# ANALYSIS OF THE RELATIONSHIP BETWEEN THE SHARE OF PEOPLE LIVING IN HOUSEHOLDS WITH VERY LOW WORK INTENSITY AND SELECTED SOCIO-ECONOMIC INDICATORS IN THE CZECH REPUBLIC IN THE PERIOD 2005 - 2016

**Dagmar Blatná**

**Abstract**

The proportion of people living in households with very low work intensity (LWI) is one of the indicators of social exclusion and social policy. LWI is defined as the percentage of the population living in households whose working-age members worked less than 20 percent of their total work potential over the previous twelve months. The proportion of people living in households with very low work intensity was significantly lower in the Czech Republic than in the European Union throughout the analysed period 2005–2016. The aim of this paper is to study the development of the LWI indicator in the Czech Republic, presenting the results of the regression analysis of the LWI and selected measurable socio-economic factors such as real per capita GDP growth rate, inflation, [un]employment, educational attainment, level of lifelong learning, social benefits, etc. The theory of multivariate non-stationary time series was used as the main tool of analysis.

**Key words:** time series, regression analysis, stationarity, non-stationarity

**JEL Code:** C32, D10, I32

## Introduction

According to the Eurostat definition "people living in households with very low work intensity" are those aged 0-59 who live in households where the adults worked on average less than 20 percent of their total work potential over the past year. A working-age person is the one aged between 18 and 59 years except for students in the 18–24-age group. The work intensity of a household is the ratio of a total number of months actually worked by all household adults during the income reference year to the number of months the same household members theoretically could have worked in the same period.

The LWI indicator is one of the factors of social exclusion referring to the third dimension of poverty and one of the headline indicators monitoring the objectives of the Europe 2020 Strategy.

The present paper aims at analysing statistically one of the sets of social exclusion and social policy indicators – the proportion of people living in households with very low intensity. It also explores the possibility of applying a multivariate time-series regression model, outlining the relationship between the proportion of people living in households with very low intensity of work and selected socio-economic factors.

# 1 Methodology

Both the dependent and explanatory variables being arranged in time series, it is necessary to decide whether the latter are stationary or non-stationary prior to using them in the regression analysis. Stationary time series are those with an autoregressive model of zero order I(0), also known as short memory ones, non-stationary time series being generated by an order one I(1) autoregressive process and referred to as long memory ones. The order of integration is the number of unit roots contained in the series, or the number of differencing operations to be performed to make the series stationary.

In the regression analysis, the time series of both variables must be of the same order integrated rate. The "classical" regression model can be used when the analysed time series are zero order stationary ones. Applying the unit root tests, it can be concluded that the series are not of the same order, showing no relationship. When unit root tests of a linear combination of two non-stationary time series do not reject the non-stationarity of the non-systematic component, there is only spurious regression. The relation between the two non-stationary time series exists only if their linear combination is stationary. Then there is a co-integration relationship considered as a long-term one.

The formal method of testing the stationarity of a series is the unit root test. We used the Augmented Dickey-Fuller test (ADF) for the verification of the null hypothesis –

$H_0$: $\phi_1 = 1$ for non-stationary I(1) time series and

$H_1$: $|<\phi_1| 1$ for stationary I(0) time series.

The ADF test statistic is defined as

$$t = \frac{\hat{\phi}_1 - 1}{S_{\hat{\phi}_1}}, \tag{1}$$

where $\hat{\phi}_1$ is an estimate of the autoregressive parameter of model $y_t = \phi_1 y_{t-1} + a_t$, $S_{\hat{\phi}_1}$ is an estimate of the standard error of $\hat{\phi}_1$ and $a_t$ is a non-systematic component with white noise characteristics, i.e. the series of uncorrelated random variables $\mathrm{cov}(a_t, a_{t-k}) = 0$, probability distribution $N(0, \sigma_a^2)$ with zero mean and constant variance $D(a_t) = \sigma_a^2$. The test statistic follows the Dickey-Fuller distribution; for critical values, see Dickey and Fuller, 1979. For details, see e.g. Arlt and Arltová (2009), Caner and Kilian (2001), Dickey and Fuller (1979, 1981), Elliot, Rothenberg and Stock (1996) and Phillips (1987).

When the non-systematic component is auto-correlated, the Autoregressive Distributed Lag (ADL) model with time-shifted variables is utilized for estimation. It can be written as

$$Y_t = c + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + a_t, \tag{2}$$

where $Y_t$ is the dependent variable in time $t$, $Y_{t-1}$ is the dependent variable in delay $t$-1, $X_t$, $X_{t-1}$ is the explanatory variable in time $t$ and delay $t$-$1$, $\alpha$ and $\beta_i$ are model parameters, $c$ is a constant and $a_t$ has a white-noise characteristic. For more details, see Arlt (1998), Arlt and Arltová (2009), Hušek (2007), Hendry, Pagan and Sargan (1984).

Validation of the calculated regression model is performed using diagnostic tests of the non-systematic component of the model. To verify the normality, the Jarque-Bera test is employed, autoregressive conditional homoscedasticity and auto-correlation being confirmed by ARCH(1) and Breusch-Godfrey LM tests, respectively. For details, see Jarque and Bera (1980).

## 2 Analysis of LWI development in the EU and the CR in 2005–2016

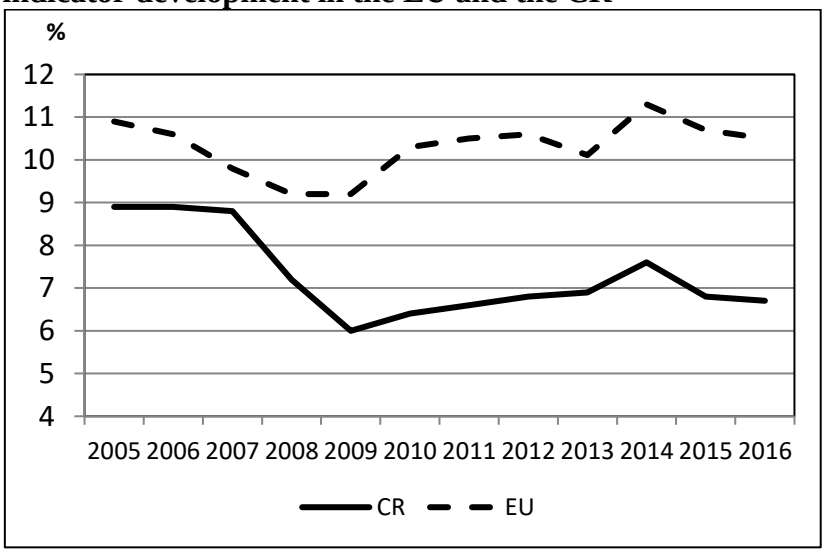
The analysis is based on 2005–2016 data of the Czech Republic taken from the EU-Statistics on Income and Living Conditions (EU-SILC) survey, the reference population being private households and their current members residing in the territory of the state. All the data as well as indicator definitions were adopted from the Eurostat database and calculations done using Excel and E-Views 9 statistical software package.

Figure 1 shows the development of the LWI indicator in the European Union and the Czech Republic from 2005 to 2016. As is evident from the figure, the proportion of people living in households with very low work intensity was decreasing steadily in the EU as well as in the Czech Republic over the period 2005–2009. The indicator hit its low in 2009 and the positive trend was reversed in the following years. The increase was largely attributed to the

economic crisis and subsequent recession, peaking in 2014. Since then, the indicator has been declining.

The values of the LWI indicator in the Czech Republic were lower than those in the European Union throughout the reporting period. Essential characteristics are displayed in Table 1. A significant difference between the level of the LWI in the EU and the CR was detected undertaking the standard two-sample t-test (t =8.6491, p = 0.0001) and non-parametric Wilcoxon rank-sum test (W = 3.0237, p = 0.0024).

**Fig. 1: LWI indicator development in the EU and the CR**




Source: Eurostat data, own elaboration

**Tab. 1: Proportion of people living in households with very low work intensity – EU and CR basic statistics**

| | Average. | Median | Variance | Standard deviation | Skewness | Kurtosis | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| *LWI_EU* | 10.308 | 10.5 | 0.408 | 0.639 | -0.6516 | -0.044 | 9.2 | 11.3 |
| *LWI_CR* | 7.3 | 6.85 | 1.043 | 1.021 | 0.797 | -0.820 | 6 | 8.9 |

Source: data Eurostat, own calculations

## 3 Regression analysis of LWI in relation to selected indicators in 2005–2016

The regression analysis of multivariate time series was carried out, the LWI representing the dependent variable. Other selected socio-economic indicators (such as the rates of employment, unemployment, employment growth, inflation, household savings, educational attainment, social benefits, etc.) were employed as explanatory variables; see Table 3.

Applying a unit root test (namely the Augmented Dickey-Fuller one) the initial analysis identified stationary and non-stationary time series, respectively. In Table 2, ADF test values and relevant p-values are presented.

**Tab. 2: Unit root tests of selected time series**

| Indicator | Abbreviation | $t_{ADF}$ | p-value | Stationarity/ non-stationarity |
|---|---|---|---|---|
| People living in households with very low work intensity | LWI | -1.2258 | 0.1885 | N |
| People living in households with very low work intensity in EU | LWI_EU | -1.9105 | 0.3160 | N |
| Real per capita GDP growth rate | GGDP | -2.0883 | 0.0375 | S |
| Inflation rate (HICP) | IR | -7.1295 | 0.0022 | S |
| Employment rate | ER | 0.9522 | 0.8948 | N |
| Employment growth | EG | -2.2283 | 0.0561 | N |
| Unemployment rate | UR | -1.2280 | 0.1721 | N |
| Lower secondary educational attainment | LSE | -1.4020 | 0.5422 | N |
| Household saving rate | HSR | -2.7565 | 0.0987 | N |
| Social benefits | SB | -1.6926 | 0.4078 | N |
| Lifelong learning | LLL | -1.5884 | 0.4349 | N |
| Young people neither in employment nor in education and training | YNE | -3.4720 | 0.0339 | S |

Source: Eurostat data, own calculations

Since the time series are to be of the same integrated process type, it can be concluded that there is no relationship between the LWI and indicators whose time series are stationary, such as GDP growth, inflation rate and the proportion of young people neither in employment nor in education and training.

Upon performing the unit root tests, non-stationarity of the LWI time series dependent variable was identified ($t_{ADF}$ = -1.2258, p =0.1885). Most of the selected explanatory variables are also non-stationary, thus being included in the regression analysis.

Many regression models with different combinations of explanatory variables were calculated. However, we present only the two best regression models satisfactory from a statistical point of view.

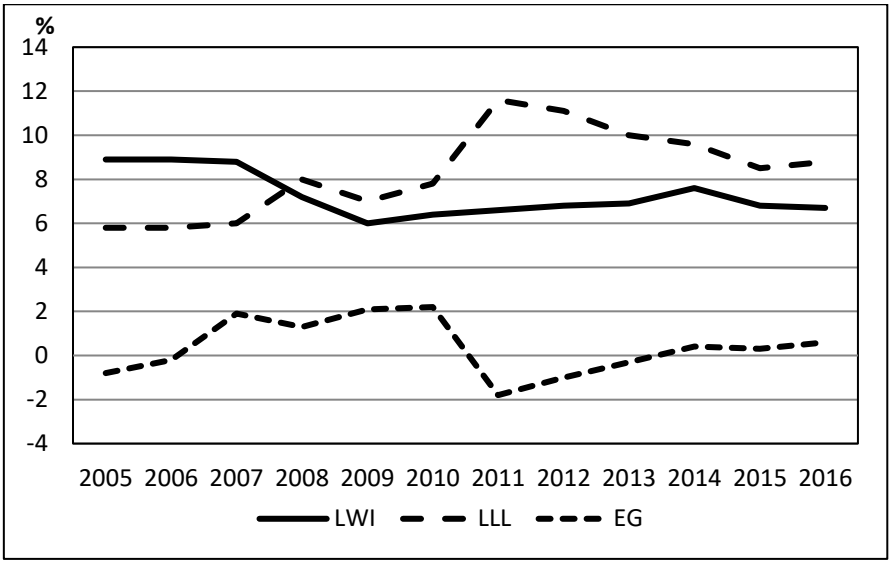
The first model contains LLL (lifelong learning) and EG (employment growth) variables; see Table 3. Owing to the non-stationarity of dependent and explanatory variables, the non-stationarity of a non-systematic component of their linear combination was tested using the ADF test. Because the AFD test rejects non-stationarity (t= -3.0352, p = 0.0061), the time series are co-integrated, the model expressing a long-term relationship. This regression model can be written in the form

$$\hat{LWI}_t = 11,5259 - 0,4836\ LLL_t - 0,5002\ EG_t \tag{3}$$

Thus, we can say that the relation of LLL and EG to the LWI is negative. An increase in the employment growth rate and proportion of people in lifelong education leads to the decline in the proportion of people living in households with very low work intensity.

The development of the dependent variable LWI and explanatory variables LLL and EG is plotted in Figure 2, the regression model being shown in Table 3.

**Fig. 2 Development of LLL, EG and LWI indicators in the Czech Republic, 2005–2016**




Source: Eurostat data, own elaboration

**Tab.3: *LWI ~ LLL + EG* regression model**

| Variable | Coefficient | Std. error | t-Statistic | Prob. |
|---|---|---|---|---|
| **C** | 11.5259 | 1.0181 | 11.3213 | 0.0000 |
| **LLL** | -0.4836 | 0.1155 | -4.1860 | 0.0024 |
| **EG** | -0.5002 | 0.1777 | -2.8136 | 0.0203 |

Source: Eurostat data, own calculations

The model explains 66.78 percent of time series LWI dynamics, its determination index being 0.6678 and the F-test significant (F = 9.0454, p = 0.0070). This model is acceptable from the statistical point of view – diagnostic tests (see Table 4) confirming that the non-systematic component has a normal distribution (Jarque-Bera test JB = 0.2550, p = 0.8803); it is homoscedastic [ARCH = 1.3312, p = 0.2783], not auto-correlated (Breuch-Godfray LM test F = 1.3549, p = 0.3181).

**Tab. 4: Diagnostic tests of model's *LWI ~ LLL + EG* non-systematic component**

| Test | Test statistic | Prob. | p-value |
|---|---|---|---|
| Breusch-Godfrey serial correlation LM test | 1.3549 | Prob. F(2.7) | 0.3181 |
| Normality test: Jarque-Bera | 0.2550 | Prob. | 0.8803 |
| Heteroscedasticity test: ARCH test | 1.3312 | Prob. F(1.9) | 0.2783 |

Source: Eurostat data, own calculations

Another regression model acceptable from all diagnostic tests is that with the explanatory variables SB (social benefits) and UR (unemployment rate). The development of the dependent variable LWI and explanatory variables SB and UR is plotted in Figure 3. This model is displayed in Table 5 and can be written as

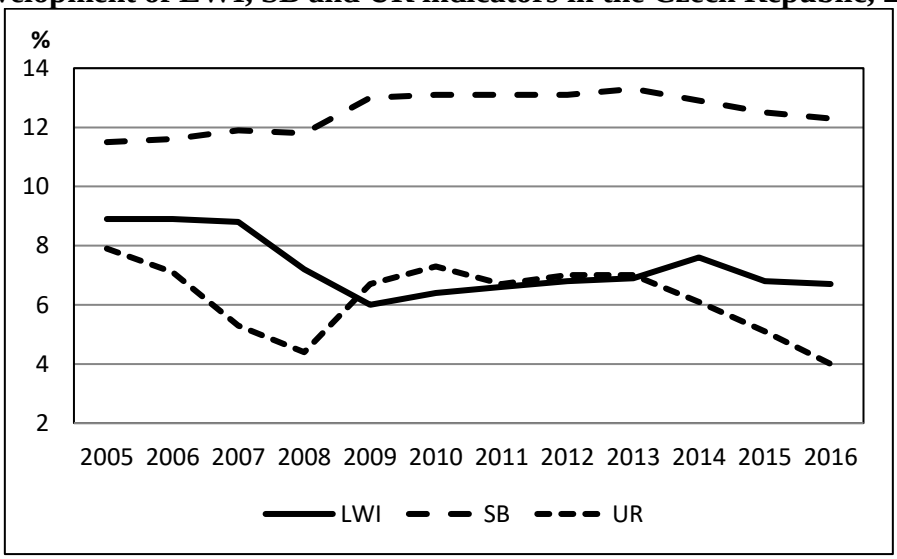
$$\hat{LWI}_t = 22{,}1951 - 1{,}3332\ SB_t + 0{,}2864\ UR_t, \qquad (4)$$

**Tab. 5: *LWI ~ SB + UR* regression model**

| Variable | Coefficient | Std. error | t-Statistic | Prob |
|---|---|---|---|---|
| C | 22.1951 | 3.3266 | 6.6719 | 0.0001 |
| SB | -1.3331 | 0.2722 | -4.8962 | 0.0009 |
| UR | 0.2864 | 0.1255 | 2.2822 | 0.0484 |

Source: Eurostat data, own calculations

**Fig. 3: Development of LWI, SB and UR indicators in the Czech Republic, 2005–2016**




Source: Eurostat data, own elaboration

The ADF test confirms the stationarity of the non-systematic component of model LWI ~ SB + UR (t = -2.5454, p=0.0164). This model also expresses the co-integration. The proportion of people living in households with very low work intensity rises with the growth of unemployment rates and declines with the growth in social benefits. The model explains 73.34 percent of the time series LWI dynamics, the F-test being significant (F = 12.3799, p = 0.0026). This model is acceptable from the statistical point of view – diagnostic tests confirming that the non-systematic component has a normal distribution (Jarque-Bera test t = 0.7637, p =0.6826); it is homoscedastic (ARCH = 0.1011; p = 0.7578), not auto-correlated (Breuch-Godfray LM test F = 0.9119, p = 0.4447).

## Conclusion

People living in households with very low work intensity are those aged 0-59 who live in households where the adults worked on average less than 20 percent of their total work potential during the last year.

The values of the LWI indicator in the Czech Republic were lower than those in the European Union over the reporting period 2005–2016. A significant difference between the levels of the LWI was verified using the standard two-sample t-test and non-parametric Wilcoxon rank-sum test.

The regression analysis of multivariate time series was conducted, the LWI representing the dependent variable. Carrying out the unit root tests, non-stationarity of dependent variable LWI time series was detected.

Since the time series are supposed to be of the same integrated process type, it can be concluded that there is no relation between the LWI and indicators whose time series are stationary such as the growth of GDP, inflation rate and the proportion of young people who are neither employed nor educated or trained.

Several regression models with different combinations of explanatory variables were calculated, the first one including explanatory variables LLL (lifelong learning) and EG (employment growth). There is a negative correlation between LLL, EG and the LWI. An increase in the employment growth and the proportion of people in lifelong education leads to a reduction of the proportion of people living in households with very low work intensity.

Another regression model acceptable from all diagnostic tests is that with explanatory variables SB (social benefits) and UR (unemployment rate). The proportion of people living in households with very low work intensity rises with the growth of unemployment rates and decreases with the growth in social benefits.

Both models presented are the cases of co-integration regression, showing long-term relationships.

## Acknowledgment

Paper was processed with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

## References

Arlt, J. (1998). Časové řady typu I(0) a I(1). Praha: *Acta oeconomica pragensia*, 6(2), pp. 7-11, Praha, ISSN 0572-3043.

Arlt,J., Arltová, M. (2009). *Ekonomické časové řady*. Praha: Professional Publishing. ISBN 978-80-247-1319-9.

Caner, M., Kilian, L. (2001). Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate. *Journal of International Money and Finance, 20*, pp. 639-657.

Dickey, D., Fuller,W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, pp. 427-431.

Dickey, D., Fuller, W. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica,* 49, 1057-1072.

Elliot, G., Rothenberg, J., Stock, J.H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64, 813-836

Hendry, D., Pagan, A., Sargan, J. (1984). *Dynamic Specifications*. Chapter 18 in Handbook of Econometrics, Vol.II (ed., Z. Griliches and M. Intriligator), North Holland

Hušek, R. (2007). *Ekonometrická analýza*. Praha: Oeconomica.

Jarque, C., Bera, A. (1980). Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals, *Economics Letters* 6, pp.255-259.

Phillips, P.C.B. (1987). Time Series Regression with a Unit Root. *Econometrica* 55, 277-301.

**Contact**

Dagmar Blatná
University of Economics, Prague
W. Churchill sq. 4
130 67 Prague 3
Czech Republic
blatna@vse.cz