



Academia de Studii Economice
București
Facultatea de Cibernetică, Statistică și
Informatică Economică



Analiza evoluției speranței de viață în perioada 1910 – 2018 în România

Proiect Serii de Timp

Profesor coordonator: Daniel Traian Pele

Realizat de
Berca Sofia Maira
Bișag Alexandru Ștefan
Bucăloiu Florian Alexandru
grupa 1066, seria C, Informatică Economică

- 1. Contextul și datele selectate**
- 2. Seria de timp – reprezentare și analiză**
- 3. Staționaritate – analiză și transformări**
- 4. Modelul ARIMA**
- 5. Predicția**
- 6. Bibliografie**

1. Contextul si datele selectate

Așa cum poate fi observat încă din titlu, proiectul analizează, folosind concepte statistice specifice, evoluția speranței de viață a persoanelor ce trăiesc în România în perioada modernă și contemporană, respectiv între anii 1910 și 2018.

Inițial, proiectul urmărea să acopere o perioadă mai restrânsă, anume perioada cuprinsă în intervalul 1945– 2018, însă, în urma observării datelor și a plasării lor în contextul istoric pe care îl descriu, am considerat că extinderea perioadei pentru a include două dintre cele mai importante evenimente istorice, cele două războaie mondiale, ar reprezenta un exercițiu interesant și ar permite înțelegerea mai exactă a fenomenului.

Totodată, am dorit să alegem pentru observare un indicator care să descrie, prin evoluția sa, evoluția economică și socială a întregii regiuni pentru care este calculat, speranța de viață fiind un etalon al dezvoltării socio-economice datorită multitudinii de factori ce îi influențează variația.

Software-ul utilizat pentru calcularea și generarea tuturor reprezentărilor ce vor fi expuse în continuare este Eviews 10 – Student Edition.

Figura 1- Valorile speranței de viață între anii 1910 – 2018 în România
Sursa date: Gapminder/ Life Expectancy
<https://www.gapminder.org/data/>

An	Speranța de viață
1910	35.8
1911	35.8
1912	35.8
1913	35.8
1914	35.8
1915	35.8
1916	35.8
1917	35.8
1918	26.5
1919	35.8
1920	35.8
1921	36.3
1922	36.8
1923	37.4
1924	37.9
1925	38.4
1926	38.9
1927	39.5
1928	40
1929	40.5
1930	41
1931	41.6
1932	42.1
1933	43.2
1934	44.2
1935	45.3
1936	46.4
1937	47.4
1938	48.5
1939	49.6
1940	50.6
1941	45.5
1942	44.8
1943	42.5
1944	41.3
1945	45.5
1946	54.1
1947	58.1
1948	59.2
1949	60.2
1950	61.3
1951	61.2
1952	61.1
1953	61.3
1954	61.5
1955	62
1956	62.6
1957	63.3
1958	64.2
1959	65.1
1960	66
1961	66.7
1962	67.3
1963	67.8
1964	68
1965	68
1966	68
1967	67.9
1968	68
1969	68.1
1970	68.4
1971	69
1972	69.2
1973	69.4
1974	69.9
1975	70.1
1976	70.4
1977	69.8
1978	69.9
1979	69.5
1980	69.7
1981	69.8
1982	70
1983	70.2
1984	70.1
1985	69.8
1986	70
1987	69.4
1988	69.8
1989	69.9
1990	69.9
1991	70.4
1992	69.9
1993	69.7
1994	69.5
1995	69.4
1996	69.2
1997	69.4
1998	70.1
1999	70.9
2000	71.2
2001	71.1
2002	71.3
2003	71.6
2004	72.1
2005	72.4
2006	72.8
2007	73.2
2008	73.2
2009	73.3
2010	73.7
2011	74.5
2012	74.6
2013	74.8
2014	74.8
2015	75.1
2016	75.3
2017	75.4
2018	75.5

2. Seria de timp – reprezentare și analiză

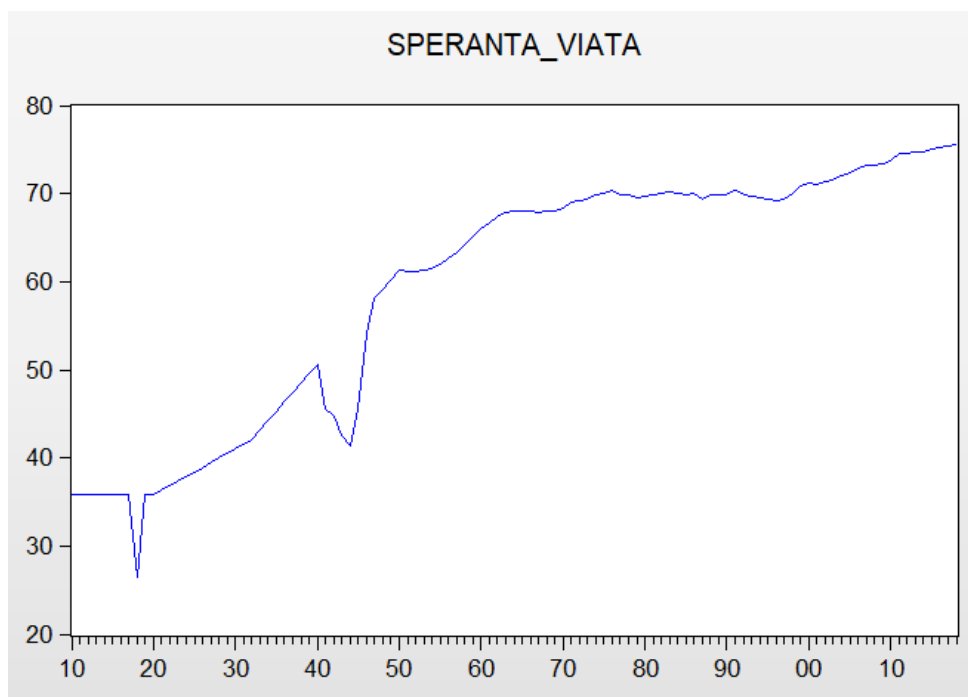


Figura 2. Speranța de viață - evoluție, 1910 – 2018, România

Reprezentarea grafică a seriei de timp selectate surprinde fenomenul dramatic de scădere a speranței de viață în perioada celor două războaie mondiale, între anii 1914 – 1918, respectiv 1939 – 1945, ce poate fi cu ușurință tradus prin numeroasele pierderi suferite, mai ales în rândul populației tinere, masculine. Totuși, diferența nivelului de trai dintre cele două perioade nefaste poate fi ușor observată, cel de-al Doilea Război Mondial având un impact mai puternic ca urmare a trend-ului pozitiv, puternic surprins în cadrul perioadei interbelice.

Din punct de vedere statistic, observăm că seria analizată prezintă un trend general pozitiv, fără sezonabilitate, iar, la prima vedere, bazându-ne pe forma graficului, putem afirma că aceasta este nestaționară atât în medie cât și în dispersie.

3. Staționaritate – analiză și transformări

Pentru a confirma sau a respinge afirmația precedentă – nestaționaritatea seriei analizate – vom utiliza două metode ale căror rezultate vor determina modul de acțiune ulterior.

Date: 05/26/19 Time: 16:23

Sample: 1910 2018

Included observations: 109

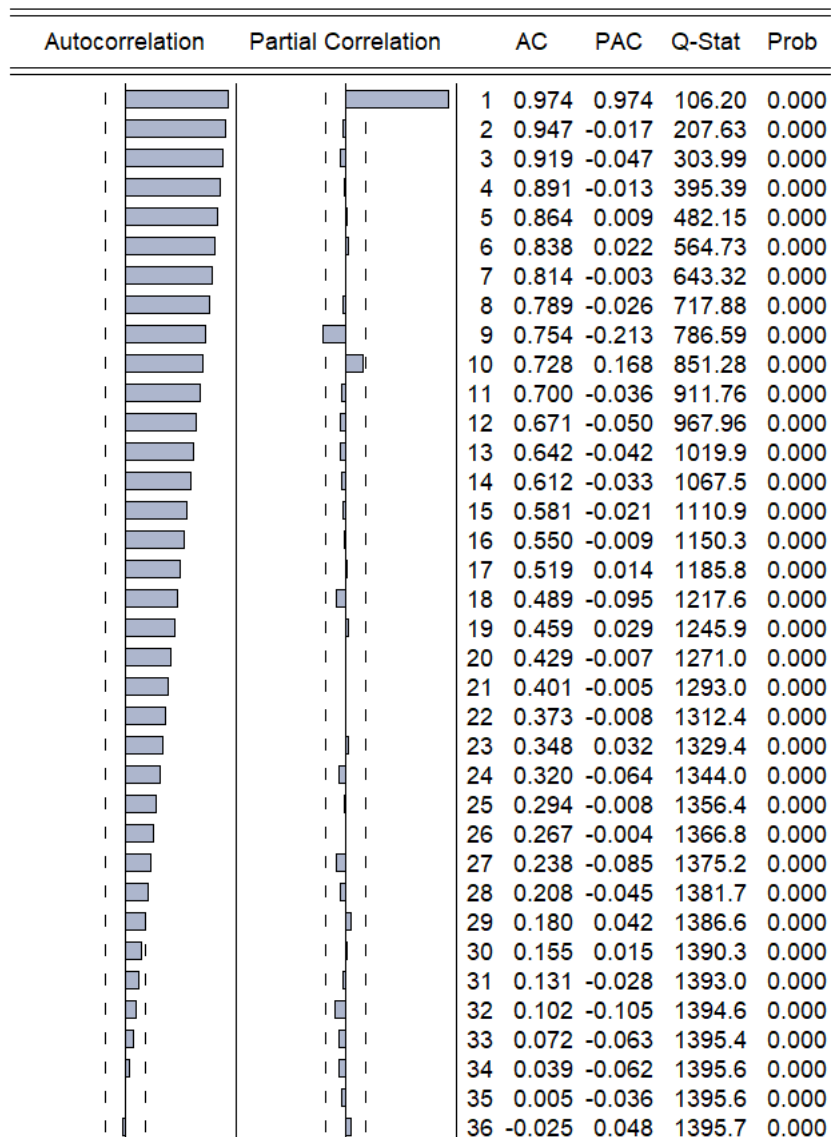


Figura 3. Corelograma seriei de date

Observând corelograma generată, se observă că lag-urile prezintă valori foarte mari, iar că tendința de scădere manifestată de acestea este foarte lentă, ajungând în proximitatea lui 0 abia în jurul lag-ului 33, fapt ce întărește concluzia anterioară, anume nestaționaritatea seriei analizate.

Totuși, pentru siguranța afirmației, am aplicat, în continuare, testul Dickey-Fuller, în toate cele 3 forme disponibile.

Augmented Dickey-Fuller Unit Root Test on SPERANTA_VIATA

Null Hypothesis: SPERANTA_VIATA has a unit root
Exogenous: None
Lag Length: 1 (Automatic - based on SIC, maxlag=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	1.820263	0.9832
Test critical values:		
1% level	-2.586753	
5% level	-1.943853	
10% level	-1.614749	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SPERANTA_VIATA)
Method: Least Squares
Date: 05/26/19 Time: 16:24
Sample (adjusted): 1912 2018
Included observations: 107 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SPERANTA_VIATA(-1)	0.005217	0.002866	1.820263	0.0716
D(SPERANTA_VIATA(-1))	-0.017731	0.098125	-0.180700	0.8570
R-squared	-0.012910	Mean dependent var		0.371028
Adjusted R-squared	-0.022557	S.D. dependent var		1.752680
S.E. of regression	1.772338	Akaike info criterion		4.000991
Sum squared resid	329.8240	Schwarz criterion		4.050950
Log likelihood	-212.0530	Hannan-Quinn criter.		4.021244
Durbin-Watson stat	1.996960			

Figura 4. Testul Dickey-Fuller pentru seria inițială (without Intercept, Trend)

Augmented Dickey-Fuller Unit Root Test on SPERANTA_VIATA

Null Hypothesis: SPERANTA_VIATA has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.251799	0.6497
Test critical values:		
1% level	-3.492523	
5% level	-2.888669	
10% level	-2.581313	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SPERANTA_VIATA)
Method: Least Squares
Date: 05/26/19 Time: 16:24
Sample (adjusted): 1912 2018
Included observations: 107 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SPERANTA_VIATA(-1)	-0.015124	0.012082	-1.251799	0.2135
D(SPERANTA_VIATA(-1))	-0.024770	0.097288	-0.254605	0.7995
C	1.278850	0.738308	1.732136	0.0862
R-squared	0.015492	Mean dependent var		0.371028
Adjusted R-squared	-0.003441	S.D. dependent var		1.752680
S.E. of regression	1.755693	Akaike info criterion		3.991242
Sum squared resid	320.5757	Schwarz criterion		4.066181
Log likelihood	-210.5314	Hannan-Quinn criter.		4.021621
F-statistic	0.818249	Durbin-Watson stat		1.997888
Prob(F-statistic)	0.444023			

Figura 5. Testul Dickey-Fuller pentru seria inițială (Intercept)

Augmented Dickey-Fuller Unit Root Test on SPERANTA_VIATA		
Null Hypothesis: SPERANTA_VIATA has a unit root		
Exogenous: Constant, Linear Trend		
Lag Length: 1 (Automatic - based on SIC, maxlag=12)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.454487	0.8389
Test critical values:		
1% level	-4.046072	
5% level	-3.452358	
10% level	-3.151673	
*MacKinnon (1996) one-sided p-values.		

Figura 6. Testul Dickey-Fuller pentru seria inițială (Intercept, Trend)

Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(SPERANTA_VIATA)				
Method: Least Squares				
Date: 05/26/19 Time: 16:24				
Sample (adjusted): 1912 2018				
Included observations: 107 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SPERANTA_VIATA(-1)	-0.048050	0.033036	-1.454487	0.1489
D(SPERANTA_VIATA(-1))	-0.003601	0.099210	-0.036294	0.9711
C	2.340599	1.235953	1.893761	0.0611
@TREND("1910")	0.016126	0.015060	1.070756	0.2868
R-squared	0.026330	Mean dependent var		0.371028
Adjusted R-squared	-0.002029	S.D. dependent var		1.752680
S.E. of regression	1.754458	Akaike info criterion		3.998863
Sum squared resid	317.0466	Schwarz criterion		4.098782
Log likelihood	-209.9392	Hannan-Quinn criter.		4.039369
F-statistic	0.928441	Durbin-Watson stat		2.000160
Prob(F-statistic)	0.429863			

Așa cum poate fi observat în urma aplicării testului Dickey-Fuller în toate cele trei forme, putem afirma că seria analizată este nestaționară.

Date fiind ipotezele H_0 – seria este nestaționară și H_1 – seria este staționară, putem observa în cele trei rezultate faptul că, pentru fiecare dintre cele trei nivele de semnificație, valoarea calculată a testului t este mai mică decât cea a valorii critice, iar prob este mai mare de 0.05. Astfel putem afirma că nu avem suficiente date pentru a accepta ipoteza H_1 și vom accepta ipoteza H_0 , seria este nestaționară.

Totuși, pentru a permite creionarea unui model ARIMA pe baza seriei noastre, trebuie să încercăm transformarea acesteia.

Primul pas îl reprezintă logaritmarea, ce are ca scop staționarizarea seriei din punct de vedere al dispersiei.

În programul EViews logaritmarea presupune generarea unei serii noi ce va fi exprimată prin formula $\log(\text{serie_inițială})$. Ulterior generării, vom aplica din nou metodele utilizate mai sus pentru a determina dacă noua serie obținută este staționară sau dacă trebuie să recurgem la o transformare suplimentară.

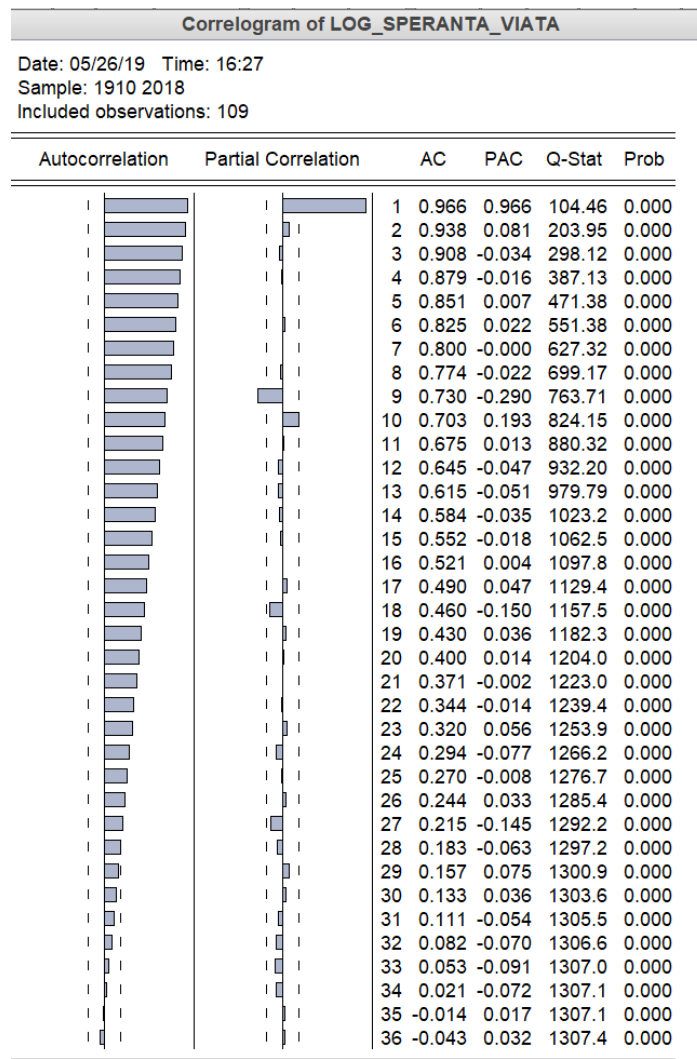


Figura 7. Corelograma
seriei logaritmata

Augmented Dickey-Fuller Unit Root Test on LOG_SPERANTA_VIATA				
Null Hypothesis: LOG_SPERANTA_VIATA has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 1 (Automatic - based on SIC, maxlag=12)				
		t-Statistic	Prob.*	
Augmented Dickey-Fuller test statistic		-1.515221	0.8185	
Test critical values:	1% level	-4.046072		
	5% level	-3.452358		
	10% level	-3.151673		
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(LOG_SPERANTA_VIATA)				
Method: Least Squares				
Date: 05/26/19 Time: 16:27				
Sample (adjusted): 1912 2018				
Included observations: 107 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG_SPERANTA_VIATA(-1)	-0.062070	0.040964	-1.515221	0.1328
D(LOG_SPERANTA_VIATA(-1))	-0.208990	0.097072	-2.152940	0.0337
C	0.240248	0.148269	1.620352	0.1082
@TREND("1910")	0.000358	0.000356	1.004627	0.3174
R-squared	0.083958	Mean dependent var	0.006974	
Adjusted R-squared	0.057277	S.D. dependent var	0.048084	
S.E. of regression	0.046687	Akaike info criterion	-3.254045	
Sum squared resid	0.224505	Schwarz criterion	-3.154126	
Log likelihood	178.0914	Hannan-Quinn criter.	-3.213539	
F-statistic	3.146761	Durbin-Watson stat	1.997204	
Prob(F-statistic)	0.028300			

Figura 8. Testul Dickey-Fuller aplicat seriei logaritmata (Intercept, Trend)

Utilizând aceleași criterii, putem observa că în cadrul corelogramei valorile nu prezintă o tendință de scădere accelerată către 0 și că, din punct de vedere al testului Dickey-Fuller, unde valorile respectă același model enunțat anterior, seria continuă să fie nestaționară.

Prin urmare, este necesară o transformare suplimentară, diferențierea de ordinul I a seriei logaritmice, ce urmărește producerea unei noi serii care să respecte principiile staționarității. Noua serie, generată în baza ecuațiilor $d(\text{serie_logaritmată})$ sau $dlog(\text{serie_initială})$, este supusă aceluiași teste.

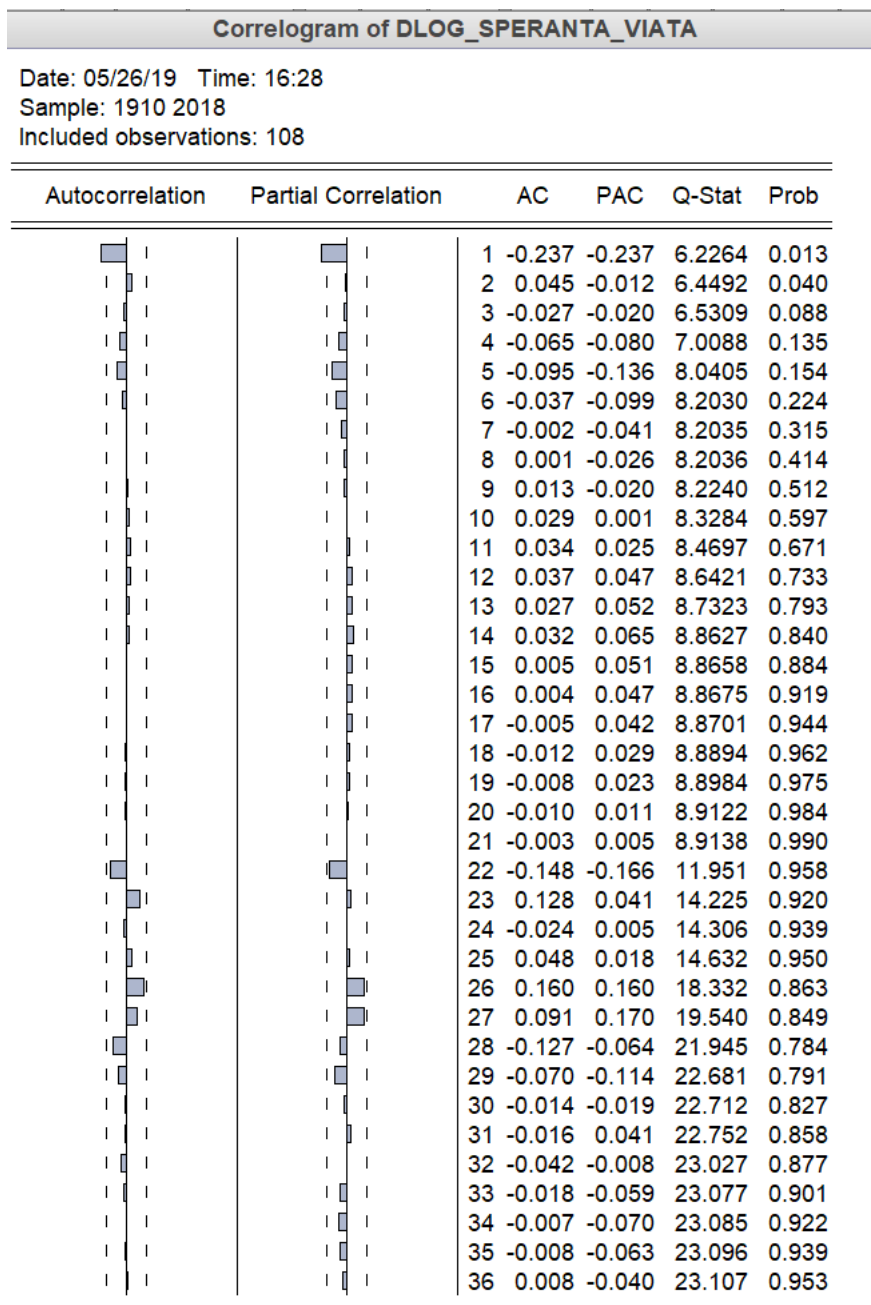


Figura 9.
Corelograma seriei
diferențiate

Figura 10.
Testul Dickey-
Fuller aplicat
seriei
diferențiate
(Intercept,
Trend)

Augmented Dickey-Fuller Unit Root Test on DLOG_SPERANTA_VIATA				
Null Hypothesis: DLOG_SPERANTA_VIATA has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 1 (Automatic - based on SIC, maxlag=12)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-8.134681	0.0000
Test critical values:	1% level		-4.046925	
	5% level		-3.452764	
	10% level		-3.151911	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DLOG_SPERANTA_VIATA)				
Method: Least Squares				
Date: 05/26/19 Time: 16:29				
Sample (adjusted): 1913 2018				
Included observations: 106 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG_SPERANTA_VIATA(-1)	-1.268596	0.155949	-8.134681	0.0000
D(DLOG_SPERANTA_VIATA(-...	0.020281	0.098906	0.205051	0.8379
C	0.016983	0.009745	1.742690	0.0844
@TREND("1910")	-0.000145	0.000151	-0.958991	0.3398
R-squared	0.622028	Mean dependent var		1.25E-05
Adjusted R-squared	0.610911	S.D. dependent var		0.075987
S.E. of regression	0.047399	Akaike info criterion		-3.223442
Sum squared resid	0.229156	Schwarz criterion		-3.122935
Log likelihood	174.8424	Hannan-Quinn criter.		-3.182706
F-statistic	55.95379	Durbin-Watson stat		2.002346
Prob(F-statistic)	0.000000			

În acest caz, primul semn al staționarității poate fi observat încă din corelogramă, valorile lag-urilor atingând un prag apropiat lui 0 încă din cea de-a doua instanță.

Testul Dickey-Fuller surprinde atât valoarea lui prob., 0.0000., ce este mai mică decât 0.05, precum și valoarea lui t-calculat ce este mai mare decât valorile critice aferente celor trei intervale de încredere, rezultând, în baza acestuia, că seria obținută în urma logaritmării și diferențierii este staționară.

Totuși, pentru a garanta acest rezultat, am ales să aplicăm testul KPSS, în baza căruia vom putea fi complet siguri de natura noii serii.

KPSS Unit Root Test on DLOG_SPERANTA_VIATA

Null Hypothesis: DLOG_SPERANTA_VIATA is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 10 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	0.091013
Asymptotic critical values*:	
1% level	0.216000
5% level	0.146000
10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.002259
HAC corrected variance (Bartlett kernel)	0.000754

KPSS Test Equation
Dependent Variable: DLOG_SPERANTA_VIATA
Method: Least Squares
Date: 05/26/19 Time: 16:29
Sample (adjusted): 1911 2018
Included observations: 108 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.012497	0.009298	1.344040	0.1818
@TREND("1910")	-0.000103	0.000148	-0.692372	0.4902
R-squared	0.004502	Mean dependent var		0.006909
Adjusted R-squared	-0.004889	S.D. dependent var		0.047864
S.E. of regression	0.047981	Akaike info criterion		-3.217699
Sum squared resid	0.244026	Schwarz criterion		-3.168030
Log likelihood	175.7557	Hannan-Quinn criter.		-3.197560
F-statistic	0.479380	Durbin-Watson stat		2.484492
Prob(F-statistic)	0.490217			

Figura 11. Testul
KPSS aplicat seriei
logaritmă și
diferențiate

Spre diferență de testul Dickey-Fuller, testul KPSS presupune o inversare a ipotezelor inițiale, H_0 , afirmând staționaritatea seriei, iar H_1 infirmând-o. În cazul curent, putem remarca faptul că valoarea calculată a lui t este mai mică decât valorile tuturor celor trei nivele de semnificație, fapt ce conduce la acceptarea ipotezei nule: seria analizată este staționară.

4. Modelul ARIMA

Obținerea unei serii staționare provenită din setul inițial de date ne permite ca, pe baza acesteia, să generăm un model ARIMA cu ajutorul căruia vom încerca, ulterior, să realizăm predicții.

Pentru a determina parametrii necesari construirii modelului am analizat corelograma seriei staționare și am ajuns la concluzia că modelul potrivit pentru datele noastre este ARIMA(1, 22).

Figura 12. Modelul ARIMA

Dependent Variable: DLOG(SPERANTA_VIATA)				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 05/26/19 Time: 16:40				
Sample: 1911 2018				
Included observations: 108				
Convergence achieved after 169 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007035	0.003659	1.922432	0.0573
AR(1)	-0.207555	0.031549	-6.578797	0.0000
MA(22)	-0.264831	0.097864	-2.706104	0.0080
SIGMASQ	0.002036	0.000114	17.91413	0.0000
R-squared	0.102926	Mean dependent var		0.006909
Adjusted R-squared	0.077049	S.D. dependent var		0.047864
S.E. of regression	0.045983	Akaike info criterion		-3.269547
Sum squared resid	0.219899	Schwarz criterion		-3.170208
Log likelihood	180.5555	Hannan-Quinn criter.		-3.229269
F-statistic	3.977499	Durbin-Watson stat		2.006033
Prob(F-statistic)	0.009977			
Inverted AR Roots	-.21			
Inverted MA Roots	.94	.90+.27i	.90-.27i	.79+.51i
	.79-.51i	.62+.71i	.62-.71i	.39+.86i
	.39-.86i	.13+.93i	.13-.93i	-.13-.93i
	-.13+.93i	-.39+.86i	-.39-.86i	-.62+.71i
	-.62-.71i	-.79+.51i	-.79-.51i	-.90+.27i
	-.90-.27i	-.94		

După cum putem observa prin observarea valorii prob, ce este mai mică de 0.05, parametrii sunt semnificativi. Pentru ca modelul să poată fi utilizat în generarea predicțiilor trebuie să ne asigurăm că reziduurile urmează o distribuție de tip zgomot alb.

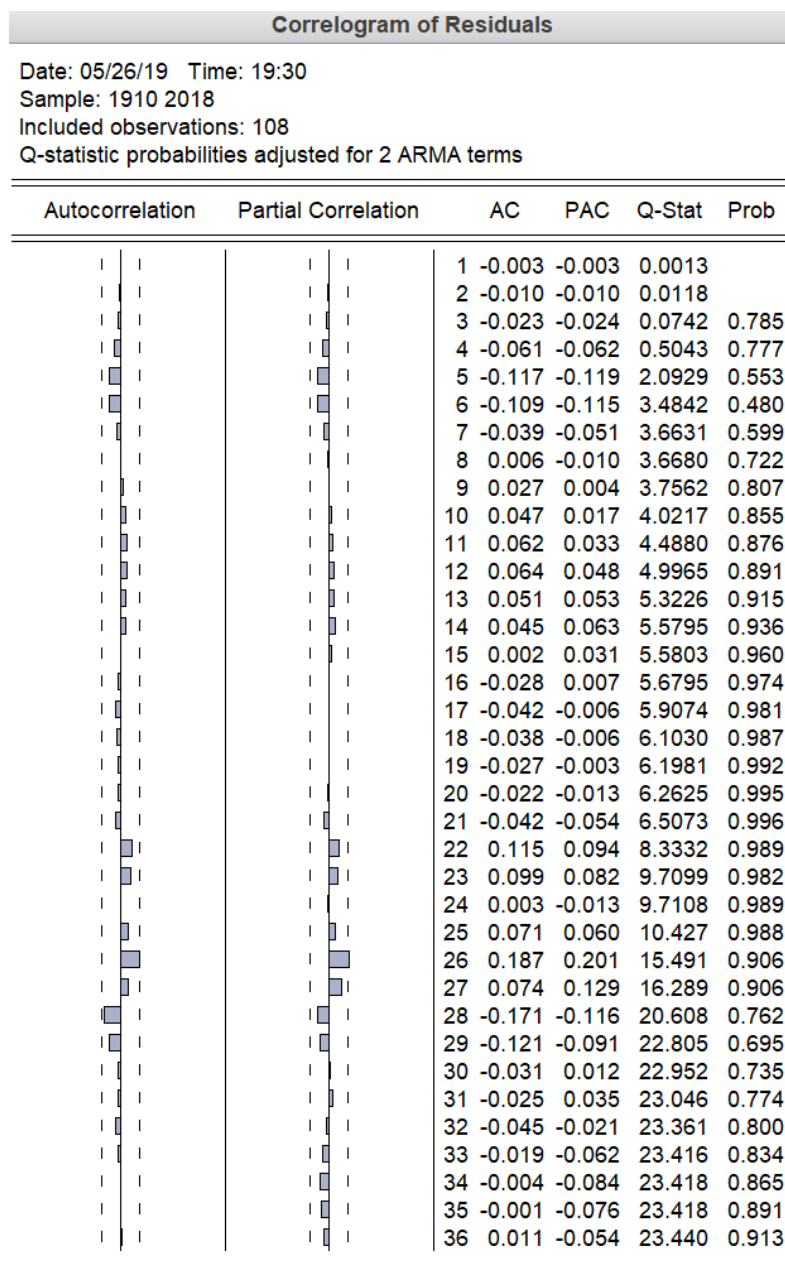


Figura 13. Corelograma reziduurilor

În ceea ce privește corelograma, putem observa că probabilitățile coeficienților sunt mai mari decât nivelul de semnificație și că, în cadrul reprezentării grafice, nu se depășește intervalul de încredere.

În baza acestor rezultate, putem afirma că reziduurile urmează o distribuție de tip zgomot alb, deci modelul poate fi utilizat pentru a realiza predicții, acesta reprezentând o aproximare acceptabilă pentru procesul pe care îl replică.

5. Predicție

Pe baza modelului la care ne-au condus pașii prezentați anterior, am generat următorul grafic de predicție.

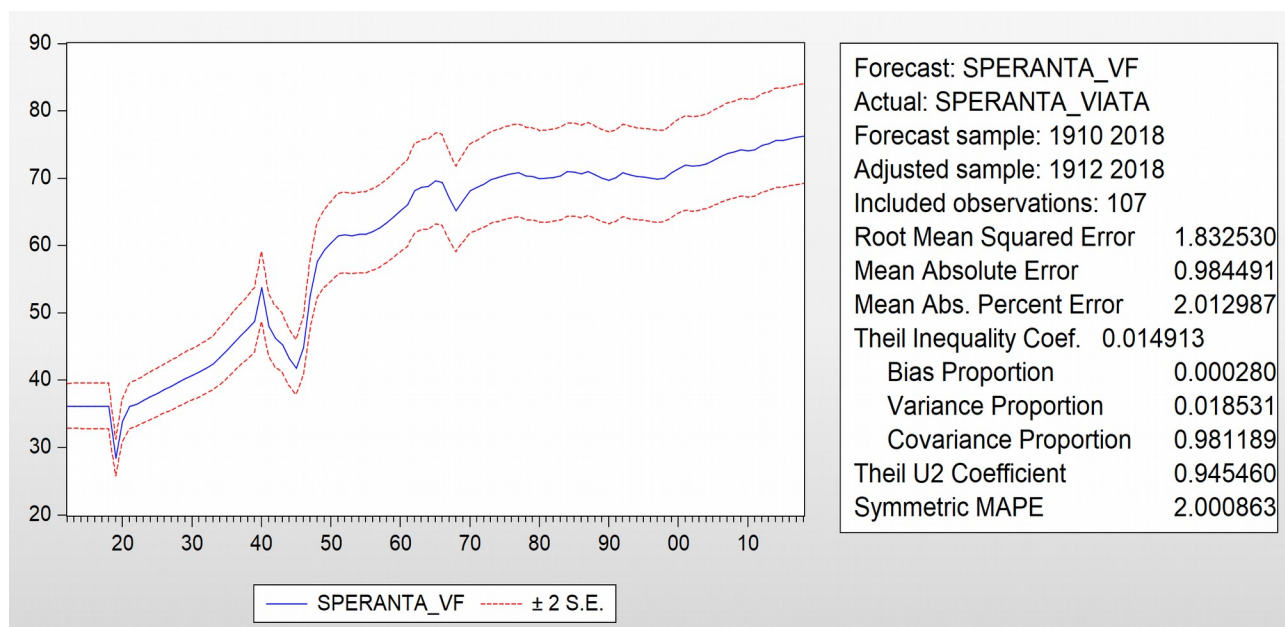
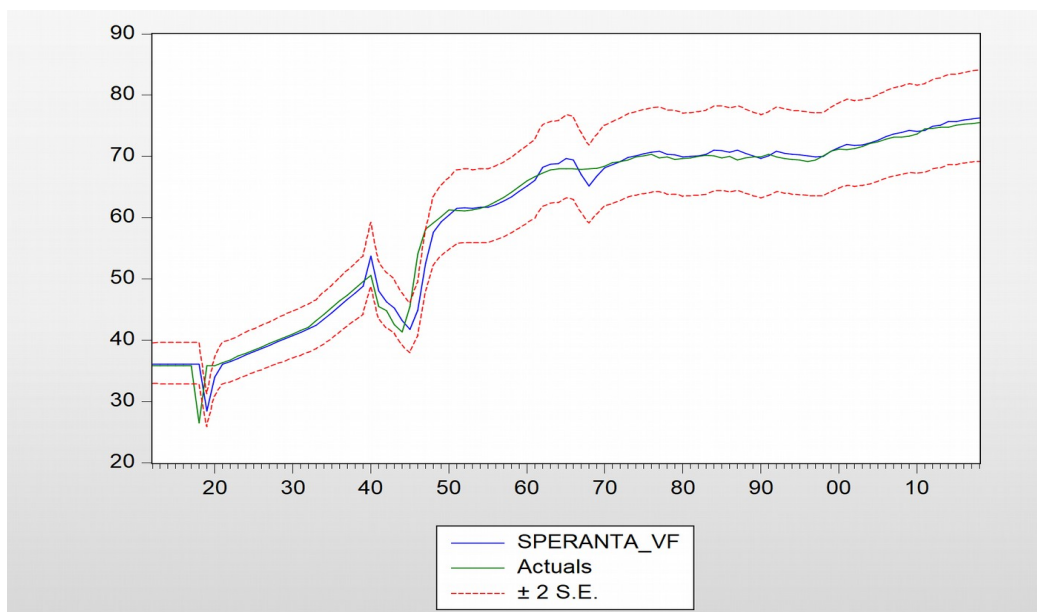


Figura 14. Graficul de predicție

În ciuda unei valori Root Mean Squared Error ce depășește nivelul de semnificație 0.05, putem observa că indicatorii Theil Inequality Coefficient, Bias Proportion și Variance Proportion au valori acceptabile, sub nivelul de semnificație, modelul fiind pregătit spre a fi comparat cu echivalentul său din realitate.

Prin comparație directă, putem observa similitudinea dintre graficele determinate de cele două modele, valorile estimate de către modelul nostru fiind, în cele mai multe dintre cazuri, în imediata proximitate a valorilor reale, existând, chiar, multiple intersecții între acestea.

Figura 15.
Comparație între
modelul obținut și
realitate



6. Bibliografie

- Pele, Daniel Traian; Curs de Serii de Timp, ASE București, 2019
- EViews – Time Series Estimation - <https://www.eviews.com/Learning/timeseries.html>
- Gapminder – <https://gapminder.org/data>