ANALYSIS OF THE RELATIONSHIP BETWEEN THE SHARE OF EARLY LEAVERS FROM EDUCATION AND TRAINING AND SELECTED SOCIO-ECONOMIC INDICATORS IN THE CZECH REPUBLIC IN THE PERIOD 2005 - 2018

Dagmar Blatná

Abstract

The aim of this paper is to analyse the development of the "early leavers from education and training" indicator (ELE). The ELE indicator is based on the EU Labour Force Survey data. This indicator is defined as the percentage of the population aged 18-24 with at most lower secondary education and who were not involved in further education or training during the last four weeks preceding the EU Labour Force Survey. The ELE is one of the headline indicators being tracked within the main socio-economic EU strategy to 2020. The present article examines the ELE rate development in the Czech Republic compared to the EU between 2005 and 2018 in terms of meeting the objectives of the Europe 2020 strategy. The focus is on regression analysis, detecting regression relations between the time series of the ELE and selected socio-macroeconomic indicators in the Czech Republic, while applying theory of multidimensional analysis of non-stationary time series.

Key words: early school leavers, Europe 2020 strategy, time series, regression analysis, stationarity/non-stationarity

JEL Code: I28, I25, C22

Introduction

Early school leaving is a serious economic and social factor posing a risk to individuals and the society. Early dropouts are generally less employable people, their lack of qualifications resulting in precarious and low earnings, thus potentially increasing demands on the social security system. Moreover, they are at risk of social exclusion.

According to the Eurostat definition, the percentage of "early leavers from education and training" is shown by the ELE indicator, representing the population aged 18–24 with only lower secondary education or less who were not involved in any further education during

four weeks prior to the launch of the EU Labour Force Survey, which reports the ELE share. The International Standard Classification of Education categorizes this population cohort into groups of those who attained [0] less than primary, [1] primary and [2] lower secondary education.

In 2010, the European Commission devised the Europe 2020 strategy "for smart, sustainable and inclusive growth". The EC proposed five headline targets for the coming decade to be achieved by 2020, the ELE indicator monitoring educational attainment in relation to employability and access to the labour market. The strategic target for 2020 is to bring the EU dropout rate below 10 %, national target for the Czech Republic being 5.5 %. (see Eurostat (2013), Soriano and Mulatero (2010)).

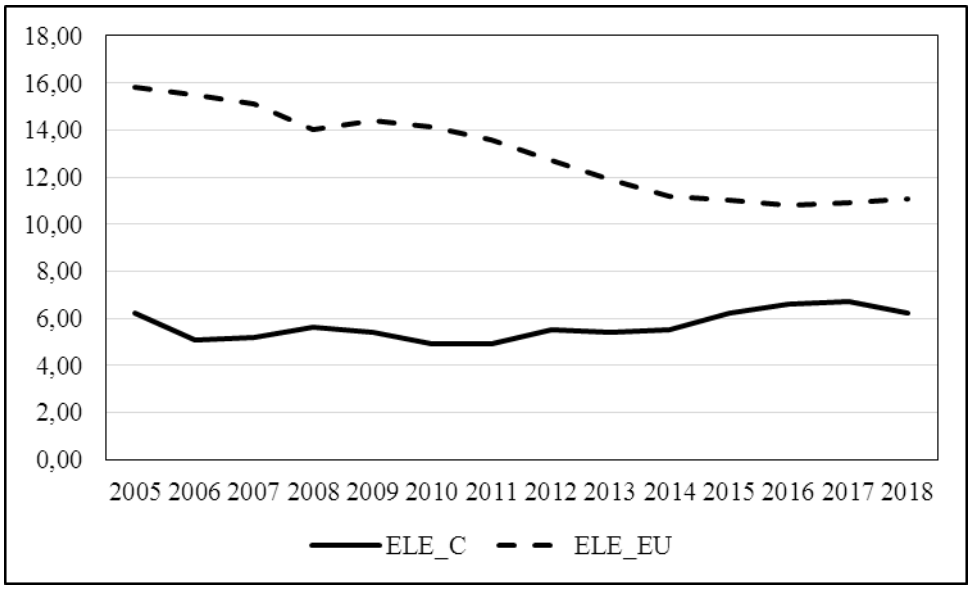
This paper presents the results of the ELE development analysis using Eurostat data for the Czech Republic and the European Union over the period between 2005 and 2018. Having applied the theory of multidimensional examination of non-stationary time series, the regression analysis was carried out, explaining regression relations between the time series of the ELE and the selected macroeconomic socio-economic indicators. Statistical calculations were done employing EViews 9, Statgraphics and Excel software packages.

1 ELE rate analysis for the EU and the Czech Republic (2005–2018)

The rate of early leaving from education and training shows some differences between the Czech Republic (ELE_C) and the European Union as a whole (ELE_EU) over time. As is obvious from Figure 1 below, the ELE in the former EU-28 member countries was decreasing steadily before the economic downturn, hitting its low in 2008. Having slightly increased in the crisis year 2009, the share of early school leavers, however, continued to decline until 2017, recent years showing a slight increase. The average annual decline in the ELE rate in the EU-28 was 2.67 percentage points throughout the research period. Assuming the 2005–2018 trend continues, the strategic 10% target by 2020 seems realistic.

The Czech Republic shows a different development, the proportion of early school dropouts being among the lowest in the EU. Throughout the monitored period, the ELE rate fluctuated between 4.9 % and 6.5 %, i.e. around the value assigned for the Czech Republic. Between 2009 and 2014, it was lower than the strategic goal, since 2015, however, the share of early leavers has been slightly increasing, thus slowly moving away from the national target.

Fig. 1: Early school leavers in the Czech Republic and the EU (2005–2018)



Source: Eurostat data, author's own elaboration

The ELE indicator values for the Czech Republic were below those of the EU-28 countries over the reporting period. Basic characteristics are displayed in Tab. 1. A significant difference between the ELE levels was verified by a standard two-sample t-test ($t = -14.1179$; $p = 0.000$) and non-parametric Wilcoxon rank-sum test ($W = 4.4794$; $p = 0.0000$).

Tab. 1: ELE basic statistics for the Czech Republic and the EU

Statistics	**ELE_C**	**ELE_EU**
Average.	5.6714	13.0071
Median	5.5	13.15
Standard deviation	0.6018	1.8487
Variance	0.3622	3.4176
Skewness	-1.0668	-1.6405
Kurtosis	0.4413	0.1309
Minimum	4.9	10.8
Maximum	6.7	15.8
Range	1.8	5.0
Coefficient of variation (%)	0.1061	0.1421
Jarque-Bera test of normality / p-value	1.0941/0.5786	1.3627/0.5059

Source: Eurostat data, author's own calculations

2 2005–2018 ELE rate regression analysis

2.1 Methodology

When using time series data to estimate the regression model, it is necessary to determine whether the variables are stationary or non-stationary. In the regression analysis, the time series of both variables must be integrated of the same order. Stationary time series are those with an autoregressive model of zero order I(0), the stationary process oscillating around a constant long-term mean, its constant variance being independent of time. Means, variances and covariances of non-stationary time series, on the other hand, change over time. These time series are generated by an autoregressive process of order one I(d).

For stationarity testing, the Augmented Dickey-Fuller (ADF) test is performed to verify 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 the model $y_t = \phi_1 y_{t-1} + a_t$, $S_{\hat{\phi}_1}$ is an estimate of $\hat{\phi}_1$ standard error, and a_t is a non-systematic component with white noise characteristics, i.e. the time series of uncorrelated random variables and probability distribution $N(0, \sigma_a^2)$ with zero mean and constant variance. For critical values of the test statistic, see Dickey and Fuller (1979). The null hypothesis of a unit root is rejected in favour of a stationary alternative if the test statistic is more negative than the critical value. For details, see Arlt and Arltová (2009), Hušek (2007), Caner and Kilian (2001), Dickey and Fuller (1981), Elliot *at al,* (1996) or Phillips (1987).

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 carried out, the null hypothesis assuming that both the skewness and excess kurtosis are zero (see, e.g., Jarque and Bera, 1980). Testing for the autocorrelation in the regression model errors, the Breusch-Godfrey Serial Correlation LM test was conducted with the null hypothesis of no serial correlation of any order up to p (cf. Breusch and Godfrey, 1986).

Autoregressive conditional homoscedasticity in the residuals was verified by the ARCH(1) LM test (Engle, 1982).

2.2 Early school leaver rate analysis: Czech Republic versus EU-28

First, it is necessary to determine whether the time series of both indicators are of the same integration order. Table 2 shows the extended Dickey-Fuller unit root test results for ELE_C and ELE_EU time series revealing that the former is non-stationary, while the latter is stationary, respectively.

Since the time series are to be of the same integrated process type, it can be concluded that there is no statistical relationship between the share of early school leavers in the EU and the Czech Republic in the period 2005–2018, the Pearson correlation coefficient not being allowed to be used for assessing the tightness of dependence between the two variables.

Tab. 2: ELE_C and ELE_EU unit root tests

Indicator	Abbreviation	t_{ADF}	p-value	Stationarity/ non-stationarity
Early leavers in the Czech Republic	ELE_C	-1.8323	0.8217	N
Early leavers in the EU	ELE_EU	-4.1786	0.0358	S

Source: Eurostat data, author's own calculations

2.3 Regression analysis: relationship between the early leaver rate and selected indicators in the Czech Republic (2005–2018)

The share of early leavers from education in the Czech Republic represents the dependent variable. The choice of socio-economic indicators employed as explanatory variables was limited by the availability and completeness of data for the whole reporting period. The Augmented Dickey-Fuller test identified stationary and non-stationary time series of the indicators used as explanatory variables for the ELE_C analysis. The ADF test values along with relevant p-values are provided in Table 3 below.

Since the time series are to be of the same integrated process type, it can be concluded that there is no relationship between the ELE_C and socio-economic indicators whose time series are stationary (such as the real GDP per capita growth rate, inflation rate, long-term unemployment rate, the share of young people in neither employment nor training and the share of households with broadband internet access).

Due to the non-stationarity of the dependent variable ELE_C, only non-stationary indicators can be included in the regression analysis as explanatory variables. Several regression models with different combinations of explanatory variables were created within the regression analysis for the ELE_C, only two statistically appropriate ones, however, being presented in this article.

Tab. 3: Unit root test of selected indicators for the Czech Republic

Indicator	Abbreviation	t_{ADF}	p-value	Stationarity/ non-stationarity
Real GDP per capita growth rate	GDPG_C	-2.0534	0.0425	S
Inflation rate (HICP)	IR_C	-4.9356	0.0129	S
Employment rate	ER_C	-1.1136	0.8810	N
Unemployment rate	UR_C	-1.6912	0.1004	N
Long-term unemployment rate	LUR_C	-2.1137	0.0374	S
Employment growth	EG_C	-1.8472	0.0635	N
Part-time employment rate	PTER_C	-0.0351	0.0067	S
Social benefits	SB_C	0.1239	0.7047	N
Young people neither in employment nor in training	YPN_C	-3.7949	0.0157	S
Employment rate of low skilled persons	ERLSP_C	-0.5434	0.8521	N
Households with broadband access	IA_C	-0.5149	0.0021	S
Lifelong learning	LL_C	-1.5025	0.7742	N
Average wage growth	AWG_C	-0.0037	0.6631	N
Minimum wage growth	MWG_C	0.3547	0.7706	N
People living in households with very low work intensity	LWI_C	-1.8352	0.5538	N

Source: Eurostat data, author's own calculations

The first statistically suitable regression model is displayed in Table 4, including only one – namely ER (employment rate) – explanatory variable. Figure 2 plots the development of the dependent ELE_C variable and the explanatory ER_C variable.

Tab. 4: ELE_C~ ER_C regression model

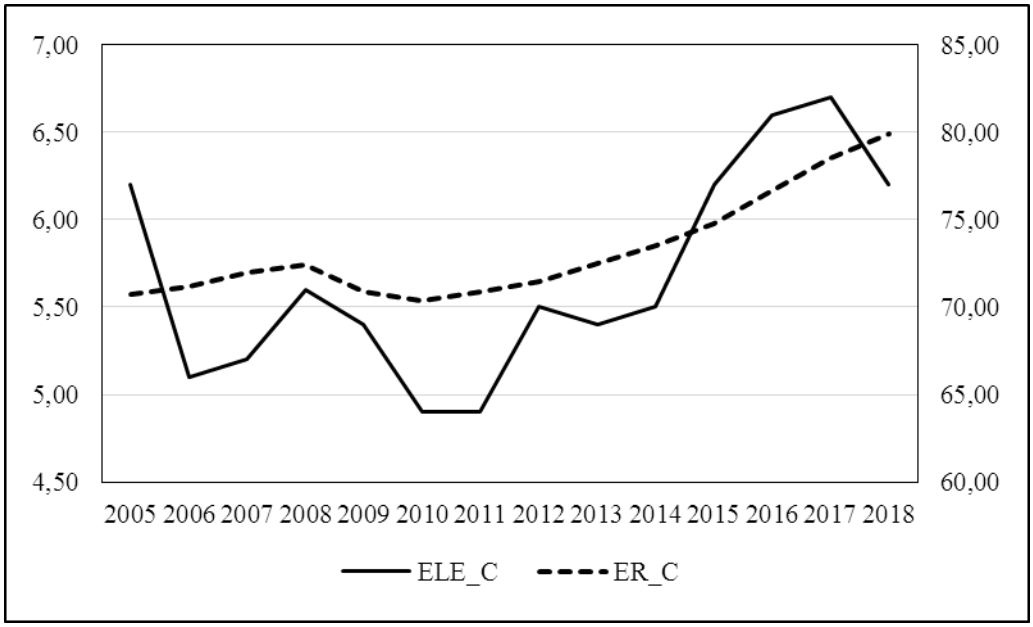
Variable	Coefficient	Std. error	t-statistic	p-value
C	-5.4603	2.6393	2.0688	0.0608
ER_C	0.1519	0.0359	4.2210	0.0012

Source: data Eurostat, author's own calculations

The share of early school dropouts in the Czech Republic grows with the increasing employment rate. The model explains 59.75 percent of the ELE_C time series dynamics, its determination index is 0.5975, the F-test being significant (F = 17.8171 p = 0.0012).

In terms of statistics, this model is acceptable, diagnostic tests (see Table 5) confirming that its non-systematic component is normally distributed (Jarque-Bera test JB = 1.9429, p = 0.3785), non-autocorrelated (Breuch-Godfray LM test F = 0.7318, p = 0.5051) and homoscedastic (ARCH = 0.0360, p = 0.8529). The ADF test confirms the stationarity of the model's non-systematic component (t_{ADF}= -3.1990; p = 0.0042). Because of the non-stationarity of all variables, the model implies a long-term relationship.

Fig. 2: Development of ELE_C and ER_C indicators for the Czech Republic (2005–2018)



Source: Eurostat data, author's own elaboration

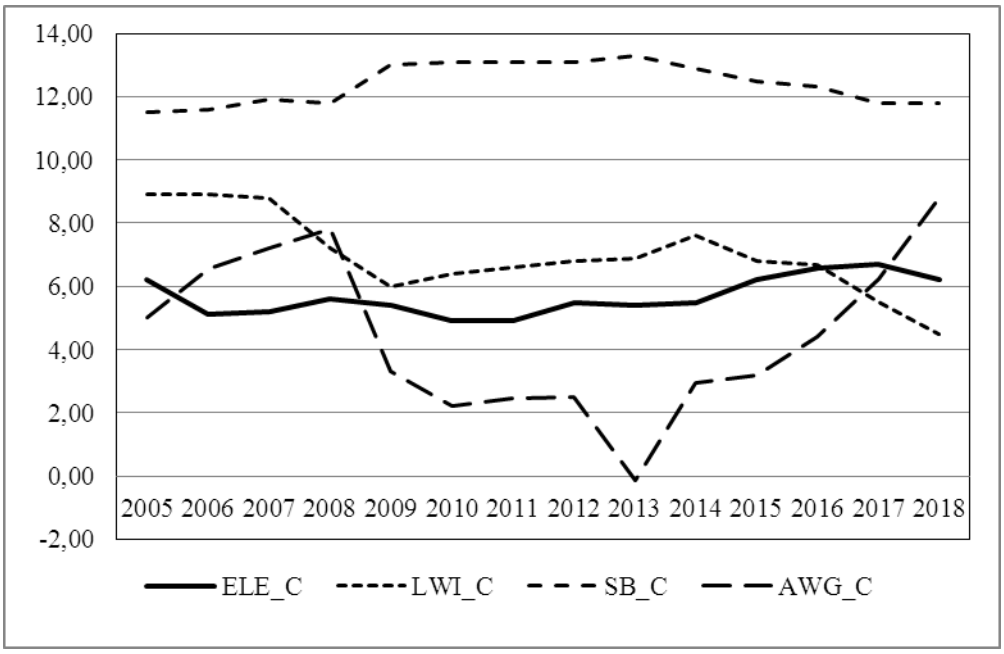
Tab. 5: Diagnostic tests of ELE_C~ ER_C model's non-systematic component

Test	Test statistic	Prob	p-value
Breusch-Godfrey serial correlation LM test	0.7318	Prob F(2,10)	0.5051
Normality test: Jarque-Bera	1.9429	Prob	0.3785
Heteroskedasticity test: ARCH test	0.0360	Prob F(1,11)	0.8529

Source: data Eurostat, author's own calculations

Another statistically relevant, diagnostically tested regression model is the one with explanatory variables LWI_C (people living in households with very low work intensity), SB_C (social benefits) and AWG_C (average wage growth). For the regression model, see Table 6 underneath. The development of the dependent variable ELE_C and explanatory variables LWI_C, SB_C and AWG_C, respectively, is illustrated in Figure 3.

Fig. 3: Development of ELE_C, LWI_C, SB_C and AWG_C indicators for the Czech Republic (2005–2018)



Source: Eurostat data, author's own elaboration

Tab. 6: ELE_C~ LWI_C+SB_C+AWG_C regression model

Variable	Coefficient	Standard error	t-statistic	Prob.
C	28.9703	5.9846	4.8408	0.0007
LWI_C	-0.3709	0.1026	-3.6143	0.0047
SB_C	-1.5688	0.4137	-3.7918	0.0035
AWG_C	-0.2794	0.1031	-2.7092	0.0220

Source: data Eurostat, author's own calculations

As you can see from Table 6, the ELE share in the Czech Republic declines with increasing SB (social benefits), LWI (people living in households with a very low intensity of work) and AWG (average wage growth), respectively. The model accounts for 67.53 percent of ELE_C time series dynamics, the F-test providing significant information ($F = 6.9324$, $p = 0.0083$). In terms of statistics, this model is acceptable, diagnostic tests proving that the non-systematic component is normally distributed (Jarque-Bera test $t = 0.2431$, $p = 0.8855$), homoscedastic (ARCH = 1.2581, $p = 0.2958$) and non-autocorrelated (Breuch-Godfray LM test $F = 0.2936$, $p = 0.7533$).

The ADF test confirms the stationarity of the non-systematic component ($t = -3.0987$, $p = 0.0048$). This model expresses cointegration, thus also suggesting a long-term relationship.

Tab. 7: Diagnostic tests of ELE_C~LWI_C+SB_C+AWG_C model's non-systematic component

Test	Test statistic	Prob.	p-value
Breusch-Godfrey serial correlation LM test	0.2936	Prob. F(2,8)	0.7533
Normality test: Jarque-Bera	0.2431		0.8855
Heteroscedasticity test: ARCH test	1.2581	F(1,11)	0.2958

Source: Eurostat data, author's own calculations

Conclusion

The early school leaving indicator is based on the EU Labour Force Survey data. It is defined as the percentage of the population aged 18 to 24 who had completed at most a lower secondary education and who were not involved in further education or training during four weeks prior to the updated EU-EFS.

The aims of this paper were to examine the development of the ELE indicator over the period 2005–2018 and find regression relations between the ELE and selected socio-macroeconomic data time series for the Czech Republic and the entire European Union. Some differences between ELE_C and ELE_EU rates over the monitored period were detected, the values of the ELE indicator for the Czech Republic being significantly lower than those for the EU-28 countries.

The Augmented Dickey-Fuller test identified the non-stationarity of the ELE_C indicator. As the time series are to be of the same integrated process type, the conclusion can be drawn that there is no relationship between the ELE_C and socio-economic indicators whose time series are stationary (the real GDP per capita growth rate, inflation rate, long-term unemployment rate, the share of young people neither in employment nor in training and the share of households with broadband internet access).

Several ELE_C regression models with different combinations of explanatory variables having been constructed, only two statistically suitable ones, however, were presented in this article. Both the models express cointegration, thus also suggesting a long-term relationship.

Acknowledgment

The paper was institutionally supported within the long-term research scheme of the Faculty of Informatics and Statistics, University of Economics, Prague.

References

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

Breusch, T., Goldfray, L.G. (1986). Data transformation Tests. *Economic Journal*, 96, pp. 47-58.

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

Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50,987-1007.

EUROSTAT (2013). *Smarter, Greener, More Inclusive? Indicators to Support the Europe 2020 Strategy*. European Union. ISBN 978-92-79-31156-7.

Hušek, R. (2007). *Ekonometrická analýza*. Praha: Oeconomica. ISBN 978-80-245-1300- 3.

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, pp. 277-301.

Soriano, F. H., F. Mulatero (2010). Knowledge Policy in the EU: From the Lisbon Strategy to Europe 2020. *Journal of the Knowledge Economy*, 1(4), s. 289–302. ISSN 1868-7865.

Contact

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