Multiple Linear Regression Equation for Economic Dimension of Standard of Living

Nicoleta Mihaela Florea[1], Georgeta Mădălina Meghişan[2], Cristina Nistor[3]
[1,2, 3]University of Craiova
[2]Faculty of Economics and Business Administration, Craiova
Scientific Researcher III, Romanian Academy
National Institute of Economic Research "Costin C. Kiritescu", Bucharest

niko.2903@yahoo.com;madalina_meghisan@yahoo.com;cristinanistor88@yahoo.com

Abstract. *The purpose of this analysis is to determine the signification of the following factors: population, population density and inflation rate to the measurement of the standard of living. The analysis takes into consideration demographic and economic data for a number of 10 EU member states. After calculating the standard of living in these countries, we analysed the impact of the mentioned factors using a multiple linear regression equation. We concluded that a part of the variation of the standard of living depends on the evolution of the three variables taken into consideration in the analysis.The standard of living of a population assesses the economic dimension of a country and the quality of life for a population.*

Keywords: living standard, European Union, inflation, population density, population, multiple linear regression equation.

JEL classification: C2, E2, E3.

1. Introduction

Many economists have long been concerned with evaluating the factors that may have an influence on the standard of living of a population. Some authors (Romer, 1986; Wolff, 1991; Mankiw et al., 1992) designed several models of convergence between the level per capita production at an initial point and the growth rate per capita production over a time period series. In their studies, Evans and Karras (1993) underlined the fact that there is a convergence between per capita consumption properties within rich countries. Ravallion's study (1994) focuses on standard of living measurement errors for estimating an individual-specific poverty line, due to imperfect information that can be found on "various consumption needs".

Other authors (Cooper et al., 2015) evaluated the standard of living within an era of fast technological change, focusing on two consumer demand systems: AIDS (Almost Ideal Demand System) of Deaton and Muellbauer (1980) and QAIDS (Quadratic Almost Ideal Demand System), which were adapted in order to estimate the income effects.

In our study, we focus on the measurement of the standard of living, using the following independent variables: population, population density and inflation rate, considered to be important measures of welfare. This analysis, connected with the studies made before by other researchers, can be part of a complex standard of living measurement model.

2. Method and results

This article focuses on the analysis of the dimension of standard of living in 10 member states of the European Union, during the period 2004-2013. In order to determine the key factors that have a contribution to this, we used panel data, which was analyzed using multiple linear regression equation.

The dependent variable, the standard of living, was determined as a ratio between annual GDP per capita and the annual final consumption expenditures per capita. The population consumption expenditures were taken into consideration, because they measure the objective economic dimension of the standard of living, while the other expenses (e.g. the access to IT, telecommunications services) emphasize more on the subjective dimension of the standard of living, meaning the quality of life.

The multiple linear regression focuses on the explanation of a dependent variable, with the help of *m* independent variables, where $m>1$. According to the mathematical model, the following relations characterize our equation (1, 2):

$$y = f(x_1, x_2, \dots, x_m) \qquad (1)$$

$$\text{where, } f(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \qquad (2)$$

with $c, c_1, c_2, \dots, c_m$ representing the parameters of the equation.

We will suppose that this is a stochastic relationship, where the econometric model is (3):

$$y = f(x_1, x_2, \dots, x_m) + \varepsilon \qquad (3)$$

According to multiple linear regression, in an analytical writing, the equation becomes (4):

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + \varepsilon_i, \quad i = 1,n \qquad (4)$$

where ε_i is the error of specification and it has a known probabilistic distribution (it is an random variable), while *n* represents the number of observations.

The forecast model is (5):

$$SOL_t = \alpha + \beta_1 POP_t + \beta_2 DENSPOP_t + \beta_3 INFL_t + \varepsilon_t \qquad (5)$$

All the data in the panel regression model is in annual frequency and takes into consideration the period 2004-2013.
Where (Table 1):

Table 1. Variables definition

Variable	Definition of the variable	Time series
SOL	Standard of living (GDP per capita/ total consumption expenditure per capita)	2004-2013
POP	Population (measured in thousands) – logarithmic function was used for this	2004-2013

	data, for a better precision.	
DENSPOP	Density of population (number of persons/ squared kilometer)	2004-2013
INFL	Inflation rate (expressed in percentage)	2004-2013

Source: Authors' own encoding

The following results were obtained, using the EViews informatics program package, in order to estimate the parameters of the model:

Dependent Variable: SOL
Method: Panel Least Squares
Date: 03/22/15 Time: 17:40
Sample: 2004 2013
Periods included: 10
Cross-sections included: 10
Total panel (balanced) observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.545030	0.509786	6.953957	0.0000
POP	-0.332416	0.085437	-3.890774	0.0002
DENSPOP	0.005348	0.001095	4.884905	0.0000
INFL	0.021120	0.009929	2.127191	0.0360

R-squared	0.203388	Mean dependent var	1.838165
Adjusted R-squared	0.178493	S.D. dependent var	0.293138
S.E. of regression	0.265692	Akaike info criterion	0.226219
Sum squared resid	6.776846	Schwarz criterion	0.330425
Log likelihood	-7.310934	Hannan-Quinn criter.	0.268393
F-statistic	8.170097	Durbin-Watson stat	0.871514
Prob(F-statistic)	0.000067		

Figure 1. Estimation of the parameters of the model
Source: Data according to Eviews informatics program package

Based on the information from the Figure 1, obtained with the Eviews informatics program package, the following statements on the regression equation can be made:

- the free element from the regression equation (*c*) is α = 3,545030 and it represents the point where all the explanatory variables (population, density of the population and inflation rate) are equal to zero and have a standard error of 0,509786;
- the value of the β_1 coefficient is -0,332416. This value can be explained the following way: when the population undertakes a raise of a unity, the standard of living decreases with 0,332416 unities. Because the value of *P* for this parameter is 0,0002 $\leq$ 0,05, we can affirm that the parameter is statistically significant;
- the value of β_2 is 0,005348. Thus, at a raise of the population density with a unity, the standard of living would rise with 0,005348 unities. The positive coefficient associated with the population density was expected because entertainment and leisure related industries thrive in dense population countries. For example, one finds

more movie theatres, bowling alleys, skate rinks, etc. in densely populated countries versus non-densely populated countries. These leisure related industries also tend to be located in more affluent areas, where the people can afford not only to pay for the goods and services with their money, but also with available time. This aspect can be mainly explained by the existing situation of two of the analyzed countries, Czech Republic and Poland, which have the highest density of the population and also the highest standard of living. The value of *P* (0,0000) is higher than 0,05, se we can affirm that the parameter is statistically significant;

- the β_3 coefficient has the value of 0,021120. In other words, at a raise of the inflation rate with a unity, the standard of living will encounter a growth of 0,021120 unities. The fact that the probability (*P*) associated to this parameter is 0,0360 (lower than 0,05) leads us to the conclusion that the parameter β_3 is statistically significant;
- the coefficient of determination (R^2) is 0,2033, which indicates that only 20,3388% of the standard of living variation can be explained by the three independent variables taken into analysis. A higher percent could be encountered by adding other variables that were omitted in this model;
- *log likelihood* = -7,310934 represents the logarithm of the likelihood function (supposing that the errors have a normal distribution); this function is determined taking into consideration the estimated values of the parameters. The function for the calculation of this indicator, used by EViews informatics program package, is (6):

$$L = (n/2)\left(1 + \ln(2\pi) + \ln\left(\sum \hat{u}_t^2 / n\right)\right) \qquad (6)$$

where: $\sum \hat{u}_t^2 =$ sum of squared errors;

k = number of exogenous variables;

n = number of observations.

This indicator is used for statistical tests that find omitted variables from an econometric model, together with some tests for finding out redundant variables from an econometric model, such as, for instance, LR test or the ratio of verisimilitudes (Likelihood Ratio).

- from the calculation of Fisher statistics (*F-statistic* = 8,170097), the hypothesis according to which the model is not valid is rejected and we can conclude that the regression model is statistically valid. However, the associated probability (*F-statistic* = 0,000067 < 0,05) strengthens the affirmation that the constructed multiple linear regression model is statistically valid;
- *mean dependent var* = 1,838165 represents the average of the dependent or endogenous variable, having the following calculation relationship (7):

$$\bar{y} = \sum_{i=1}^{n} y_i \Big/ n \qquad (7)$$

- Akaike criterion is used for comparing two or more econometric models. The calculation relationship for it, used by the EViews informatics program package, is (8):

$$AIC = -2L/n + 2k/n \qquad (8)$$

where *L* = Log likelihood;

The decision rule used for this test is the following: it will be chosen the econometric model for which the lowest value for this indicator was obtained.

- Schwartz criteria is also used in order to compare two or more econometric models. The calculation relation for this, used by EViews informatics program package, is (9):

$$SC = -2L/n + k \ln n / n \qquad (9)$$

In this case also, it will be chosen the econometric model for which the lowest value for this indicator was obtained.

The low Durbin-Watson statistic indicates serial correlation. An AR (1) process could offset the serial correlation. However, after taking into consideration the AR (1) process (Figure 2), the model estimates become spurious. The presence of serial correlation in the model does not take away its significance. The model remains non-bias and consistent. Serial correlation only affects parameter estimates. If anything, the presence of serial correlation downplays the significance of the model.

Dependent Variable: SOL
Method: Panel Least Squares
Date: 03/23/15 Time: 22:55
Sample (adjusted): 2005 2013
Periods included: 9
Cross-sections included: 10
Total panel (balanced) observations: 90
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.702752	0.712018	5.200361	0.0000
POP	-0.382257	0.117256	-3.260016	0.0016
DENSPOP	0.005632	0.001441	3.909761	0.0002
INFL	-0.007427	0.003543	-2.095973	0.0391
AR(1)	0.745756	0.035581	20.95912	0.0000

R-squared	0.876270	Mean dependent var	1.788292
Adjusted R-squared	0.870447	S.D. dependent var	0.252556
S.E. of regression	0.090904	Akaike info criterion	-1.904081
Sum squared resid	0.702395	Schwarz criterion	-1.765203
Log likelihood	90.68366	Hannan-Quinn criter.	-1.848077
F-statistic	150.4943	Durbin-Watson stat	1.967292
Prob(F-statistic)	0.000000		

Figure 2. Model significance

Source: Data according to Eviews informatics program package

3. Conclusions

According to the analysis made, the evolution of the standard of living in EU member states is influenced by the three independent variables chosen: population, density of the population and inflation rate. However, the multiple linear regression equation can be improved by the inclusion of other variables such as: poverty rate, unemployment rate, personal bankruptcy and by the enlargement of the number of observations. These variables could not be included in the regression equation, because available data for the analyzed states could not be found.

References

Cooper, R.J., McLaren, K.R., Rehman, F., Szewczyk, W.A., 2015. Economic welfare evaluation in an era of rapid technological change. J. Economics Letters, in press-accepted manuscript.
Deaton, A., Muellbauer, J., 1980. An almost ideal demand system. Am. Econ. Rev. 70 (3), 312–326.
Evans, P., Karras, G., 1993. Do standards of living converge?: Some cross-country evidence. J. Economics Letters. 43 (2), 149-155.
Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. Quarterly Journal of Economics, 107, 407–438.
Ravallion, M., 1994. Poverty rankings using noisy data on living standards. J. Economics Letters. 45, 481-485.
Romer, P.M., 1986. Increasing returns and long-run growth, Journal of Political Economy, 94, 1002–1037.
Wolff, E.N., 1991. Capital formation and productivity convergence over the long term. American Economic Review, 81, 565–579.