

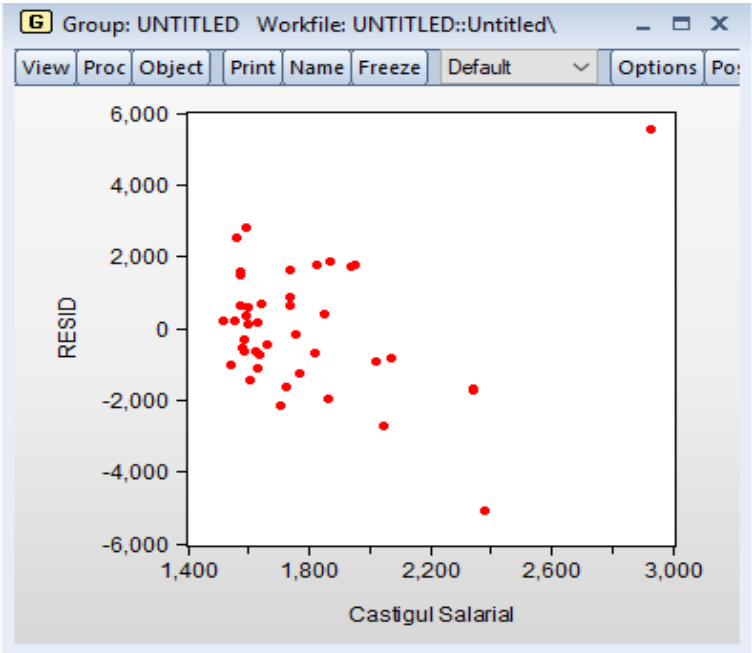


Figura3: Reziduri versus Câștigul salarial nominal net

Testul White

Modelul de regresie linear original:

$$Y = \beta_0 + \beta_1 \cdot x + \varepsilon \Leftrightarrow y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \Leftrightarrow$$

$$\text{Emigranți}_i = \beta_0 + \beta_1 \cdot \text{CastigulS}_i + \varepsilon_i$$

atunci modelul de regresie auxiliar este:

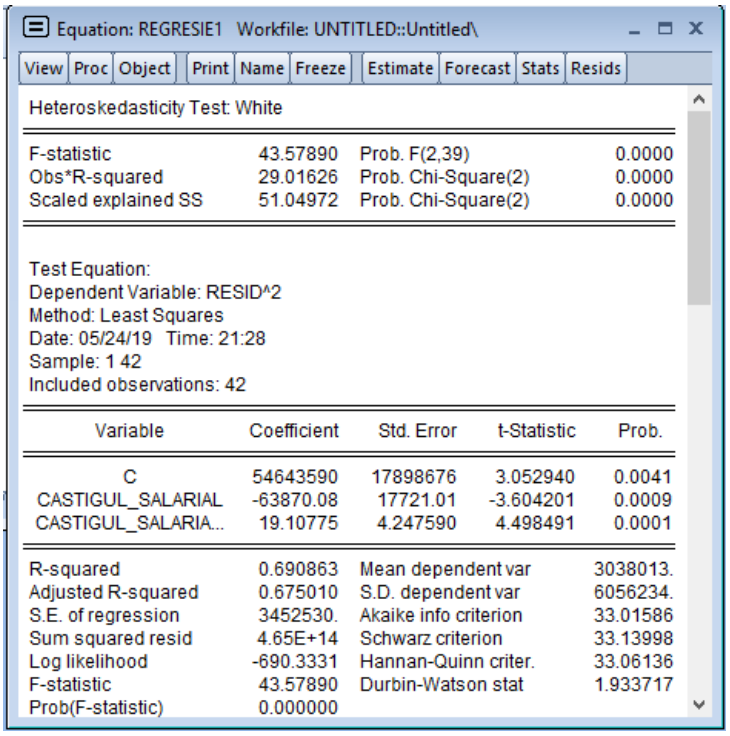
$$\widehat{\varepsilon}_i^{\,2} = \propto_0 + \propto_1 \cdot x_i + \propto_2 \cdot x_i^{\,2} + \eta_i \Leftrightarrow$$

$$\widehat{\varepsilon}_i^{\,2} = \propto_0 + \propto_1 \cdot C\hat{a}\hat{s}tigulS_i + \propto_2 \cdot C\hat{a}\hat{s}tigulS_i^{\,2} + \eta_i$$

Ipotezele sunt:

H_0: $\propto_1 = \propto_2 = 0$ (nu există heteroscedasticitate sau erorile aleatoare sunt homoscedastice)

H_1: $(\exists)$ $\propto_i \neq 0$ (există heteroscedasticitate sau erorile aleatoare sunt heteroscedasctice)



Equation: REGRESIE1 Workfile: UNTITLED::Untitled\

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

Heteroskedasticity Test: White

F-statistic	43.57890	Prob. F(2,39)	0.0000
Obs*R-squared	29.01626	Prob. Chi-Square(2)	0.0000
Scaled explained SS	51.04972	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/24/19 Time: 21:28
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	54643590	17898676	3.052940	0.0041
CASTIGUL_SALARIAL	-63870.08	17721.01	-3.604201	0.0009
CASTIGUL_SALARIA...	19.10775	4.247590	4.498491	0.0001

R-squared	0.690863	Mean dependent var	3038013.
Adjusted R-squared	0.675010	S.D. dependent var	6056234.
S.E. of regression	3452530.	Akaike info criterion	33.01586
Sum squared resid	4.65E+14	Schwarz criterion	33.13998
Log likelihood	-690.3331	Hannan-Quinn criter.	33.06136
F-statistic	43.57890	Durbin-Watson stat	1.933717
Prob(F-statistic)	0.000000		

Figura4: Output Testul White

Ecuația de regresie auxiliară:

$$\widehat{\varepsilon}_i^{\,2} = 54643590 - 63870.08 \cdot C\hat{a}\hat{s}tigulS_i + 19.10775 \cdot C\hat{a}\hat{s}tigulS_i^{\,2} + \eta_i$$

Se observă:

Valoarea calculată a testului statistic F, clașică, adică $F_{calculat} = 43.57890\sim$

$F - statistic$

cu probabilitatea asociată Prob. $F(2,39) = 0.0000 < \propto = 0.05$.

Valoarea calculată a testului statistic $W_{calculat} = nR_a{}^2 = 29.01626$ cu probabilitatea asociată Prob. $Chi - Square = 0.0000 < \propto = 0.05$.

Decizie:

Pentru Prob. $F(2,39) = 0.0000 < \propto = 0.05$ și Prob.$Chi - Square = 0.0000 < \propto = 0.05$, atunci respingem ipoteza nulă, H_0: $\propto_1 = \propto_2 = 0$ și acceptăm ipoteza alternativă, H_1: $(\exists)$ $\propto_i \neq 0$ rezultă erorile aleatoare sunt heteroscedastice.

Dacă $W_{calculat} = nR_a{}^2 = 29.01626 >$ $\chi_{critic(0.05,2)} = 5{,}99$, atunci respingem ipoteza nulă, H_0: $\propto_1 = \propto_2 = 0$ și acceptăm ipoteza alternativă, H_1: $(\exists)$ $\propto_i \neq 0$ rezultă erorile aleatoare sunt heteroscedastice.

Testul Glejser

Modelul de regresie linear original:

$$Y = \beta_0 + \beta_1 \cdot x + \varepsilon \Leftrightarrow y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \Leftrightarrow$$

$$\text{Emigranți}_i = \beta_0 + \beta_1 \cdot \text{CâștigulS}_i + \varepsilon_i$$

atunci modelul de regresie auxiliar este:

$$|\hat{\varepsilon}_i| = \beta_0 + \beta_1 \cdot x_i + \upsilon_i \Leftrightarrow$$

$$|\hat{\varepsilon}_i| = \beta_0 + \beta_1 \cdot Câș tigulS_i + \upsilon_i$$

Equation: EQ_03 Workfile: UNTITLED::Untitled\

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

Dependent Variable: ABSOLUTE_RESID
Method: Least Squares
Date: 05/24/19 Time: 22:31
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3717.921	825.9246	-4.501525	0.0001
CASTIGUL_SALARIAL	2.823068	0.457570	6.169689	0.0000

R-squared	0.487607	Mean dependent var	1314.735
Adjusted R-squared	0.474797	S.D. dependent var	1158.198
S.E. of regression	839.3565	Akaike info criterion	16.34960
Sum squared resid	28180776	Schwarz criterion	16.43234
Log likelihood	-341.3415	Hannan-Quinn criter.	16.37993
F-statistic	38.06506	Durbin-Watson stat	1.769738
Prob(F-statistic)	0.000000		

Equation: REGRESIE1 Workfile: UNTITLED::Untitled\

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

Heteroskedasticity Test: Glejser

F-statistic	38.06506	Prob. F(1,40)	0.0000
Obs*R-squared	20.47949	Prob. Chi-Square(1)	0.0000
Scaled explained SS	23.13552	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 05/24/19 Time: 22:48
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3717.921	825.9246	-4.501525	0.0001
CASTIGUL_SALARIAL	2.823068	0.457570	6.169689	0.0000

R-squared	0.487607	Mean dependent var	1314.735
Adjusted R-squared	0.474797	S.D. dependent var	1158.198
S.E. of regression	839.3565	Akaike info criterion	16.34960
Sum squared resid	28180776	Schwarz criterion	16.43234
Log likelihood	-341.3415	Hannan-Quinn criter.	16.37993
F-statistic	38.06506	Durbin-Watson stat	1.769738
Prob(F-statistic)	0.000000		

Testam ipotezele:

H_0:$\beta_1 = 0$ (nu există heteroscedasticitate, adică există homoscedasticitate)

H_1:$\beta_1 \neq 0$ (există hetersocesaticitate)

$$|\hat{\varepsilon}_i| = -3717.912 + 2.823068 \cdot CastigulS_i$$

se = (825.9246) (0.457570)

$R^2 = 0.487607$ $t_{statistic} = 6.1696$

Decizie:

Pentru Prob(p-*value*) = 0.0000 < $\propto = 0.05$ se respinge ipoteza nulă,deci acceptăm ipoteza alternativă, β_1 este semnificativ statistic,atunci erorile aleatoare ale modelului original de regresie sunt heteroscedastice.

Corectarea heteroscedasticităţii erorilor aleatoare

MCMMP-Ponderată

Conşiderăm cazul în care dispersiile (varianţele) perturbaţiilor sunt necunoscute: ${\sigma_i}^2$ necunoscut.

Presupunem că dispersia (varianţa) erorilor este proporţională cu pătratul unei variabile explicative, adică ${\sigma_i}^2 \cdot {x_i}^2$, atunci modelul de regresie $y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$ se transforma prin împărțirea la $x_i \Longrightarrow$

$$\frac{y_i}{x_i} = \frac{\beta_0}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i}$$

Equation: CORECTARE2 Workfile: UNTITLED::Untitled\

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

Dependent Variable: 1/EMIGRANTI_TEMPORAR
Method: Least Squares
Date: 05/25/19 Time: 02:20
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000165	9.37E-05	-1.764392	0.0853
1/CASTIGUL_SALARIAL	0.709272	0.162519	4.364245	0.0001

R-squared	0.322569	Mean dependent var	0.000240
Adjusted R-squared	0.305634	S.D. dependent var	9.15E-05
S.E. of regression	7.62E-05	Akaike info criterion	-16.07875
Sum squared resid	2.33E-07	Schwarz criterion	-15.99600
Log likelihood	339.6537	Hannan-Quinn criter.	-16.04842
F-statistic	19.04663	Durbin-Watson stat	1.948110
Prob(F-statistic)	0.000087		

Deoarece Prob(F-Statistic) = 0.000087 < α = 0.05, atunci putem afirma faptul că modelul este valid, erorile aleatoare sunt homoscedastice.

Non-autocorelarea erorilor aleatoare

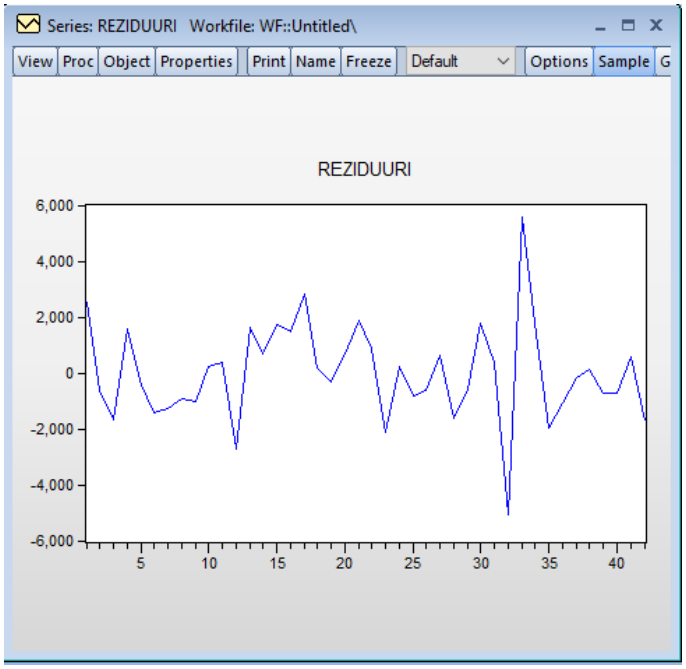


Figura5: Evoluția seriei rezidurilor celor 42 de județe.

Testul Durbin Waston

Ipotezele statistice:

$H_0: \rho = 0$ (nu există autocorelarea erorilor)

$H_0: \rho \neq 0$ (există autocorelarea erorilor)

Equation: REGRESIE1 Workfile: UNTITLED::Untitled\

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

Dependent Variable: EMIGRANTI_TEMPORAR
Method: Least Squares
Date: 05/24/19 Time: 17:55
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7610.079	1757.452	-4.330178	0.0001
CASTIGUL_SALARIAL	7.041274	0.973646	7.231864	0.0000

R-squared	0.566630	Mean dependent var	4942.333
Adjusted R-squared	0.555796	S.D. dependent var	2679.773
S.E. of regression	1786.033	Akaike info criterion	17.85983
Sum squared resid	1.28E+08	Schwarz criterion	17.94258
Log likelihood	-373.0564	Hannan-Quinn criter.	17.89016
F-statistic	52.29986	Durbin-Watson stat	2.209005
Prob(F-statistic)	0.000000		

$$DW_{calculat} = 2.209005 \cong 2.21$$

Regiunea în care se încadrează $DW_{calculat} = 2.209$ este Regiunea 3, deoarece

$d_2 < DW_{calculat} < 4 - d_2$ (1.350 < 2.21 < 2.650), rezultă că nu respingem ipoteza nulă

$H_0: \rho = 0$, nu există autocorelare a erorilor aleatoare.

Testul BREUSCH-GODFREY

Prin acest test se verifică dacă erorile aleatoare admit o reprezentare autoregresivă de ordinul r, AR(r), adică:

$$\varepsilon_t = \rho_1 \cdot \varepsilon_{t-1} + \rho_2 \cdot \varepsilon_{t-2} \ldots\ldots + \rho_r \cdot \varepsilon_{t-r} + \upsilon_t,$$

unde υ_t este zgomot alb, adică un şir de v.a. i.i.d., necorelate, cu media şi dispersia constantă.

Ipotezele testate:

$H_0: \rho_1 = \rho_2 = \ldots = \rho_3 = 0$ (erorile nu sunt autocorelate sau nu există autocorelare de ordin r a erorilor aleatoare).

$H_1: \rho_s \neq 0, (\forall) s \ni \{1,2, \ldots, r\}$ (există autocorelarea de ordin r a erorilor aleatoare).

Equation: REGRESIE1 Workfile: UNTITLED::Untitled\

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

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.943727	Prob. F(2,38)	0.1571
Obs*R-squared	3.897899	Prob. Chi-Square(2)	0.1424

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 05/25/19 Time: 01:03
Sample: 1 42
Included observations: 42
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	347.9851	1777.524	0.195770	0.8458
CASTIGUL_SALARIAL	-0.186773	0.984722	-0.189671	0.8506
RESID(-1)	-0.189344	0.163020	-1.161480	0.2527
RESID(-2)	-0.272813	0.158125	-1.725304	0.0926

R-squared	0.092807	Mean dependent var	-1.19E-12
Adjusted R-squared	0.021187	S.D. dependent var	1764.118
S.E. of regression	1745.330	Akaike info criterion	17.85767
Sum squared resid	1.16E+08	Schwarz criterion	18.02316
Log likelihood	-371.0110	Hannan-Quinn criter.	17.91833
F-statistic	1.295818	Durbin-Watson stat	1.805991
Prob(F-statistic)	0.289876		

Estimarea modelului auxiliar de regresie:

$$\hat{\varepsilon}_t = \alpha_0 + \alpha_1 \cdot x_t + \rho_1 \cdot \hat{\varepsilon}_{t-1} + \rho_2 \cdot \hat{\varepsilon}_{t-2} + \nu_t \Rightarrow$$

$$\hat{\varepsilon}_t = \hat{\alpha}_0 + \hat{\alpha}_1 \cdot x_t + \hat{\rho}_1 \cdot \hat{\varepsilon}_{t-1} + \hat{\rho}_2 \cdot \hat{\varepsilon}_{t-2}$$

$$\hat{\varepsilon}_t = 347.9851 - 0.186773 \cdot x_t - 0.189344 \cdot \hat{\varepsilon}_{t-1} - 0.272813 \cdot \hat{\varepsilon}_{t-2}$$

$$LM_{calculat} = nR_a{}^2 = Obs * R - squared = 0.092807$$

Valoarea critică: $\chi^2{}_{critic(0.05,2)} = 5.9914$

Decizie:

Daca $LM_{calculat} = 0.092807 < \chi^2{}_{critic(0.05,2)} = 5.9914$, atunci acceptăm ipoteza nulă $H_0: \rho_1 = \rho_2 = \ldots = \rho_3 = 0$, adică erorile nu sunt autocorelate sau nu există autocorelare de ordin r a erorilor aleatoare.

Daca $Prob.Chi - Square(2) = 0.1424 > \alpha = 0.05$, atunci acceptăm ipoteza nulă $H_0: \rho_1 = \rho_2 = \ldots = \rho_3 = 0$, adică erorile nu sunt autocorelate sau nu există autocorelare de ordin r a erorilor aleatoare.

În concluzie erorile aleatoare nu sunt autocorelate.

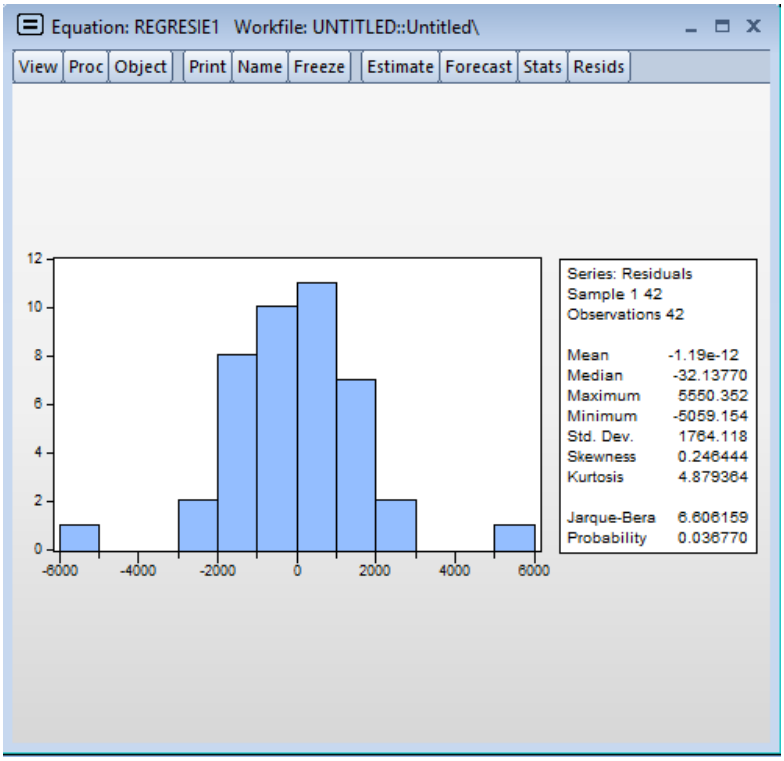
Normalitatea erorilor aleatoare

Testul Jarque-Bera (JB) privind distribuţia normală a erorilor

Ipotezele statistice:

H_0: Reziduurile provin dintr-o distribuţie normală (S=0 și K=3), unde S=Skewness, K=Kurtosis

H_1: Reziduurile nu provin dintr-o distribuţie normală



Jarque-Bera=6.606159

Probability=0.036770

Deoarece Probabilitatea asociată statisticii JB este < 0,05, adică Probability = 0.036770 < 0.05, respingem ipoteza nulă și acceptăm alternativa, reziduurile nu provin dintr-o distribuţie normal.

Previziune

Daca valoarea variabilei independente(X), câștigului salarial nominal net, crește cu 10%, atunci care este valoarea variabilei dependente(Y), emigrant temporar, pentru fiecare județ al României?

Modelul estimate de regresie: $\boldsymbol{\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i + \hat{\varepsilon}_i}$

$$\hat{y}_i = -7610.0793 \ + 7.041274 \cdot x_i$$

$$\hat{y}_0 = -7610.0793 \ + 7.041274 \cdot x_0$$

$$\hat{y}_0 = -7610.0793 \ + 7.041274 \cdot 0.1$$

$\hat{y}_0$ =-7610.2958726 nr. persoane

Raspunsul mediu al variabilei $\hat{y}_0$ =-7610.2958726, atunci când variabila $x_0 = 0.1$.

$$CI_{100(1-\alpha)\%} = \left[\hat{y}_0 \pm t_{critic} \cdot \hat{\sigma}_\varepsilon \cdot \sqrt{\frac{1}{n} + \frac{x_0 - \bar{x}}{\sum_1^{42}(x_i - \bar{x})^2}}\right]$$

$$CI_{95\%} = \left[-7610.29 \pm 2.02 \cdot 1786.033 \cdot \sqrt{\frac{1}{42} + \frac{0.1 - 4942.33}{\sum_1^{42}(0.1 - 4942.33)^2}}\right]$$

Intervalul de incredere la un nivel de semnificație $\alpha = 0.05$, atunci când variabila independentă crește cu un procent de 10%(0.1 lei).

Modelul econometric multifactorial

Y- variabila care arată valoarea emigrațiilor temporari pentru fiecare județ, exprimată în număr de persoane (variabila dependentă sau variabila explicată sau variabila endogenă sau variabila răspuns).

X_1- variabila care arată valoarea câștigului salarial nominal net pentru fiecare județ, exprimată în Lei (variabila independentă sau variabila explicativă sau variabila exogena sau variabila factor).

X_2- variabila care arată numarul șomerilor înregistrați pentru fiecare județ, exprimată în numar de persoane (variabila independentă sau variabila explicativă sau variabila exogena sau variabila factor).

Definirea modelului econometric

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \varepsilon \Leftrightarrow$$

$$y_i = \beta_0 + \beta_1 \cdot x_{i,1} + \beta_2 \cdot x_{i,2} + \varepsilon_i \Leftrightarrow$$

$$\text{Emigranți}_i = \beta_0 + \beta_1 \cdot \text{Câștigul}S_{i,1} + \beta_2 \cdot \text{Șomeri}_{i,2} + \varepsilon_i$$

Pentru fiecare dintre cele n=42 județe selectate s-au înregistrat valorile celor trei variabile, X_1 , X_2 și Y, cu k=2 variabile explicative .

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.859092107
R Square	0.738039248
Adjusted R Square	0.724605364
Standard Error	1406.292954
Observations	42

ANOVA					
	df	SS	MS	F	Significance F
Regression	2	217299856.3	108649928.1	54.93863208	4.52487E-12
Residual	39	77128735.05	1977659.873		
Total	41	294428591.3			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-9119.547534	1415.682547	-6.441802616	1.26216E-07	-11983.03577	-6256.059301
castigul	6.651231256	0.770511046	8.632233492	1.38621E-10	5.09272556	8.209736952
someri	0.221408815	0.043829191	5.051629032	1.06399E-05	0.132755908	0.310061722

RESIDUAL OUTPUT		
Observati	Predicted	Residuals
1	3081.916	2830.084
2	2511.339	412.6609
3	8220.032	-1023.03
4	2755.197	2288.803
5	3264.332	367.6682
6	2750.613	-489.613
7	4529.766	-951.766
8	6376.37	-676.37
9	2132.327	81.6735
10	2717.082	595.9176
11	5623.854	199.1461
12	5953.934	-1860.93
13	5670.656	586.344
14	3099.479	1010.521
15	6685.182	1080.818
16	3829.5	1130.5
17	4675.885	1738.115
18	5419.681	-1418.68
19	3342.065	-87.065
20	5610.909	-972.909
21	5632.207	1820.793
22	6635.786	-1131.79
23	3227.557	-966.557
24	2999.489	547.511
25	7357.474	-1189.47
26	3323.134	-120.134
27	5440.364	-194.364
28	3411.93	-494.93
29	3046.412	-101.412
30	6666.231	1232.769
31	5004.881	-1043.88
32	7144.787	-3055.79
33	14457.71	4078.285
34	8739.923	-1712.92
35	5500.167	-1969.17
36	3943.575	-1173.58
37	5838.388	-1284.39
38	3334.22	443.7802
39	4072.137	419.8626
40	2302.778	881.2225
41	3911.202	337.7978
42	7337.524	-165.524

Output Excel Model Regresie Multifactorială

Ecuatia de regresie estimată:

$$\hat{y}_i = -9119.547534 + 6.651231256 \cdot x_{i,1} + 0.221408815 \cdot x_{i,2}$$

R = 0.8590 (o valoare apropiată de 1), arată că între numarul emigranților temporari și cele doua variabile luate în studiu (câștigul slarial nominal net și șomerii înregistrați) există o legatură puternică, legatură pusă în evidență prin modelul de regresie liniară multifactorială.

R^2 =0.7380, arată că la nivelul eșantionului, 73,80% din variația totală a numarului de emigranți temporari este explicată de influența celor două variabile independente, iar restul de 26,20% din variația totală a numarului de emigranți este dată de factorii reziduali.

$\widehat{\sigma}_{\varepsilon} =$ 1406.2929, abaterea standard de selecție a variabilei reziduale sau abaterea medie patratică a erorilor.

Intercept este termenul liber, deci coeficientul estimat $\hat{\beta}_0$ este -9119.547534 și arată, pe baza datelor din eșantion, nivelul mediu al variabilei dependente atunci când nivelul tututor variabilelor explicative este egal cu 0 unități. Deoarece valoarea calculată a statisticii testului t pentru testarea ipotezei $H_0{:}\beta_0 = 0$ contra $H_1{:}\beta_0 \neq 0$ este $t_{calculat} = -6.4418$, iar pragul de semnificație calculat al testului, P-*value*, este 1.26216E - 07 < 0.05 = α, înseamnă că parametrul β_0 este semnificativ.

Coeficientul $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ este 6.651231256 ceea ce înseamnă că la o creștere a nivelului câștigului salarial nominal net cu o unitate , menținând nemodificat nivelul celeilalte varabile explicative din model, numarul de emigranți temporari va crește în medie cu 6.651231256 (≅ 7 persoane). Deoarece valoarea calculată a statisticii testului t pentru testarea ipotezei $H_0{:}\beta_1 = 0$ contra $H_1{:}\beta_1 \neq 0$ este $t_{calculat} = 8.6322$,iar pragul de semnificație calculat al testului, P-*value*, este 1.38621E - 10 < 0.05 = α, înseamnă că avem suficiente motive să respingem ipoteza nulă, deci β_1 este semnificativ statistic.

Coeficientul $\widehat{\boldsymbol{\beta}}_{\mathbf{2}}$ este 0.221408815 ceea ce înseamnă că la o creștere a numarului de șomeri înregistrați cu o unitate , menținând nemodificat nivelul celeilalte varabile explicative din model, numarul de emigranți temporari va crește în medie cu 0.221408815 (≅ 1 $persoană$). Deoarece valoarea calculată a statisticii testului t pentru testarea ipotezei $H_0{:}\beta_2 = 0$ contra $H_1{:}\beta_2 \neq 0$ este $t_{calculat}$ =5.051629032, iar pragul de semnificație calculat al testului, P-*value*, este 1.06399E – 05 < 0.05 = α, înseamnă că avem suficiente motive să respingem ipoteza nulă, deci β_2 este semnificativ statistic.

Testarea validitatii modelului de regresie multifactoriala

Ipotezele statistice:

$H_0 : \beta_1 = \beta_2 = 0$ (modelul de regresie nu este valid statistic, adică toți coeficienții variabilelor explicative nu sunt semnificativ diferiți de 0).

$H_1 : \beta_k \neq 0\ (\forall) k \ni \{1,2\}$ (modelul de regresie este valid sau semnificativ statistic; cel puțin un coefficient diferă semnificativ de 0).

Întrucât $F_{calculat} = 54.938$, iar *Significace F* = 4.52487E – 12 < 0.05 = α, atunci respingem H_0 la un nivel de semnificație de 5% și concluzionăm că datele sunt în favoarea ipotezei alternative, adică modelul de regresie construit este valid statistic.

Multicoliniaritatea variabilelor explicative, Normalitatea erorilor

Equation: REGRESIE Workfile: BAZA_DE_DATE11111::Baza_de_date1...

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

Dependent Variable: EMIGRANTI
Method: Least Squares
Date: 05/25/19 Time: 14:31
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9119.548	1415.683	-6.441803	0.0000
CASTIGUL	6.651231	0.770511	8.632233	0.0000
SOMERI	0.221409	0.043829	5.051629	0.0000

R-squared	0.738039	Mean dependent var	4942.333
Adjusted R-squared	0.724605	S.D. dependent var	2679.773
S.E. of regression	1406.293	Akaike info criterion	17.40405
Sum squared resid	77128735	Schwarz criterion	17.52817
Log likelihood	-362.4851	Hannan-Quinn criter.	17.44955
F-statistic	54.93863	Durbin-Watson stat	2.217786
Prob(F-statistic)	0.000000		

Detectarea multicoliniarităţii pe baza coeficienţilor de corelaţie dintre variabilele explicative:

Table: UNTITLED Workfile: BAZA_DE_DATE11111::Baza_d...

View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comment

Correlation

	A	B	C	D
1		EMIGRANTI	CASTIGUL	SOMERI
2				
3	EMIGRANTI	1.000000	0.752748	0.487364
4	CASTIGUL	0.752748	1.000000	0.100208
5	SOMERI	0.487364	0.100208	1.000000
6				
7				

Între variabilele X_1 și X_2 există o legătură liniară directă, deoarece coeficientul de corelație liniară Pearson este pozitiv,

$r_{x_1,x_2} = 0.100208$ variabilele X_1 și X_2 sunt corelate.

Model de regresie1: CâștigulS C Șomeri

Equation: EQ Workfile: BAZA_DE_DATE11111::Baza_de_date11111\

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

Dependent Variable: CASTIGUL
Method: Least Squares
Date: 05/25/19 Time: 16:10
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1725.928	99.61815	17.32544	0.0000
SOMERI	0.005700	0.008949	0.636977	0.5278

R-squared	0.010042	Mean dependent var	1782.690
Adjusted R-squared	-0.014707	S.D. dependent var	286.4815
S.E. of regression	288.5805	Akaike info criterion	14.21427
Sum squared resid	3331148.	Schwarz criterion	14.29702
Log likelihood	-296.4997	Hannan-Quinn criter.	14.24460
F-statistic	0.405740	Durbin-Watson stat	1.757001
Prob(F-statistic)	0.527769		

Model de regresie2: Someri C CâștigulS

Equation: EQ1 Workfile: BAZA_DE_DATE11111::Baza_de_date11111\

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

Dependent Variable: SOMERI
Method: Least Squares
Date: 05/25/19 Time: 16:19
Sample: 1 42
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6817.561	4992.020	1.365692	0.1797
CASTIGUL	1.761642	2.765629	0.636977	0.5278

R-squared	0.010042	Mean dependent var	9958.024
Adjusted R-squared	-0.014707	S.D. dependent var	5036.304
S.E. of regression	5073.204	Akaike info criterion	19.94778
Sum squared resid	1.03E+09	Schwarz criterion	20.03053
Log likelihood	-416.9034	Hannan-Quinn criter.	19.97811
F-statistic	0.405740	Durbin-Watson stat	1.876218
Prob(F-statistic)	0.527769		

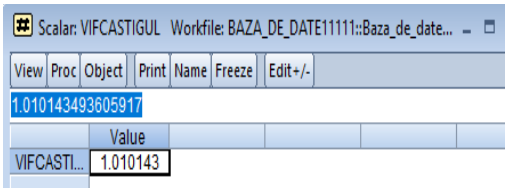
Table: UNTITLED Workfile: BAZA_DE_DA...

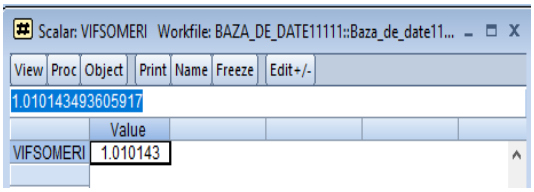
View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/-

Correlation

	A	B	C
1		CASTIGUL	SOMERI
2			
3	CASTIGUL	1.000000	0.100208
4	SOMERI	0.100208	1.000000
5			
6			

Criteriul factorului de inflaţie a varianţei (Variance Inflationary Factor):





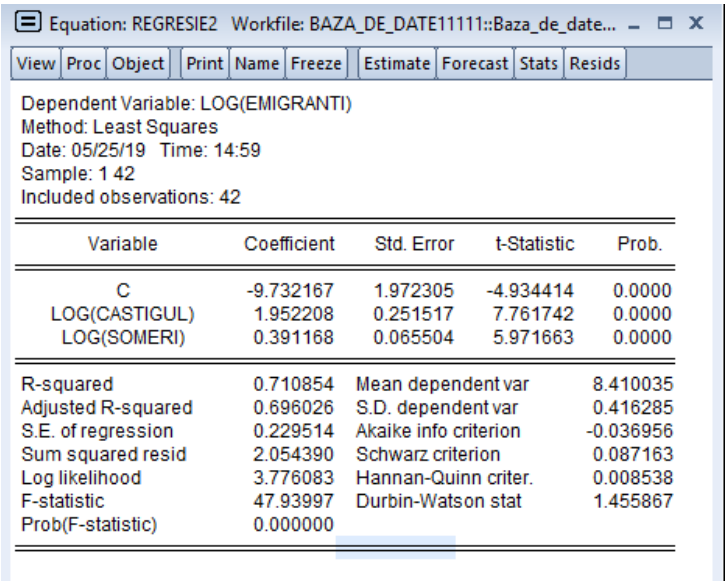
Variabilele X_1 și X_2 includ multicoliniaritate de nivel redus conform criteriului de inflație a varianței.

Remedii pentru Multicoliniaritate: Transformarea variabilelor

Transformarea modelului de regresie:

$$y_i = \beta_0 + \beta_1 \cdot x_{i,1} + \beta_2 \cdot x_{i,2} + \varepsilon_i \Leftrightarrow$$

$$ln(yi) = \beta_0 + \beta_1 \cdot ln(x_{i,1}) + \beta_2 \cdot ln(x_{i,2}) + \varepsilon_i$$



Întrucât F-statistic = 47.93997 și Prob(F-statistic) = 0.0000 < 0.05 = α, la un nivel de semnificație de 5%, modelul statistic este valid.

Pentru coeficientul pantă β_1 se observă valoarea statistică a probabilității Prob.(p-*value*) = 0.0000 < 0.05 = α, deci putem afirma faptul că β_1 este semnificativ statistic.

Pentru coeficientul panta $\beta_2$se observă valoarea statistică a probabilității Prob.(p-*value*) = 0.0000 < 0.05 = α, deci putem afirma faptul că $\beta_2$este semnificativ statistic.

Table: UNTITLED Workfile: BAZA_DE_DATE11111::Baza_d...

View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comment

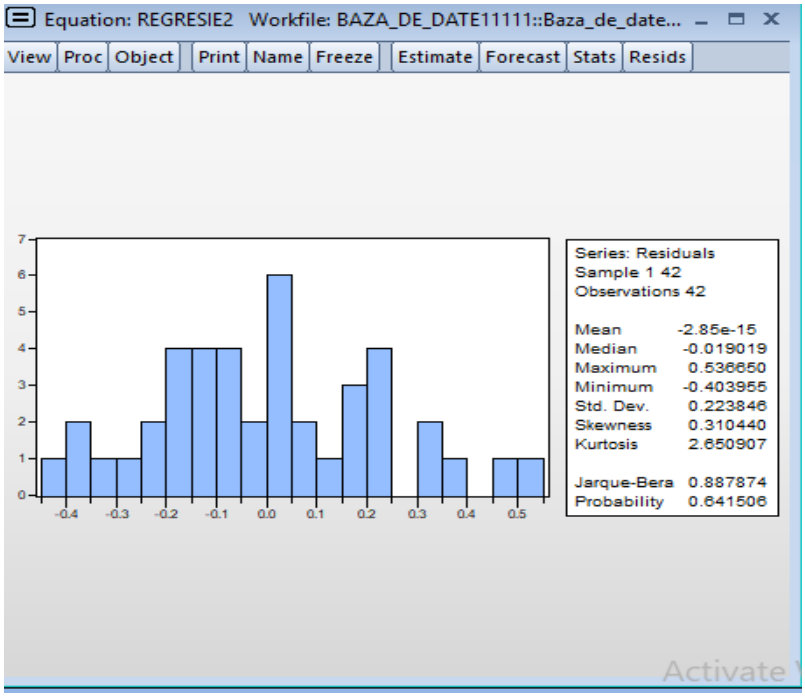
Correlation

	A	B	C	D
1		LOG(EMIGR...	LOG(CASTI...	LOG(SOMERI)
2				
3	LOG(EMIGR...	1.000000	0.668180	0.514004
4	LOG(CASTI...	0.668180	1.000000	-0.000275
5	LOG(SOMERI)	0.514004	-0.000275	1.000000
6				
7				

Între variabilele X_1 și X_2 nu există o legătură liniară directă, deoarece coeficientul de corelație liniară Pearson este negativ,

$r_{x_1,x_2} = -0.000275$ variabilele X_1 șiX_2 sunt negativ correlate .

Testul Jarque-Bera (JB) privind distribuţia normală a erorilor



Jarque-Bera = 0.887874

Deoarece Probabilitatea asociată statisticii JB este > 0,05, adică Probability = 0,64 > 0,05, putem afirma faptul că reziduurile provin dintr-o distribuţie normală.

Concluzii

În urma analizei modelului de regresie, atât unifactorial cât și multifactorial putem concluziona faptul că în România, pentru fiecare județ, în anul 2016, atât câștigul salarial nominal net cât și numarul șomerilor înregistrați pentru fiecare județ are influență asupra numarului emigranților temporari.

Din modelul de regresie multifactorial reiese că un procent de 73,80% din variația totală a numarului emigranților este explicat de variabilele independente ale modelului, câștigul salarial nominal net și numărul șomerilor înregistrați, restul de 26,20% este explicat de alți factori reziduali.

În România emigrațiile sunt într-un procent mai mare decat imigrațiile, tinerii fiind principalii emigranți.

Bibliografie

1. http://statistici.insse.ro:8077/tempo-online
2. http://www.insse.ro/cms/
3. https://ec.europa.eu/eurostat/statistics-explained
4. http://revistaș inteza.ro/