Analysis of the labour market in Romania in relation with working time

Nicoleta CARAGEA
Ciprian ALEXANDRU
Ecological University of Bucharest and National, Institute of Statistics, Romania

ABSTRACT

In this research study there were applied Regression models to examine the socio-economic factors that could influence the working time on the labor market. As a result of the regression models applied, the most significant factors to consider here are the sex of the employed people, age, education level, residence areas where they live, the activity and the occupational status on the labour market. For computing the regression models it was used the glm function from the nlme package in R.

This research will contribute to know the determinants of working time in Romania. The data from a Romanian labor force survey 2013 is used for this study.

Key words: Labor Force Survey, GLM Regression, Working time, R, Packages

JEL Classification: J21, C50, C87

INTRODUCTION

The relationships established between two or more economic variables could be described using mathematical models. Although economic and social phenomena have a high degree of complexity, being most often oversized in terms of influence factors, the econometric models simplify the real world helping us to understand them more easily. It is important, however, that the premises on which it is based models do not depart so far from reality that the results are useless, being unreliable. The main theme of the study was to analyze the working time using econometric estimation methods and techniques based on the main socio-economic characteristics of the population in Romania. The results of regression models show the effects of socio-demographic characteristic of the labour market on the usual and actual duration or working week at national level.

VARIABLES, DATA SOURCES AND SOFTWARE

The econometric methods developed in the linear regression models (LM) and generalized linear regression models (GLM) were formulated by John Nelder and Wedderburn Robert as a way to unify the various statistical models, including logistic regression and Poisson regression.

For analysis were used data Labour Force Survey (LFS) in 2013, conducted by the National Statistics Institute. To achieve regression models were used nlme package glm function of R. The package provides summary of results and includes coefficients and standard deviations by which they were estimated and significance tests of the estimated values.

Dependent Variables are following:

- dure – actual duration or working week (in hours);
- duro – usual duration or working week (in hours);

Independent Variables (Predictors) are following:

- *Gender* – is a dummy variable for gender that is, Male = 1 and Female = 2
- *Age (Age Groups)* - Age variable was available in LFS 2013 as a continuous variable that was further converted into a categorical variable with different group showing six different stages of life [Age Group (1=15 to 19, 2=20 to 24, 3=25 to 34, 4=35 to 44, 5=45 to 54, 5=55 to 64)].
- *Residence area* – is a dummy variable for gender that is, Urban = 1 and Rural = 2
- *Education* - a categorical Variable of Education with 6 categories [Education (1=no formal education, 2=primary and lower secondary education, 3=upper secondary education, 4=post-secondary, non-tertiary education, 5=apprenticeships, technical or vocational education, 6=tertiary education)].
- *Ethnicity* a categorical Variable of Ethnicity with 5 categories [Ethnicity (1=Romanian, 2=Hungarian, 3=Romanic, 4=German, 5=other ethnicity)].
- *Professional status* - a categorical Variable of Professional status with 3 categories [Professional status (0=no workers, 1=salary workers, 2=other non-salary workers)].
- *Economic activity* – a categorical Variable of Economic activity with 3 categories [Economic activity (0=no economic activity, 1=agriculture, 2=industry, 3=construction, 4=commercial services, 5=social services)].
- *Occupations* - Occupation are classified into following 10 different major categories, set by International Standard Classification of Occupation ISCO 2008.

Categorical Variable of Occupations with 11 categories as follows:

No occupation	0
Managers	1
Professionals	2
Technicians and associate professionals	3
Clerical support workers	4
Service and sales workers	5
Skilled agricultural, forestry and fishery workers	6
Craft and related trades workers	7
Plant and machine operators, and assemblers	8
Elementary occupations	9
Armed forces occupations	10

- *Regions* - according to Romanian administrative territorial structure, there are 8 Regions - Categorical Variable [Regions (1=North-West, 2=Center, 3= North-East, 4=South-East, 5=South, 6=Bucharest-Ilfov, 7=South-West, 8=West)].

RESULTS OF THE MODELS

In this section of the paper there are presented the results obtained from the regression models; the labour time, expressed in hours actually worked per week, respectively hours usually worked per week, is estimated using certain explanatory variables. Regression models are analytical tools whose main purpose in understanding the role of variables to explain a phenomenon of interest (working time, in this case). Another purpose of this model based approach was to establish the importance of predictors in relation of the dependent variable. In this respect, the explanatory variables are selected using the correlation criteria: those for which there is statistically significant correlation with the dependent variable in hierarchical order of importance.

Variable selection

It was used a stepwise regression model. Forward stepwise selection starts with the intercept, and then sequentially adds into the model the predictor that most improves the fit. Backward-stepwise selection starts with the full model, and sequentially deletes the predictor that has the least impact on the fit. The candidate for dropping is the variable having poor correlation with the dependent variable (in the R package[1] the step function uses the AIC criterion for weighing the choices, which takes proper account of the number of parameters fit; at each step an add or drop will be performed that minimizes the AIC score).

The relationship between the usual hours of work and socio-demo-economic variables

The main goal is to estimate the usual week duration of work, measured in hours, based on factors that could produce variation in labour time duration: i.e. sex, age, residence area, education, nationality, economic activity, occupation and employment status.

The regression equation, when the dependent variable is the usual working time, it is of the form:

$$duro = \beta_0 + \beta_1 sex + \beta_2 mediu + \beta_3\ var\ sta + \beta_4 nivs + \beta_5 nat + \beta_6 act + \beta_7 ocup + \beta_8 stap$$

Using R, nlme package glm function returns the following results:

1. Other more traditional packages base the selection on z-score pr F-statistics, adding "significant" terms, and dropping "non-significant" terms. These are out of fashion, since they do not take proper account of the multiple testing issues. R software packages implement hybrid stepwise-selection strategies that consider both forward and backward moves at each step, and select the "best" of the two.

```
Call:
glm(formula = duro ~ sex + mediu + varsta + nivs + nat + act + ocup + stap, data =
amigo_TIMP)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-34.879   -1.265   -0.548    0.440   32.320

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 45.035154   0.302387 148.932  < 2e-16 ***
sex         -1.221927   0.060117 -20.326  < 2e-16 ***
mediu       -0.224892   0.034666  -6.487 8.86e-11 ***
varsta      -0.096263   0.025692  -3.747 0.000179 ***
nivs        -0.088604   0.023521  -3.767 0.000166 ***
nat          0.162165   0.081994   1.978 0.047963 *
act          0.326181   0.025804  12.641  < 2e-16 ***
ocup         0.002367   0.015024   0.158 0.874836
stap        -2.699496   0.089904 -30.026  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 26.87144)

    Null deviance: 947119  on 32357  degrees of freedom
Residual deviance: 869264  on 32349  degrees of freedom
AIC: 198331

Number of Fisher Scoring iterations: 2
```

```
> Regresia_mea <- step(glm(duro ~ sex + mediu +varsta+ nivs + nat+ act + ocup+stap, data = amigo_TIMP))
Start:  AIC=198331.3
duro ~ sex + mediu + varsta + nivs + nat + act + ocup + stap

         Df Deviance    AIC
- ocup    1   869265 198329
<none>       869264 198331
- nat     1   869369 198333
- varsta  1   869641 198343
- nivs    1   869645 198343
- mediu   1   870395 198371
- act     1   873558 198489
- sex     1   880366 198740
- stap    1   893491 199219

Step:  AIC=198329.3
duro ~ sex + mediu + varsta + nivs + nat + act + stap

         Df Deviance    AIC
<none>       869265 198329
- nat     1   869370 198331
- varsta  1   869643 198341
- nivs    1   869730 198345
- mediu   1   870403 198370
- act     1   873906 198500
- sex     1   880915 198758
- stap    1   894402 199250
```

Step procedure removes the variable charge (occupation) of the model as parameter regression equation is not statistically significant (p> 0.05).

The model equation is:

$$duro = \beta_0 + \beta_1 sex + \beta_2 mediu + \beta_3\ var\ sta + \beta_4 nivs + \beta_5 nat + \beta_6 act + \beta_7 stap$$

It is noted that sex, residence area, age, education level and employment status are inversely correlated with the usual hours of work, while nationality and economic activity are directly correlated.

Increasing by one year of age, the estimated usual working time week could be reduced with 0.096 hours. Also, the higher the level of education of the employees is higher, the estimated hours per week on average decrease. For people employed in rural areas compared to those employed in urban areas, the working week duration is less (with 0.22 hours), regardless of gender, age, nationality, education level, economic activity and employment status. Regarding the relationship between usual hours of work and employment status, it is observed that persons employed, but are not employed, compared with employees, working an average of 2.69 hours less each week, while the other variables remain constant.

3.3. The relationship between the actual duration of the working week and socio-demo-economic variables

The regression equation, when the dependent variable is the actual working time, it is of the form:

$$dure = \beta_0 + \beta_1 sex + \beta_2 mediu + \beta_3\ var\ sta + \beta_4 nivs + \beta_5 nat + \beta_6 ocup + \beta_7 act + \beta_8 stap$$

Using R, nlme package glm function returns the following results:

```
Call:
glm(formula = dure ~ sex + mediu + varsta + nivs + nat + act +
    ocup + stap, data = amigo_TIMP)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-41.724   -1.104   -0.032    1.198   33.982

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 45.11105    0.40064 112.598  < 2e-16 ***
sex         -1.53919    0.07965 -19.325  < 2e-16 ***
mediu       -0.29161    0.04593  -6.349 2.19e-10 ***
varsta       0.03654    0.03404   1.073   0.2831
nivs        -0.09584    0.03116  -3.075   0.0021 **
nat          0.10547    0.10863   0.971   0.3316
act          0.38149    0.03419  11.158  < 2e-16 ***
ocup         0.02105    0.01990   1.057   0.2903
stap        -3.38389    0.11912 -28.409  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 47.16994)

    Null deviance: 1644297  on 32357  degrees of freedom
Residual deviance: 1525900  on 32349  degrees of freedom
AIC: 216539

Number of Fisher Scoring iterations: 2
```

```
> Regresia_mea <- step(glm(dure ~  sex + mediu +varsta+ nivs + nat+ act + ocup+stap, data = amigo_TIMP))
Start:  AIC=216538.9
dure ~ sex + mediu + varsta + nivs + nat + act + ocup + stap

         Df Deviance    AIC
- nat     1  1525945 216538
- ocup    1  1525953 216538
- varsta  1  1525955 216538
<none>       1525900 216539
- nivs    1  1526346 216546
- mediu   1  1527802 216577
- act     1  1531773 216661
- sex     1  1543515 216908
- stap    1  1563968 217334

Step:  AIC=216537.8
dure ~ sex + mediu + varsta + nivs + act + ocup + stap

         Df Deviance    AIC
- varsta  1  1525995 216537
- ocup    1  1525999 216537
<none>       1525945 216538
- nivs    1  1526398 216545
- mediu   1  1527834 216576
- act     1  1531805 216660
- sex     1  1543563 216907
- stap    1  1564009 217333

Step:  AIC=216536.9
dure ~ sex + mediu + nivs + act + ocup + stap

        Df Deviance    AIC
- ocup   1  1526048 216536
<none>      1525995 216537
- nivs   1  1526450 216545
- mediu  1  1527925 216576
- act    1  1531812 216658
- sex    1  1543627 216907
- stap   1  1564016 217331

Step:  AIC=216536
dure ~ sex + mediu + nivs + act + stap

        Df Deviance    AIC
<none>      1526048 216536
- nivs   1  1526740 216549
- mediu  1  1527932 216574
- act    1  1532041 216661
- sex    1  1544919 216932
- stap   1  1565986 217370
```

Step procedure removes the following variables: age (age), nationality (nat) and employment (occupation), because parameters are not statistically significant ($p > 0.05$).

The model equation is:

$$duro = \beta_0 + \beta_1 sex + \beta_2 mediu + \beta_3 nivs + \beta_4 act + \beta_5 stap$$

It is noted that sex, residence, education level and employment status are inversely correlated with the usual hours of work, while economic activity is correlated directly.

Females actually work, on average, less than 1.53 hours per week compared to the male. Also, if the level of education of employees is higher, it is expected to decrease the working hours per week. For people employed in rural areas compared to those employed in urban areas during the actual working week is less, on average, 0.29 hours, regardless of gender, level of education, economic activity and employment status. Regarding the relationship between the actual duration of the working week and professional status, it is observed that persons employed, but are not salary workers, work an average of 3.38 hours less each week, while the other variables remain constant.

CONCLUSIONS

The relations established between two or more socio-economic phenomena can be described and understood more easily by using mathematical models. Regression models used in this paper lead to results that can help to quantify the dependence between working time and the factors that influence the number of hours spent for economic activities by labour market employment. The results also identify the intensity regression models examined the link between the dependent variables and factors of influence.

Actual weekly working time does not depend on factors such as age of employed labor market, nor the nationality and their occupation. Sex, residence, education level and employment status are inversely correlated with the usual hours of work, while economic activity is correlated directly.

The level of education of employment affects the duration of working week, meaning that low-skilled people tend to work more hours per week than highly qualified. The effects of so-called "constraints" working time can be multiple: first, an extended working time is reflected in reduced time spent outside the labour market, which is the main component of leisure, but also the time for family activities.

Both the actual duration and usual hours of work depends inversely by professional status of the people working in the labour market, meaning that people who are employed but they are not salary workers spent for work more than 2 hours less each week. This result leads to the idea that employers and self-employed allocate less time for economic activities compared with the salary workers. This is not true! The results should be interpreted taking into account the following: the number of

employers is very small in the sample because in Romania many employers declare their professional status being employees (the minimum wage) and the self-employed people are, in general, workers in agriculture is seasonal activity.

By gender, the situation is as follows: females actually work in economic activities on the labour market, an average of 1.53 hours per week less than the male. This affirmation is supported mainly by the fact that the employment rate among females is lower than for men.

ACKNOWLEDGMENT

This paper has been financially supported within the project entitled „SOCERT. Knowledge society, dynamism through research", contract number POSDRU/159/1.5/S/132406. This project is co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013. Investing in people!"

Bibliografie

1. Aldrich, J. H., Nelson, F. D., 1984, Linear probability, logit, and probit models, Thousand Oaks, CA: Sage.
2. Caragea, N., 2015, Statistică - concepte şi metode de analiză a datelor, Editura Mustang, Bucuresti, ISBN 978-606-652-063-8.
3. Eeckhout, J., Jovanovic, B., 2010, Occupational Choice and Development, Journal of Economic Theory. 147(2), 657-683
4. Fox, J., 2008, Applied Regression Analysis and Generalized Linear Models. 2nd edition. Sage, Thousand Oaks CA
5. Friedman, J., Hastie, T., Tibshirani, R., 2010, Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33 (1).
6. Garson, G. D., 2011, Logistic Regression, from Statnotes: Topics in Multivariate Analysis. http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.
7. Liebenberg, A., Hoyt, R., 2003, The determinants of enterprise risk management: evidence from the appointment of chief risk officers, Risk Management and Insurance Review. 6, 37–52
8. Lisaniler, G.F., Bhatti, F., 2005, Determinants of Female Labour Force Participation: A Study of North Cyprus. Review of Social, Economic & Business Studies. 5(6), 209 – 226
9. Pisică, S., Vasile, V., 2012, Piata muncii intre formal si informal, in Revista Romana de Economie, Vol. 34 Issue 1, pp.191.