



Economia subterană la nivel de Europa



Cuprins

Introducere

Literature Review

Metodologie

Aplicația 1

Aplicația 2

Introducere

Scop: studierea economiei informale in Europa

Metode folosite: analiza unui model de regresie simpla, multifactoriala, cu dummy si termen de interactiune, analiza unui model de tip panel

Obiective:

- Testarea mai multor ipoteze, studiind relatia dintre regresor si regresant.
- Studierea impactului pe care variabilele independente (rata saraciei, accesul la informatie, cheltuieli cu sanatatea, educatia) il are asupra inegalitatii veniturilor.

Literature review

- Primele cercetări cu privire la economia subterană au încercat să îi estimeze mărimea și să definească conceptul de economie subterană
- Include activitățile economice, producere de bunuri și furnizare de servicii, care sunt ascunse de autoritățile publice, fie legale, fie ilegale
- Motivul pentru care economia subterană apare este pentru că indivizii/firmele vor să evite plata taxelor, a contribuțiilor la sistemul de sănătate, să îndeplinească anumite cerințe obligatorii minime de funcționare

- Studiile au aratat că o economie informală mare poate să țină pe loc dezvoltarea economiei unei țări
- Numeroase studii ale economiei informale cercetează cauzele apariției și ce soluții ar putea exista pentru minimizarea ei
- Studiile arată că economiile emergente au o rată mai mare a economiei informale, iar inegalitatea veniturilor poate determina emergența economiei informale

Metodologie

Aplicatia 1

Regresie Simpla

$$y = b_0 + b_1x_1 + \varepsilon$$

y - var explicata
coeficientul GINI
x - var explicativa
rata saraciei

Regresie Multipla

$$y = b_0 + b_1x_1 + b_2x_2 + \varepsilon$$

y - coef GINI
x1 - rata saraciei
x2 - acces la informatie

Regresie multipla-dummy

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

y - coef GINI
x1 - rata saraciei
x2 - accesul la informatie
x3 - variabila dummy
ECE = tari din Europa centrala si de est

Aplicatia 2

Regresie cu date panel

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

y - coef GINI
x1 - salariul minim pe economie
x2 - capitalul uman
x3 - cheltuieli cu educatie
x4 - cheltuieli cu sanatatea

Aplicatia 1

Regresia simpla

variabila cea mai
semnificativa este
rata saraciei

Scop: analiza impactului pe care
rata saraciei il are asupra inegalitatii
veniturilor

```
Call:
lm(formula = `Coeficientul GINI` ~ `Rata saraciei`, data = df)

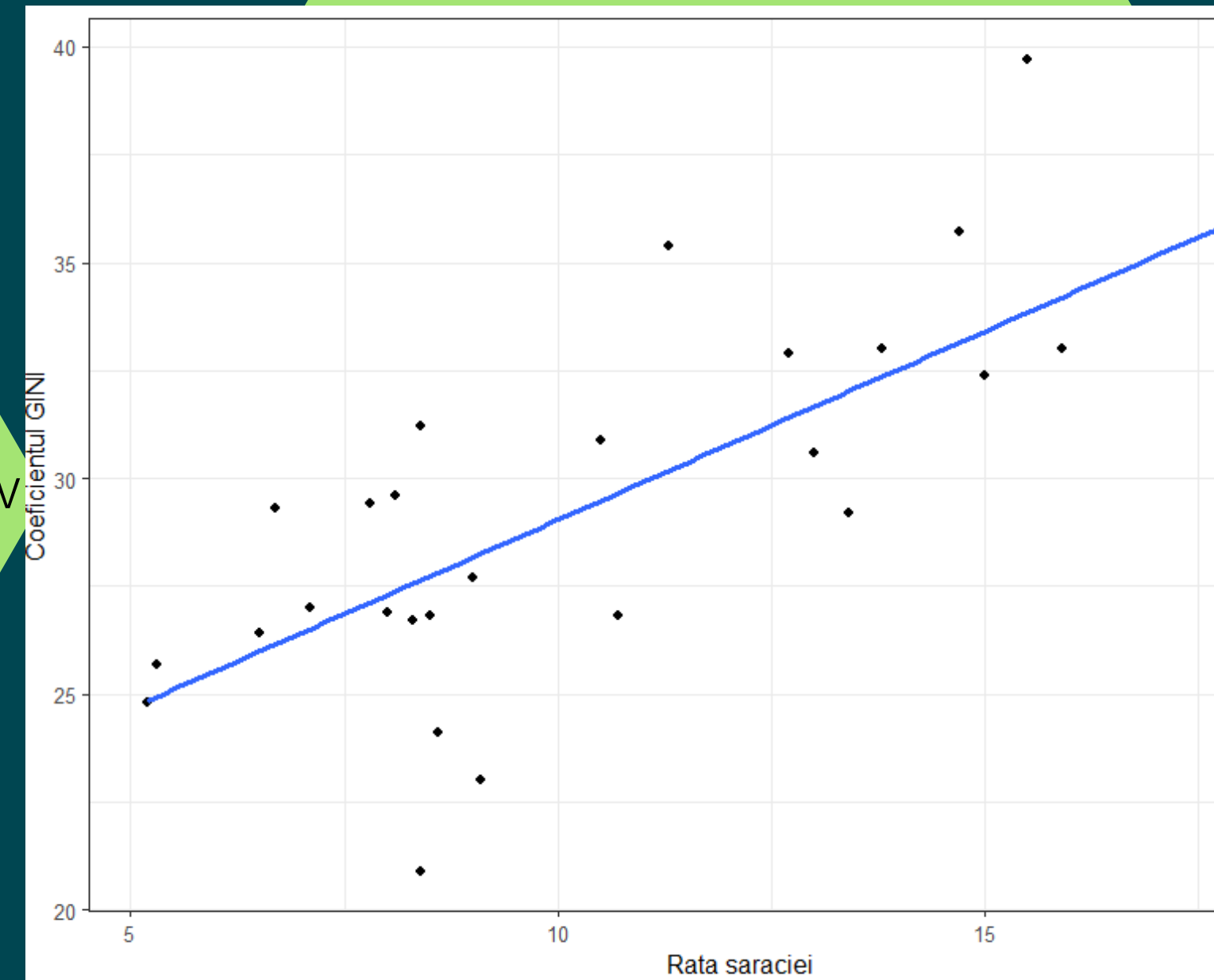
Residuals:
    Min       1Q   Median       3Q      Max
-6.744 -1.131 -0.047  1.858  5.851

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    20.302     1.778   11.416 2.08e-11 ***
`Rata saraciei`  0.874     0.162    5.396 1.34e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.981 on 25 degrees of freedom
Multiple R-squared:  0.538,    Adjusted R-squared:  0.5195
F-statistic: 29.11 on 1 and 25 DF,  p-value: 1.343e-05
```

R² 53.8
semnificativ
la 99%

Graficul valorilor observate



Testarea principalelor ipoteze pe regresia simpla

Bonitatea modelului:
 $R^2 = 0.538$

Variația factorului GINI este explicată
în proporție de 53.8% de rata saraciei.

Erori homoschedasticitate

- Breusch-Pagan
- White

p-value > 0.05 => nu avem
heteroschedasticitate

```
> bptest(model) # nu avem hetero
      studentized Breusch-Pagan test
data:  model
BP = 0.19123, df = 1, p-value = 0.6619
> white_test(model) # nu avem hetero
White's test results
Null hypothesis: Homoskedasticity of the residuals
Alternative hypothesis: Heteroskedasticity of the residuals
Test Statistic: 0.95
P-value: 0.622768
> ArchTest(df$uhat) # nu avem hetero conditionata
      ARCH LM-test; Null hypothesis: no ARCH effects
data:  df$uhat
Chi-squared = 12.45, df = 12, p-value = 0.4102
```

Erori non autocorelate

- Testele Durbin-Watson
- Breusch-Godfrey de ordin 2 și 3

p-value > 0.05 => nu
avem autocorelare

```
Durbin-Watson test
data:  model
DW = 1.4811, p-value = 0.08091
alternative hypothesis: true autocorrelation is greater than 0
> bgtest(model, order = 2) # nu avem
      Breusch-Godfrey test for serial correlation of order up to 2
data:  model
LM test = 3.6354, df = 2, p-value = 0.1624
> bgtest(model, order = 3) # nu avem
      Breusch-Godfrey test for serial correlation of order up to 3
data:  model
LM test = 3.6479, df = 3, p-value = 0.3021
```

Normalitate

- Testele Shapiro-Wilk
- Jarque-Bera

p-value > 0.05 =>
normal distribuite

```
Shapiro-Wilk normality test
data:  df$uhat
W = 0.98579, p-value = 0.9635
> jarque.bera.test(df$uhat)
      Jarque-Bera Test
data:  df$uhat
X-squared = 0.13105, df = 2, p-value = 0.9366
```

Am impartit setul de date in set de antrenament (80%) si set de testare (20%) si testam acuratetea pe setul de testare

```
> RMSE(y_pred, test.data$`Coeficientul GINI`)  
[1] 2.072679  
> MAE(y_pred, test.data$`Coeficientul GINI`)  
[1] 1.589718  
> mse(y_pred, test.data$`Coeficientul GINI`)  
[1] 4.295999  
> MAPE(y_pred, test.data$`Coeficientul GINI`)  
[1] 0.05149286  
> |
```

Forecast

Acuratetea modelului

Pnetru o valoare a ratei sărăciei de 16.8, putem sa previzionam ca coeficientul Gini va avea valoarea de 34.98 pentru Muntenegru

Modelul este semnificativ la 99%, mai bine decat modelul pe tot esantionul

Forecast Coeficientul Gini

din setul de testare

27.38

31.25

27.47

27.99

Model de regresie multipla

$y = b_0 + b_1x_1 + b_2x_2$
y - coeficientul GINI
x1 - rata saraciei
x2 - accesul la informatie

**Ipoteze
acceptate**



**Ipoteze
respinse**



Liniaritate	Numar observatii	Model corect specificat	X pozitiv
Media rezidurilor e 0	Multicoliniaritate	Rezidurile nu sunt corelate	Reziduri nu sunt autocorelate
Reziduri homoschedastice	Reziduri normal distribuite		



avem reziduri
heteroschedastice

Indicator de bonitate
Rpatrat = 0.54

Corectare ipoteze

Corectare heteroschedasticitate



**Breusch
Pagan**

Valori initiale	Valori dupa corectie
pValue=0.0624 < 0.1	pValue=0.105 > 0.1

White

pValue=0.66 > 0.1	pValue = 0.518 > 0.1
-------------------	----------------------

logaritmam cele 2 var
independente => cu
formula log-log

Model de regresie multipla cu variabila dummy

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

y - coeficientul GINI

x1 - rata saraciei

x2 - accesul la informatie

x3 - variabila dummy

ECE = tari din Europa centrala si de est

Reziduri homoschedastice

Breusch
Pagan



p value = 0.1062 > 0.1

White



p value = 0.66 > 0.1

Rezidurile nu sunt autocorelate

Durbin-
Watson



p value = 0.09039 < 0.1

Breusch-
Godfrey



p value = 0.2118 > 0.1

Rezidurile sunt normal distribuite

Jarque
Bera



Shapiro



Termen de interactiune: **Accesul la informatie x var dummy**

Bonitatea modelului

Rpatrat = 57.61%

Rajustat = 52.09%

Modele de penalizare

Implementam modelele Ridge, Lasso si Elastic Net Regression si comparam rezultatele obtinute

Regresia Ridge

```
5 x 1 sparse Matrix of class "dgCMatrix"
s0
(Intercept)      30.8565853
Rata saraciei    0.6246211
Accesul la informatie 0.2328026
Tari ECE        -1.5894900
Satisfactia vietii -1.1802684
```

Regresia Lasso

```
s0
(Intercept)      27.4985980
Rata saraciei    0.7342591
Accesul la informatie .
Tari ECE        -1.5463189
Satisfactia vietii -0.7865263
```

Algoritmul Boruta

```
Residuals:
    Min       1Q   Median       3Q      Max
-5.5258 -1.7354  0.5682  1.6428  3.9677

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    32.6917     6.6776   4.896 7.67e-05 ***
`Rata saraciei`  0.4711     0.2358   1.998  0.0588 .
`Efectul de redistribuire` -0.2707     0.1501  -1.804  0.0856 .
`Cheltuieli publice cu cercetarea` -0.2007     4.5526  -0.044  0.9653
`Chletuieli private cu cercetarea`  0.7980     2.1520   0.371  0.7145
`Rata saraciei in randul varstnicilor` 0.2355     0.1957   1.203  0.2423
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.767 on 21 degrees of freedom
Multiple R-squared:  0.6657,    Adjusted R-squared:  0.5861
F-statistic: 8.363 on 5 and 21 DF,  p-value: 0.0001769
```

Elastic Net Regression

```
s0
(Intercept)      28.3361989
Rata saraciei    0.7065715
Accesul la informatie .
Tari ECE        -1.5655110
Satisfactia vietii -0.8733373
```

Concluzii:

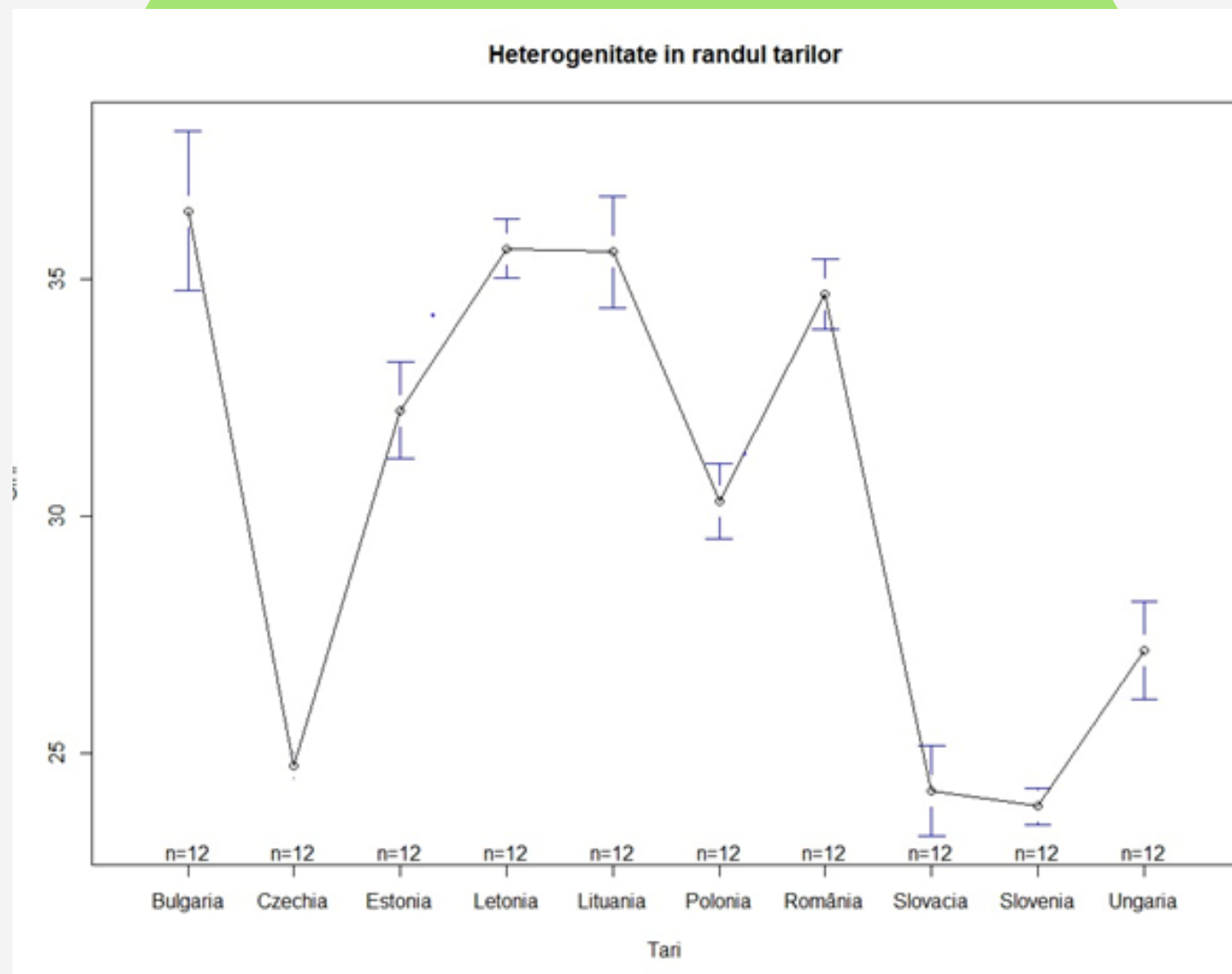
- Regresia Ridge marcheaza ca fiind nesemnificative Rata saraciei si Acces la informatie
- Regresia Lasso si Elastic Net Regression considera ca accesul la informatie nu e o variabila imporanta
- Algoritmul Boruta ne arata ce variabile sunt semnificative

Aplicatia 2 - Date panel

10 tari 11 ani

Explorarea
heterogenitatii:
presupune ca
avem diferente
intre unitatile
studiate

Heterogenitate transversala



Heterogenitate temporala



Interpretare:

=> avem
hetero trans

=> avem
hetero temp,
dar mai mica

Aplicatia2- Date panel

Model de regresie liniara

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  50.51087    2.70270   18.689  < 2e-16 ***
mmwics       -0.21479    0.06090   -3.527  0.000605 ***
em_ed_ter     0.37534    0.04581    8.193  3.99e-13 ***
ch_ed        -1.67710    0.31806   -5.273  6.38e-07 ***
ch_he        -2.54115    0.22045  -11.527  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.921 on 115 degrees of freedom
Multiple R-squared:  0.6795,    Adjusted R-squared:  0.6684
F-statistic: 60.95 on 4 and 115 DF,  p-value: < 2.2e-16
```

Model efecte fixe

Toate variabilele sunt semnificative

```
Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
mmwics       0.158873    0.062143   2.5566 0.011987 *
em_ed_ter    -0.185807    0.070523  -2.6347 0.009682 **
ch_ed        -0.875220    0.488848  -1.7904 0.076249 .
ch_he         1.045263    0.503052   2.0778 0.040140 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    244.84
Residual Sum of Squares: 217.42
R-Squared:              0.11199
Adj. R-Squared:         0.003078
F-statistic: 3.34185 on 4 and 106 DF, p-value: 0.012803
```

Model efecte random

```
Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  45.77130    4.32026  10.5946 4.163e-05 ***
em_ed_ter     0.57134    0.12237   4.6688 0.003436 **
ch_ed         -3.75323    0.92661  -4.0505 0.006722 **
ch_he         -2.47397    0.50682  -4.8814 0.002762 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    234.78
Residual Sum of Squares: 20.784
R-Squared:              0.91147
Adj. R-Squared:         0.86721
F-statistic: 20.5923 on 3 and 6 DF, p-value: 0.0014663
> |
```

H0: RE
H1: FE

Hausman Test

```
data:  Gini ~ mmwics + em_ed_ter + ch_ed + ch_he
chisq = 61.235, df = 3, p-value = 3.201e-13
alternative hypothesis: one model is inconsistent
```

se recomanda

Alegere
model FE

Testare efecte fixe in timp

H0: nu sunt necesare efecte fixe in timp

H1: sunt necesare efecte fixe in timp

pFtest

plmtest

F test for individual effects

```
data: Gini ~ mmwics + em_ed_ter + ch_ed + ch_he + factor(an)
F = 1.5761, df1 = 11, df2 = 95, p-value = 0.1183
alternative hypothesis: significant effects
```

```
> plmtest(fe, c('time'), type = 'bp')
```

Lagrange Multiplier Test - time effects (Breusch-Pagan)

```
data: Gini ~ mmwics + em_ed_ter + ch_ed + ch_he
chisq = 0.00040164, df = 1, p-value = 0.984
alternative hypothesis: significant effects
```

Concluzii

p-value > 0.05
nu sunt necesare efecte
fixe in timp

Ipotezele modelului cu efecte fixe

Ipoteze respinse

Dependenta
transversala



autocorelare



Dependenta
temporala



heteroschedasticitate



Ipoteze acceptate

Testarea efectelor
fixe in timp



Va multumim!