

## ASSESSMENT OF LIVING STANDARD INDICATORS IN THE OECD MEMBER COUNTRIES

**Diana Bílková**

University of Economics, Prague,  
Faculty of Informatics and Statistics, Department of Statistics and Probability,  
W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic  
e-mail: [diana.bilkova@vse.cz](mailto:diana.bilkova@vse.cz)

### Abstract

The research database for the present article consists of the OECD countries (except Turkey and Lithuania). The primary objective of the study is to group the countries according to twelve variables – average wage, minimum wage, GDP per capita, rates of unemployment, exchange and inflation rates, income tax, GDP per hour worked, indices of industrial, construction and manufacturing production and retail trade index, determining which of them significantly affect the average wage and defining the type and strength of such a relationship. The average wage, minimum wage and GDP per capita are used after their conversion into purchasing power parity, allowing for the comparison of price levels and PPP in different countries. Another important aim is to develop forecasts of the wage level for OECD countries by 2020. With regard to the countries' clustering according to the above criteria, Czechia always ranks alongside other post-communist countries (except Slovenia). The only explanatory variables affecting the average wage significantly are GDP per capita, income tax and manufacturing and retail trade indices, GDP exerting a major influence. Simple regression analysis of the dependence between the average wage and GDP per capita indicates that its course is best captured by the concave parabola with the peak at 77,252 PPP USD. The selected second-order polynomial regression explains approx. 89 percent of the variability of the observed values of the average annual wage. Wage growth by 2020 is expected in virtually all the OECD countries.

### Keywords

Average annual wage; GDP per capita; Cluster analysis; Ward method; Euclidean distance metric; Linear regression hyperplane; Quadratic regression function.

### Introduction

There is no doubt that the development of the average wage is related to that of the gross domestic product and other relevant indicators, such as income tax, total manufacturing production index and total retail trade index. In a period of GDP growth, the real wage growth is usually also expected. This, however, is not always the case, because GDP represents the total monetary value of goods and services provided over a given period in a particular country. Other factors, such as profit, interest rates and housing rents, come into play. Therefore, it is possible that GDP increases, but the average wage stagnates or even decreases – if, for instance, the profit has a larger share of GDP. Economic growth may also lead to an increase in the profits of companies which, however, do not share them with their employees. Many authors examine the link between wage behaviour and GDP and other labour market indicators. An obvious mismatch between per capita GDP development and real wages in pre-industrialized Europe is addressed, e.g. in [3]. It becomes apparent that the two indicators start

to change if there are any changes to the three following factors – income distribution, labour supply per capita and relative prices. Regarding the changes in GDP, the labour market adjustment mechanism was examined by [1], the results indicating that the adjustments lag behind the growth of GDP.

The importance of the subject matter researched and the necessity of research in the given area imply from above, as statistical analysis in this area can be an incentive for decision-making in the area of state budget and social policy. The results obtained can be a basis for decision-making of entrepreneurs when considering their sale opportunities, these knowledges may further help, for example, in various considerations regarding the level of tax burden. For the reasons described above, the economic situation of the developed world economies and the economies of post-communist countries has been widely addressed in professional literature. A significant part of the research works in this context is also concerned with the issue of regional economic inequalities, especially with regard to wages and incomes of the population. The authors [5] study the effects of inward foreign direct investment and trade on relative skilled labour demand in Poland, Hungary, and the Czech Republic. Their estimates show significant heterogeneity in the foreign direct investment effect across the three economies. The authors [8] present the complex picture of EU countries at the outset of the recent crisis (2007) in terms of the components of earnings differentials, with particular emphasis on the dimensions of labour market flexibility identifiable with contractual arrangements and self-employment. [9] studies the forecasting ability of various Phillips curve specifications for one year ahead headline and underlying core inflation measures for three open currently inflation targeting countries: Sweden, Canada and New Zealand. The authors [10] use the panel data approach to investigate the relationship between energy consumption and economic growth for V4 countries and for the 14 EU “old” Member States from 1995 to 2012. The authors [11] compare vulnerability to crises of the Czech Republic and Slovakia, which had operated as Czechoslovakia prior to 1993. The objective of the article [12] is to investigate whether a parallel process of tertiarisation has been taking place in the V4 countries. The researchers [13] examine the impact of maturity of economy on an activities’ focus of cluster organisations. Article [15] assesses whether the intensity of product market competition is a factor affecting economic growth (measured by the growth rate of real GDP per capita) and whether this impact depends on the model of capitalism. The study covers the 1997–2015 period and all EU 28 countries. The researchers [17] provide an empirical analysis of factors that might determine a stable economic growth in five countries (V4 countries and Romania). The researchers [18] study the determinants of firm resilience in the regions of Eastern Europe during the period 2007–2011 using a novel, dynamic, spatial and broad conceptual framework aspect.

In general, the reason why the average wage may be related to the unemployment rate is the fact that the reduction in the latter variable can affect wage growth due to the labour shortage in certain occupations, the prospects of collective bargaining thus being boosted.

The difference between the wage as the costs of labour and those of other production factors lies mainly in the exclusive nature of the workforce, resulting from the interplay of some indicators (economic, social, cultural) which determine the position of the labour force in the production process and its behaviour in the labour market. Wage level and development is the result of economic performance (expressed as GDP) and functioning of the labour market. At the same time, it is the basic determinant of the living standard of the population.

This research article aims to identify statistically significant indicators that affect wage levels in the OECD countries. The country grouping addresses the question of the existence of “different worlds” in the era of political convergence and economic globalization, post-

communist countries having integrated into OECD. The study also focuses on the prediction of wage levels in individual OECD countries by 2020.

The data for this research come from the official OECD website, see [19]. There are several research objectives that have been pursued. OECD member states (except Turkey and Lithuania due to unavailability of necessary data) were grouped into clusters of countries that are as close as possible to each other in terms of the following twelve variables for the year 2016 – average wage (gross), minimum wage (real), per capita GDP (all the above in constant prices in USD after PPP conversion), unemployment rate (in percentage terms), exchange rate (per USD, period average), inflation rate (annual CPI), income tax (pct. of labour costs for a childless person), GDP per hour worked, industry (industrial production index), construction (construction production index), manufacture (manufacturing production index) and retail trade (retail trade index); for all the above indices, 2010 = 100. The Dunn validation index being used to determine the optimal number of clusters, the OECD states were divided into seven groups. The Ward method with the Euclidean, squared Euclidean and city-block distances was employed for the construction of clusters. Using the Euclidean and city-block distance metrics, Czechia forms clusters together with five other post-communist countries, namely Estonia, Hungary, Latvia, Poland and Slovakia. Applying the squared Euclidean distance metric, Czechia constitutes a cluster along with the same countries plus Israel.

An important goal of this article is to investigate the dependence of the average wage on the other eleven variables for 2016, thus determining which of them statistically significantly affect the explained variable and indicating the type and strength of such a dependence. Normality of the variables was verified both visually and with the use of the Kolmogorov-Smirnov, Chi-Square and Shapiro-Wilks tests. Having employed simple correlation coefficients between the chosen explanatory variables, no problems with multicollinearity were identified. Inspecting visually and using the Glejser test, no heteroscedasticity was detected either. The suitability of the constructed model was verified by t- or general F- tests, determination coefficient and the Durbin-Watson test statistic. The regression hyperplane with eleven explanatory variables was considered in the first step. The methods of stepwise regression and forward selection were used to choose the appropriate set of explanatory variables (backward selection leading to the same selection of variables). It was found that only four explanatory variables – GDP per capita, income tax, manufacture and retail trade – statistically significantly affected the explained variable at a five percent significance level, GDP having the greatest impact. Therefore, a simple dependence of the average wage on per capita GDP was investigated. The concave regression parabola was chosen as the most appropriate model, allowing for the explanation – along with per capita GDP – of almost 89 percent of the variability of the average wage values observed.

Another important aim of this study was to make forecasts of the average wage for each country by the year 2020. The predictions were created analysing the relevant time series from 2000 to 2016. Exponential smoothing was used, the statistical software automatically evaluating the best combinations of equalizing constants. (The advantage of exponential smoothing lies in the fact that the latest observations are the most significant.) Appropriate exponential smoothing was selected using interpolation criteria. In all cases, sample residual autocorrelation and partial autocorrelation functions show that the non-systematic component does not indicate autocorrelation. Values of the Durbin-Watson statistic are close to the value two in all cases. Random failures can be therefore considered as independent. Model quality assessment was also performed using the Theil coefficient of non-compliance.

The research database for the present article consists of the OECD countries (except Turkey and Lithuania). Table 1 gives an overview of the 34 member countries along with their two-letter codes defined in ISO 3166-1 alpha-2.

**Tab. 1:** OECD country codes (ISO 3166-1 alpha-2)

Code	Country	Code	Country	Code	Country
AU	Australia	FR	France	LV	Latvia
AT	Austria	GB	Great Britain	MX	Mexico
BE	Belgium	GR	Greece	NL	Netherlands
CA	Canada	HU	Hungary	NO	Norway
CH	Switzerland	IE	Ireland	NZ	New Zealand
CL	Chile	IL	Israel	PL	Poland
CZ	Czechia	IS	Iceland	PT	Portugal
DE	Germany	IT	Italy	SE	Sweden
DK	Denmark	JP	Japan	SI	Slovenia
EE	Estonia	KR	South Korea	SK	Slovakia
ES	Spain	LU	Luxembourg	US	United States
FI	Finland				

Source: <http://ec.europa.eu/eurostat>

## 1 Theory and Methods

### 1.1 The Essence of Cluster Analysis

Cluster analysis was used to divide the OECD member states into relatively homogeneous groups according to the 2016 data on the twelve variables mentioned above. Multidimensional observations can be applied when classifying objects into several relatively homogeneous clusters. We have a data matrix  $X$  of  $n \times p$  type, where  $n$  is the number of objects and  $p$  the number of variables. Assuming various decompositions  $S^{(k)}$  of the set of  $n$  objects into  $k$  clusters, we look for the most appropriate ones. The aim is to identify the objects as similar as possible to each other within each cluster that are at the same time most different from those in other clusters, only decompositions with disjunctive clusters and tasks with a specified number of clusters being allowed.

#### 1.1.1 Decomposition Quality Assessment Criteria

The general task is to assess the level of achievement of the objective of cluster analysis in a given situation applying a specific algorithm. Several criteria (decomposition functions) were proposed, the most frequently used ones exhibiting the following characteristics. They are the matrices of internal cluster variance

$$E = \sum_{h=1}^k \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}}_h) \cdot (\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)' \quad (1)$$

and between-cluster variance

$$B = \sum_{h=1}^k n_h \cdot (\bar{\mathbf{x}}_h - \bar{\mathbf{x}}) \cdot (\bar{\mathbf{x}}_h - \bar{\mathbf{x}})' \quad (2)$$

whose sum is the matrix of total variance

$$T = \sum_{h=1}^k \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}}) \cdot (\mathbf{x}_{hi} - \bar{\mathbf{x}})' \quad (3)$$

There are vectors of the observations for the  $i^{\text{th}}$  object and  $h^{\text{th}}$  cluster  $\mathbf{x}_{hi}$ , the averages for the  $h^{\text{th}}$  cluster  $\bar{\mathbf{x}}_h$  and those for the total set  $\bar{\mathbf{x}}$ . There are  $p^{\text{th}}$ -membered vectors,  $E$ ,  $B$  and  $T$  being symmetric square matrices of the  $p^{\text{th}}$  order. The principal aim, consisting in the creation of the most distant compact clusters, is fulfilled by reaching the minimum of the total sum of the deviation squares of all values of corresponding cluster averages

$$C_1 = \text{st } \mathbf{E} = \sum_{h=1}^k \sum_{i=1}^{n_h} \sum_{j=1}^p (x_{hij} - \bar{x}_{hj})^2, \quad (4)$$

i.e. the Ward criterion<sup>1</sup>. Since the st  $\mathbf{T}$  is the same for all decompositions, the minimization of the st  $\mathbf{E}$  means the same as the maximization of the st  $\mathbf{B}$ . In order to become independent on the used units of measurement (or, more generally, the invariance to the linear transformations), it is recommended to minimize the determinant of the matrix of the internal cluster variance

$$C_2 = |\mathbf{E}|$$

or to maximize the trace criterion

$$C_3 = \text{st } (\mathbf{B}\mathbf{E}^{-1}) \text{ or else } C_4 = \text{st } (\mathbf{B}\mathbf{T}^{-1}).$$

The criteria mentioned above are employed not only retrospectively to assess the decomposition quality accomplished, changes in criterion values also guiding the creation of clusters. Since the criteria ultimately reach the limits ( $C_1$  and  $C_2$  the minimum,  $C_3$  and  $C_4$  the maximum) at  $k = n$ , it is necessary to find the extreme of the purpose function that properly includes the loss following from the growth in the number of clusters. The Ward criterion, for instance, is proposed to move towards the minimization of the quantity

$$Z_1 = C_1 + z \cdot k, \quad (5)$$

where constant  $z$  represents the loss resulting from an increase in the number of clusters by one.

### 1.1.2 Distance and Similarity of Objects

Having selected the variables characterizing the properties of the clustered objects and found their values, we decided on the method of the evaluation of distance or similarity of objects, the calculation of appropriate measures for all pairs of objects often being the initial stage of clustering algorithm implementation. The symmetric square matrix of  $n \times n$  type has zeros or ones on the diagonal, depending on whether it is the matrix of distance  $\mathbf{D}$  measures or that of similarity  $\mathbf{A}$  measures, respectively. Let us now focus on measuring the distance of the objects described by quantitative variables. The Hemming distance (also called Manhattan distance or city-block distance) can be used when individual variables are roughly on the same level or at least expressed in the same units of measurement

$$D_H(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^p |x_{ij} - x_{i'j}|. \quad (6)$$

The Euclidean distance can be applied in the same case

$$D_E(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\sum_{j=1}^p (x_{ij} - x_{i'j})^2} \quad (7)$$

as well as the Chebyshev distance

$$D_C(\mathbf{x}_i, \mathbf{x}_{i'}) = \max_j |x_{ij} - x_{i'j}|. \quad (8)$$

All the above mentioned measurements have some common drawbacks – the dependence on the used measuring units that sometimes hinders the meaningful acquisition of any sum for different variables and the fact that if the variables are considered in sum with the same weights, the strongly correlated variables have a disproportionately large effect on the

<sup>1</sup> Sign “st  $\mathbf{E}$ ” denotes the trace of matrix  $\mathbf{E}$ .

outcome. The starting point is the transformation of variables. The adverse effect of the measuring units can be removed by dividing all the values by the balancing factor, which can be presented with the corresponding average  $\bar{x}_j$ , standard deviation  $s_j$  or the range after deletion of extremes

$$\max_i x_{ij} - \min_i x_{ij}.$$

Particular variables can be also assigned more weight – having decided subjectively or on the basis of relevant information – their values then appearing in the formulas for the calculation of distance. Other measurements of distance and similarity of objects for numerical, ordinal, nominal and alternative variables are described in the professional literature. When dealing with variables of a different type, the Lance-Williams distance is recommended

$$D_{LW}(x_i, x_j) = \frac{\sum_{j=1}^p |x_{ij} - x_{i'j}|}{\sum_{j=1}^p (x_{ij} + x_{i'j})}. \quad (9)$$

Another possible way of expressing the relationship between two objects is the square Euclidean distance

$$D_{ES}(x_i, x_{i'}) = \sum_{j=1}^p (x_{ij} - x_{i'j})^2. \quad (10)$$

### 1.1.3 Algorithm for the Creation of Hierarchical Sequence of Decompositions

The creation of a hierarchical sequence of decompositions belongs to the most widely used techniques applied in the cluster analysis, occurring sequentially in the following steps:

- 1)  $D$  matrix calculation of appropriate measurements of distances;
- 2) the start of the decomposition process  $S^{(n)}$  from  $n$  clusters, each of them containing one object;
- 3) the assessment of the symmetric matrix  $D$  (a lower or upper triangle), finding two clusters (the  $h^{\text{th}}$  and  $h'^{\text{th}}$  ones) whose distance  $D_{hh'}$  is minimal;
- 4) the connection of the  $h^{\text{th}}$  and  $h'^{\text{th}}$  clusters into a new  $g^{\text{th}}$  cluster, the deletion of the  $h^{\text{th}}$  and  $h'^{\text{th}}$  row and column in the matrix  $D$  and their replacement with those of the new cluster, the order of the matrix being reduced by one;
- 5) recording the order of the cycle  $l = 1, 2, \dots, n - 1$ , the identification of the connected objects  $h, h'$  and the level of the connection  $d_l = D_{hh'}$ ;
- 6) returning to step (3) if the creation of decompositions has not been completed by connecting all objects into a single cluster  $S^{(1)}$ .

A divisive hierarchical procedure, contrary to the agglomerative hierarchical one, is less-used, starting from a single cluster  $S^{(1)}$ , splitting one of the clusters into two in each step and obtaining  $S^{(n)}$  at the end of the process. The results of hierarchical cluster procedures can be effectively displayed in the form of a graphical tree dendrogram. Given the choice of variables  $x_1, x_2, \dots, x_p$  and the matrix of distances  $D$ , the results of applying the described algorithm vary according to the way the distance between clusters is evaluated.

### 1.1.4 Nearest Neighbour Method

Within the nearest neighbour method, both clusters, whose connection is considered, are represented by objects that are the closest to each other. The  $D_{hh'}$  distance between the  $h^{\text{th}}$  and

$h^{\text{th}}$  clusters therefore represents the minimum of all  $q = n_h n_{h'}$  distances between their objects, the procedure of the third phase of the above algorithm thus being specified. In the fourth step, the  $h^{\text{th}}$  and  $h'^{\text{th}}$  row and column in the distance matrix are replaced with the row and column of distances of the new  $g^{\text{th}}$  cluster from other clusters. In the  $l^{\text{th}}$  cycle, total  $n - l - 1$  distances are determined by

$$D_{gg'} = \min (D_{g'h} \cdot D_{g'h'}) . \quad (11)$$

If the way of evaluation of the proximity or similarity of clusters is given, which also determines the conversion of the distance matrix in each cycle, the above algorithm allows for the creation of a hierarchical sequence of decompositions and construction of the dendrogram. When using this method, even considerably distant objects can get together in the same cluster if a large number of other objects create a kind of bridge between them. This typical chaining of objects is considered as a drawback, especially if there is a reason for the clusters to acquire the usual elliptical shape with a compact core. This method, however, possesses many positive features that outweigh the above disadvantage.

### 1.1.5 Farthest Neighbour Method

The method of the farthest neighbour is based on the opposite principle. The criterion for the connection of clusters is the maximum of  $q$  possible between-cluster distances of objects. When editing the matrix of distances, we proceed according to

$$D_{gg'} = \max (D_{g'h} \cdot D_{g'h'}) . \quad (12)$$

An adverse chain effect does not occur in this case. On the contrary, there is a tendency towards the formation of compact clusters, not extraordinarily large, though.

### 1.1.6 Average Linkage Method (Sokal-Sneath Method)

As a criterion for the connection of clusters, this method applies an average of the  $q$  possible between-cluster distances of objects. When recalculating the distance matrix, we use

$$D_{gg'} = \frac{n_h \cdot D_{g'h} + n_{h'} \cdot D_{g'h'}}{n_h + n_{h'}} . \quad (13)$$

The method often leads to similar results as the farthest neighbour one.

### 1.1.7 Centroid Method (Gower Method)

Unlike the above methods, this one is not based on summarizing the information on between-cluster distances of objects, the criterion being the Euclidean distance of centroid

$$D_E(\bar{x}_h, \bar{x}_{h'}) = \sum_{j=1}^k (\bar{x}_{hj} - \bar{x}_{h'j})^2 . \quad (14)$$

The recalculation of the distance matrix is done as follows

$$D_{gg'} = \frac{1}{n_h + n_{h'}} \left( n_h \cdot D_{g'h} + n_{h'} \cdot D_{g'h'} - \frac{n_h \cdot n_{h'}}{n_h + n_{h'}} \cdot D_{hh'} \right) . \quad (15)$$

### 1.1.8 Ward Method

The method uses a functional of the decomposition quality  $C_1$  in formula (4). The criterion for the cluster connection is an increment to the total intra-group sum of the squares of observation deviations from the cluster average, thus

$$\Delta C_1 = \sum_{i=1}^g \sum_{j=1}^p (x_{gij} - \bar{x}_{gj})^2 - \sum_{i=1}^h \sum_{j=1}^p (x_{hij} - \bar{x}_{hj})^2 - \sum_{i=1}^{h'} \sum_{j=1}^p (x_{h'ij} - \bar{x}_{h'j})^2 . \quad (16)$$

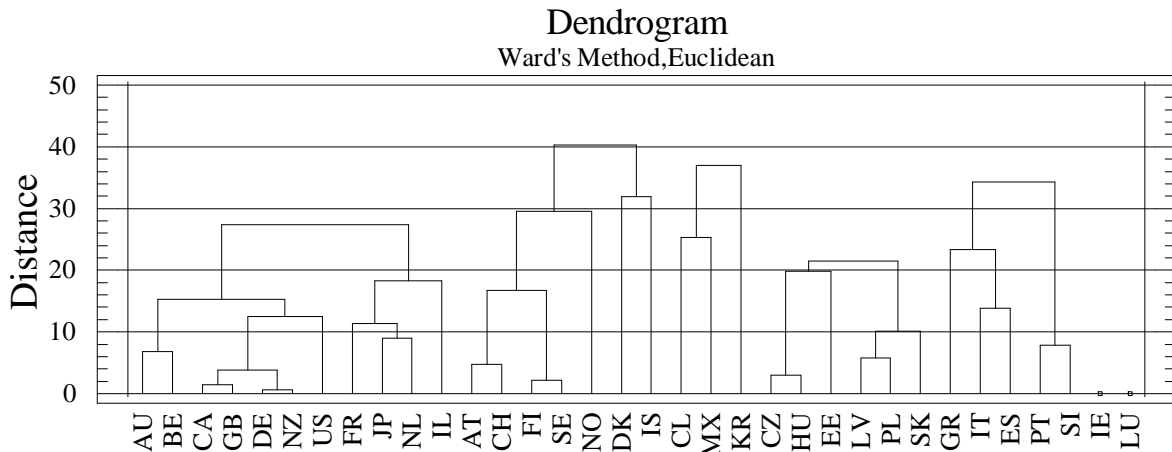
The increment is expressed as a sum of squares in an emerging cluster which is reduced by the sums of squares in both vanishing clusters. Using arithmetic modifications, the expression can be simplified into the form

$$\Delta C_1 = \frac{n_h \cdot n_{h'}}{n_h + n_{h'}} \cdot \sum_{j=1}^p (\bar{x}_{hj} - \bar{x}_{h'j})^2 . \quad (17)$$

This equation is a product of the Euclidean distance between the centroids of clusters considered for the connection and a coefficient depending on the cluster size. The value of this coefficient grows with an increasing size of clusters, and for fixed  $n_h + n_{h'}$  it represents the maximum in the case of the same-size ( $n_h = n_{h'}$ ) clusters. Since we create the connections to ensure the minimization of the criterion  $\Delta C_1$ , the Ward method tends to eliminate small clusters, i.e. to form those of roughly the same size, which is often a desirable property. Starting from the matrix of Euclidean distances between objects in the process of its modification, we can use the formula

$$D_{gg'} = \frac{1}{n_h + n_{h'} + n_{g'}} \cdot [(n_h + n_{g'}) \cdot D_{hg'} + (n_{h'} + n_{g'}) \cdot D_{h'g'} - n_{g'} \cdot D_{hh'}] . \quad (18)$$

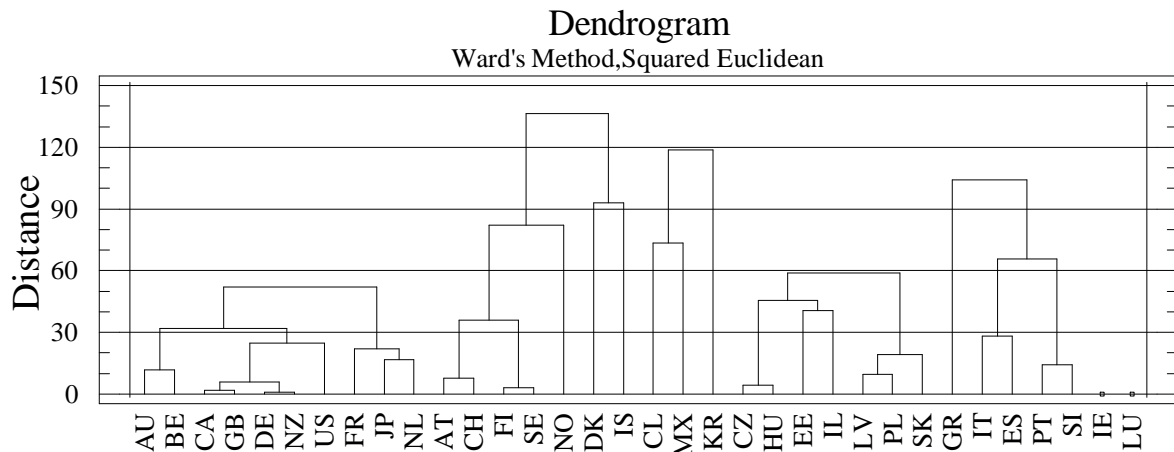
The essence of this multidimensional statistical method is explained in detail in [16]. In cluster analysis, there are different approaches and views on how to determine the optimal number of clusters, no definite conclusions being provided since cluster analysis is basically an exploratory method, not a statistical test. Commentary on and interpretation of the resulting hierarchical structure depends on the context, and theoretically there are often several possible solutions. Nevertheless, there are ways that help determine the optimal number of clusters, validation indices in particular. The proven Dunn index is one of them, representing the ratio of the smallest to the largest intra-cluster distance and taking values from zero to infinity; high index values indicating the optimal number of clusters.



Source: Own research; output from software

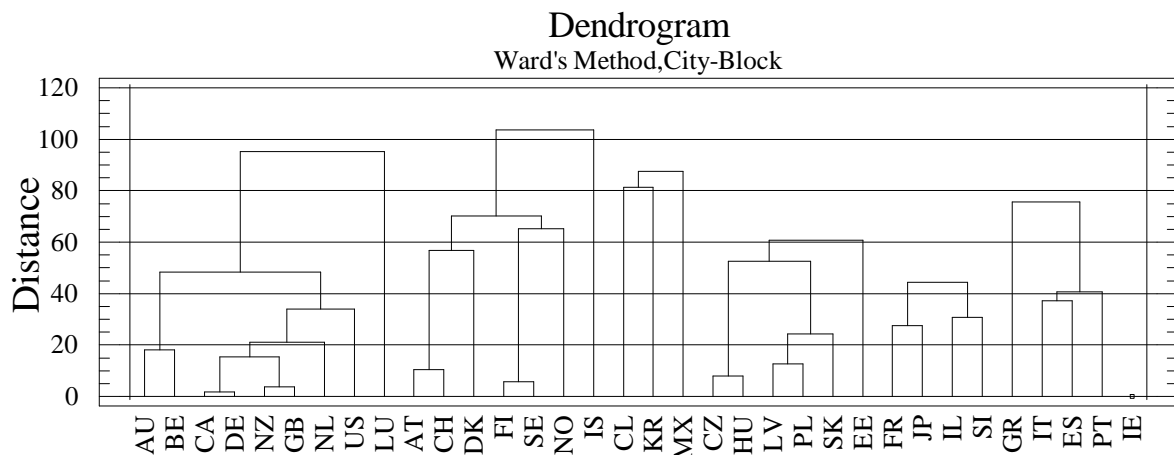
**Fig. 1:** Dendrogram of seven clusters (cluster analysis using Ward method and Euclidean distance)





Source: Own research; output from software

**Fig. 2:** Dendrogram of seven clusters (cluster analysis using Ward method and squared Euclidean distance)



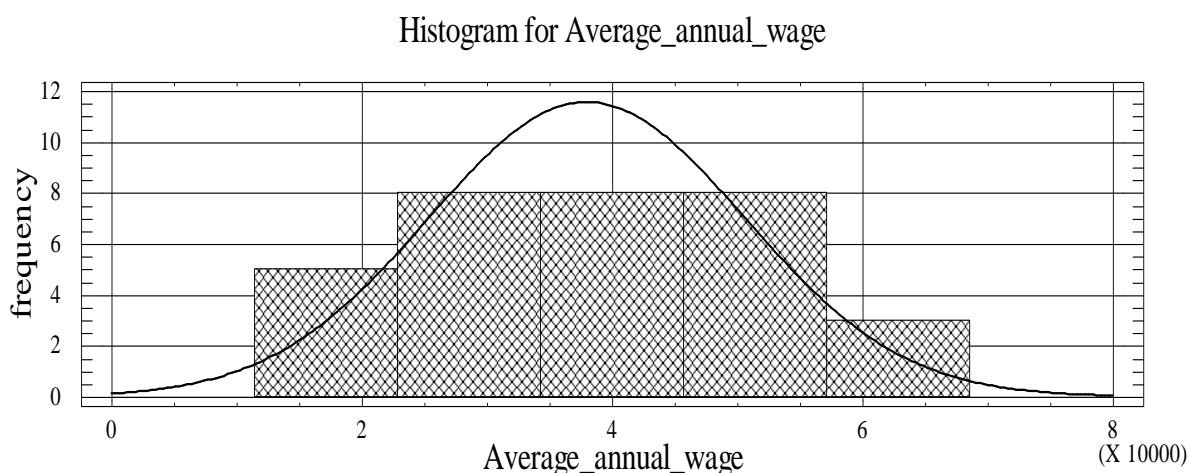
Source: Own research; output from software

**Fig. 3:** Dendrogram of seven clusters (cluster analysis using Ward method and city-block distance)

As mentioned above, the Ward method tends to remove too small clusters, so there is a tendency to create clusters of approximately the same size, which is a desirable feature in terms of clustering of the OECD countries. This is why the Ward method was utilized in the present clustering analysis. Since there is no need to strengthen the influence of any variable that might have an impact on the sum of squared distances (the points with the same distance from the centre lying on a circle), the Euclidean distance was chosen. It was completed with the square Euclidean and Hemming (city-block) distance metrics, allowing for a comparison of the results obtained. Figures 1–3 represent dendrograms of seven clusters analysed using the Ward method, Euclidean, squared Euclidean and city-block distances.

## 1.2 Regression and Correlation Analysis and Requirements for Methods Used

The regression and correlation analysis (cf., e.g. [7]) was carried out to examine data for the year 2016. The average wage represents an explained variable, the remaining eleven variables being explanatory ones.



Source: Own research; output from software

**Fig. 4:** Frequency histogram of average wage distribution normality (2016)

The normality of all variables was verified both visually and using Kolmogorov-Smirnov, Chi-Square and Shapiro-Wilks goodness-of-fit tests. Figure 4 and Table 2 illustrate the normality validation procedure for the average wage variable, which is then applied to the other eleven variables. Although the variable “wage” usually follows a lognormal distribution (with a positive skew), the “average wage” has a symmetrical distribution, which is an argument in favour of the normal distribution; see Figure 4. Considering the three goodness-of-fit tests, the smallest P-value is 0.0630961 for the Chi-Square test; see Table 2. This means that the null hypothesis, assuming the normality of the average wage distribution, cannot be rejected at a five percent level of significance. The normality of the remaining eleven variables was verified in the same way.

**Tab. 2:** Average wage normality assessment using Kolmogorov-Smirnov, Chi-Square and Shapiro-Wilks goodness-of-fit tests

Kolmogorov-Smirnov goodness-of-fit test Statistic: 0.119269 P-value: 0.718792
Chi-Square goodness-of-fit test Statistic: 21.5294 P-value: 0.0630961
Shapiro-Wilks goodness-of-fit test Statistic: 0.950289 P-value: 0.15913

Source: Own research; output from software

When constructing a regression model, a regression hyperplane was considered in the first step. The so-called stepwise regression and then forward selection were used to determine a set of explanatory variables that have a statistically significant effect on the explained variable (backward selection leading to the same results); see Table 3. It is obvious from the table that only four explanatory variables exert a statistically significant effect on the average wage at a five percent significance level. They are GDP per capita, income tax, manufacture and retail trade. It can be also seen in Table 3 that all t-tests and the general F-test are statistically significant at a five percent level of significance, the multiple determination coefficient indicating that 86.48 percent of the variability of the observed average wages were explained by the selected regression hyperplane. The Durbin-Watson statistic value is 1.83808, thus being close to the value 2 (in the interval 1.6, 2.4). This means that

autocorrelation does not pose any problem. Figure 5 shows the residues of the corresponding model with all four selected explanatory variables. The nature of these residues can be considered as accidental and therefore satisfactory. Apart from the visual assessment, the Glejser test was conducted, showing no heteroscedasticity present. For these reasons, we can see the selected regression hyperplane with four explanatory variables as satisfactory. Table 4 presents the matrix of simple correlation coefficients between the explanatory variables. None of these correlation coefficients' absolute values is higher than 0.5. Thus, harmful multicollinearity does not occur between the explanatory variables. The resulting regression hyperplane has the following form

$$\text{Average\_wage} = 7956.69 + 0.778627 * \text{GDP\_per\_capita} + 404.909 * \text{Income\_tax} - 183.614 * \text{Manufacture} + 154.08 * \text{Retail\_trade}.$$

Since GDP per capita was the first explanatory variable inserted into the model, we will still deal with a simple regression analysis of the dependence of the average wage on GDP per capita. Tables 5 and 6 show the results of a simple linear and quadratic regression analysis.

**Tab. 3:** Results of multidimensional linear regression analysis using stepwise regression and forward selection

Multiple regression analysis					
Dependent variable: Average_wage					
Parameter	Estimate		Standard error	T statistic	P-value
CONSTANT	7956.69		8684.74	0.91617	0.0071
GDP_per_capita	0.778627		0.0742635	10.4847	0.0000
Income_tax	404.909		148.992	2.71766	0.0110
Manufacture	-183.614		57.6366	-3.18572	0.0034
Retail_trade	154.08		56.8301	2.71125	0.0111
Analysis of variance					
Source	Sum of squares	D.f.	Mean square	F-ratio	P-value
Model	4.90372E9	4	1.22593E9	46.39	0.0000
Residual	7.66429E8	29	2.64286E7		
Total	5.67015E9	33			
(Corr.)					
R-squared = 86.4831 percent					
R-squared (adjusted for d.f.) = 84.6187 percent					
Standard error of est. = 5140.87					
Mean absolute error = 3949.32					
Durbin-Watson statistic = 1.83808					
Stepwise regression					
Method: forward selection					
F-to-enter: 4.0					
F-to-remove: 4.0					
Step 0:					
0 variables in the model		33 d.f. for error			
R-squared = 0.00%		Adjusted R-squared = 0.00%		MSE = 1.71823E8	
Step 1:					

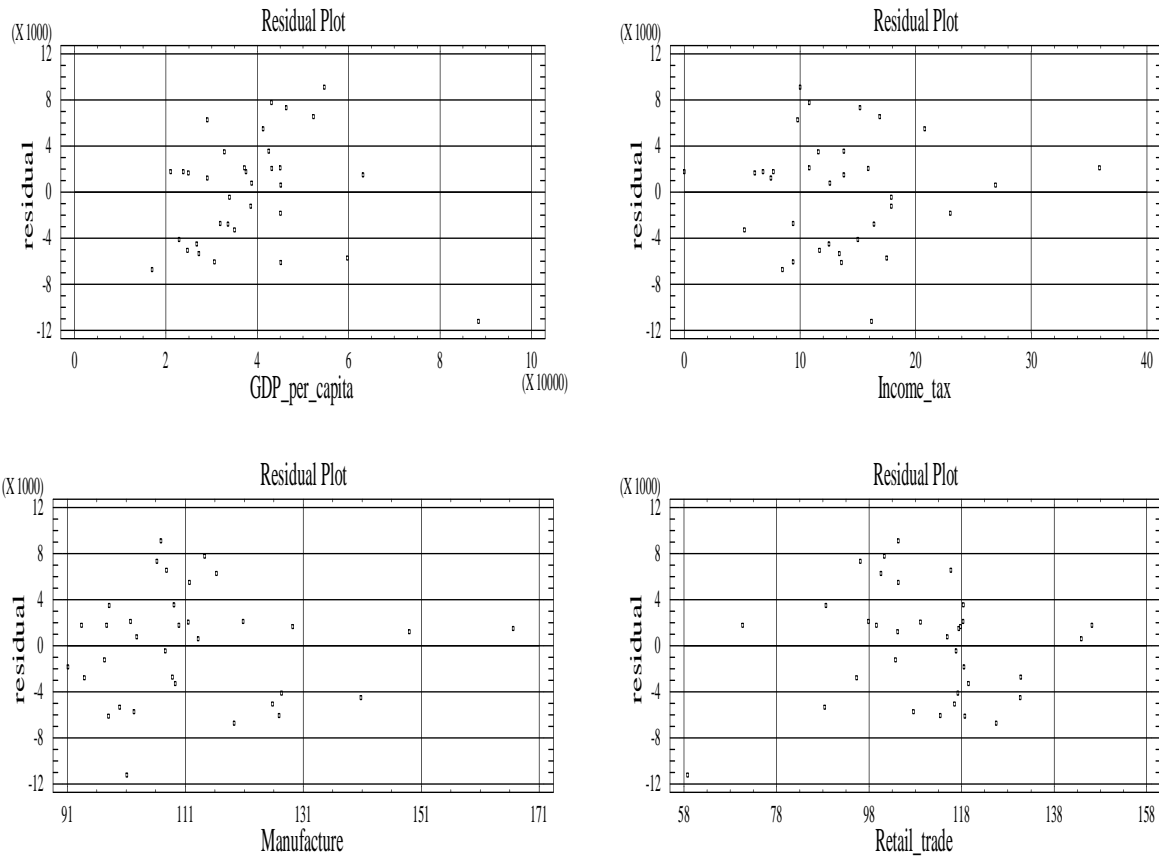
Adding variable GDP_per_capita with F-to-enter = 99.4004 1 variable in the model. 32 d.f. for error. R-squared = 75.65%      Adjusted R-squared = 74.89%      MSE = 4.31517E7
<b>Step 2:</b> Adding variable Income_tax with F-to-enter = 7.09553 2 variables in the model. 31 d.f. for error. R-squared = 80.18%      Adjusted R-squared = 78.90%      MSE = 3.62471E7
<b>Step 3:</b> Adding variable Manufacture with F-to-enter = 5.08877 3 variables in the model. 30 d. f. for error. R-squared = 83.06%      Adjusted R-squared = 81.36%      MSE = 3.20234E7
<b>Step 4:</b> Adding variable Retail_trade with F-to-enter = 7.35085 4 variables in the model. 29 d. f. for error. R-squared = 86.48%      Adjusted R-squared = 84.62%      MSE = 2.64286E7
Final model selected.

Source: Own research; output from software

**Tab. 4:** Matrix of simple correlation coefficients between explanatory variables

	<b>GDP_per_capita</b>	<b>Income_tax</b>	<b>Manufacture</b>	<b>Retail_trade</b>
<b>GDP_per_capita</b>	1.0000	-0.4060	-0.0485	0.3441
<b>Income_tax</b>	-0.4060	1.0000	0.0869	-0.0822
<b>Manufacture</b>	-0.0485	0.0869	1.0000	-0.3019
<b>Retail_trade</b>	0.3441	-0.0822	-0.3019	1.0000

Source: Own research; output from software



Source: Own research; output from software

**Fig. 5:** Residual course of linear model with four selected explanatory variables

It is obvious from the above tables that all t-tests and general F-tests are statistically significant at a five percent significance level and the Durbin-Watson statistic is near the value 2 for both regression dependencies, i.e. in the interval (1.6, 2.4). The adjusted determination index of the regression line is 74.89 percent, while that of the regression parabola reaches 87.82 percent. Figures 6 and 7 depict the course of the two dependencies considered and Figures 8 and 9 show corresponding residue graphs. The former figure indicates that in the case of a linear regression function, residues follow a non-random pattern. Regarding quadratic regression, the character of residues can be considered random and therefore satisfactory; see Figure 9. In addition to the visual assessment, the Glejser test was carried out, proving the absence of heteroscedasticity. Therefore, the second-order polynomial regression function can be considered as a more appropriate model of the average wage dependence on GDP per capita. The regression parabola is in the form

$$\text{Average\_wage} = -19286.0 + 2.13272 * \text{GDP\_per\_capita} - 0.0000138036 * \text{GDP\_per\_capita}^2.$$

**Tab. 5: Results of simple linear regression analysis**

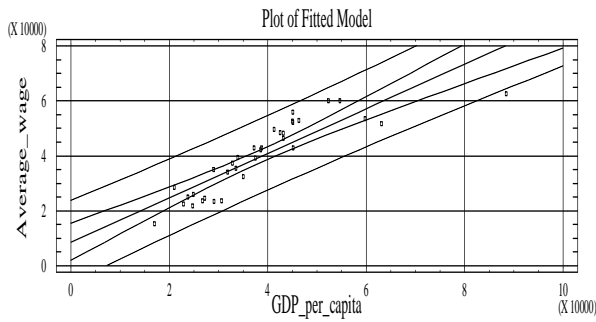
Linear regression analysis					
Dependent variable: Average_wage					
Parameter	Estimate	Standard error	T statistic	P-value	
CONSTANT	8573.63	3325.27	2.57833	0.0147	
GDP_per_capita	0.808932	0.0811368	9.96998	0.0000	
Analysis of variance					
Source	Sum of squares	D.f.	Mean square	F-ratio	P-value
Model	4.2893E9	1	4.2893E9	99.40	0.0000
Residual	1.38085E9	32	4.31517E7		
Total	5.67015E9	33			
(Corr.)					
R-squared = 75.6470 percent					
R-squared (adjusted for d.f.) = 74.8859 percent					
Standard error of est. = 6568.99					
Mean absolute error = 5347.66					
Durbin-Watson statistic = 1.74208					

Source: Own research; output from software

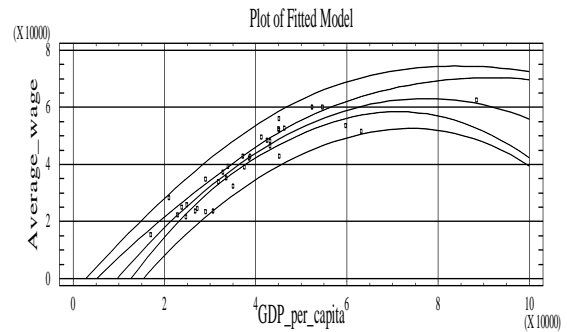
**Tab. 6: Results of simple quadratic regression analysis**

Polynomial regression analysis					
Dependent variable: Average_wage					
Parameter	Estimate	Standard error	T statistic	P-value	
CONSTANT	-19286.0	5247.08	-3.67556	0.0009	
GDP_per_capita	2.13272	0.230759	9.24218	0.0000	
GDP_per_capita^2	-0.0000138036	0.00000233299	-5.9167	0.0000	
Analysis of variance					
Source	Sum of squares	D.f.	Mean square	F-ratio	P-value
Model	5.02164E9	2	2.51082E9	120.02	0.0000
Residual	6.48511E8	31	2.09197E7		
Total	5.67015E9	33			
(Corr.)					
R-squared = 88.5627 percent					
R-squared (adjusted for d.f.) = 87.8248 percent					
Standard error of est. = 4573.81					
Mean absolute error = 3497.44					
Durbin-Watson statistic = 2.24407					

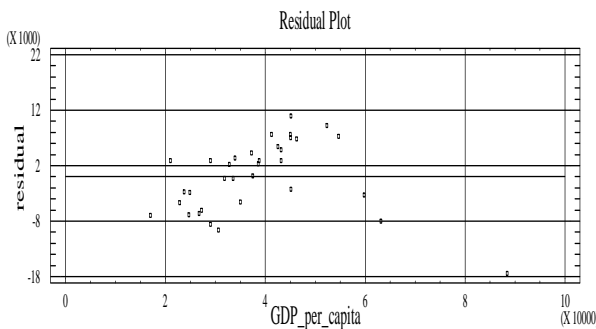
Source: Own research; output from software



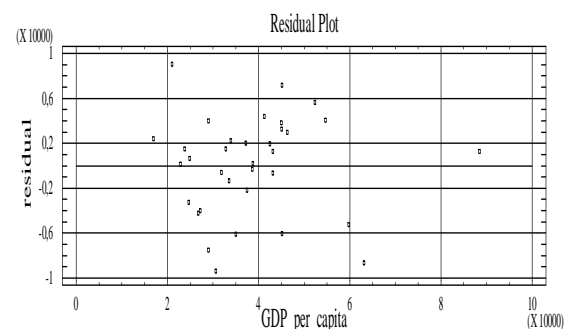
Source: Own research; output from software  
**Fig. 6:** Course of simple linear regression function



Source: Own research; output from software  
**Fig. 7:** Course of simple quadratic regression function



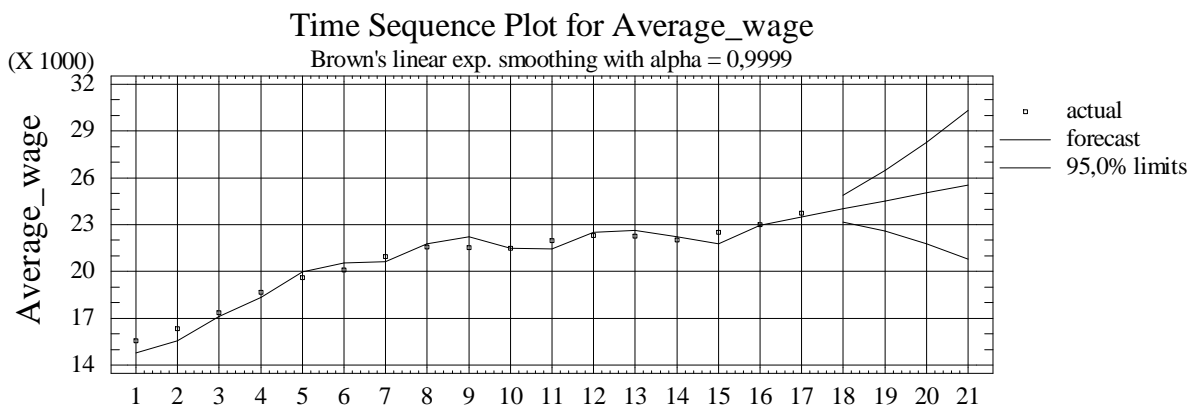
Source: Own research; output from software  
**Fig. 8:** Residual course – simple linear regression function



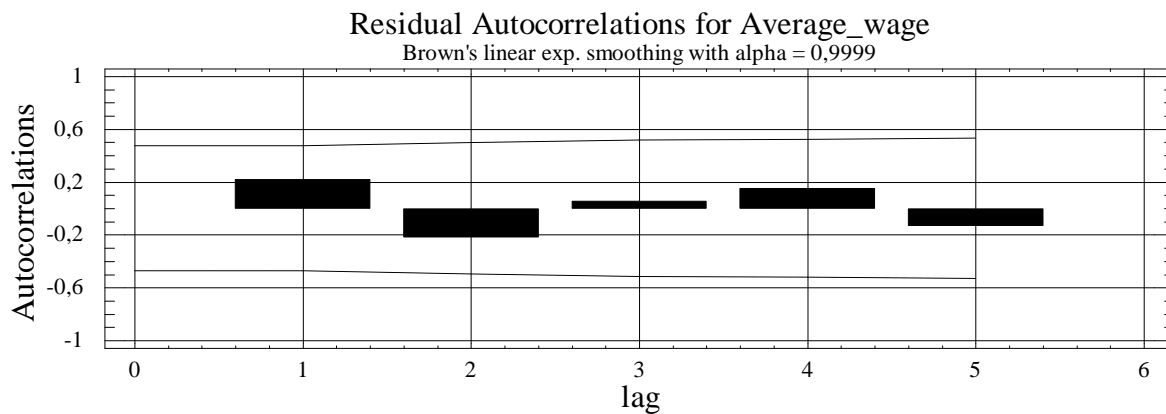
Source: Own research; output from software  
**Fig. 9:** Residual course – simple quadratic regression function

### 1.3 Time Series Analysis and Assumptions

Figures 10–12 represent Brown's linear exponential smoothing of the average wage in Czechia and residual (partial) autocorrelation functions, respectively, each of the figures illustrating an example of the procedure. (The essence of time series analysis is described in detail in [4].)

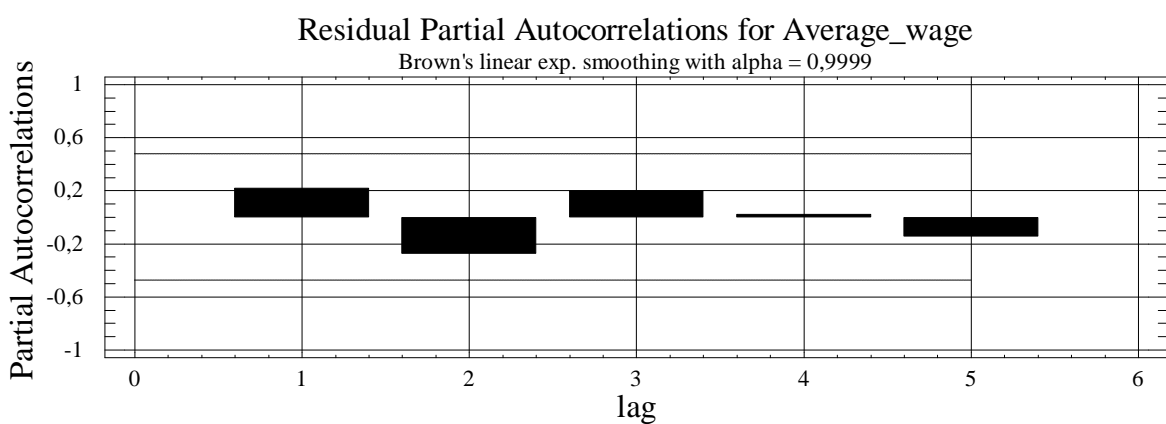


Source: Own research; output from software  
**Fig. 10:** Brown's linear exponential smoothing of average wage time series for Czechia ( $\alpha = 0.9999$ )



Source: Own research; output from software

**Fig. 11:** Residual autocorrelation function for time series of average wage in Czechia



Source: Own research; output from software

**Fig. 12:** Residual partial autocorrelation function for time series of average wage in Czechia

We focus on the average wage development in individual OECD member states over the period 2000–2016. Trend exponential smoothing was applied in time series analysis to construct average wage predictions by 2020. Exponential smoothing is one of adaptive approaches to modelling time series, employing the weighted least squares method, where the weights decrease exponentially towards the past, its advantage being that the most recent observations have the highest weights. Appropriate exponential alignment was selected using interpolation criteria.

As an illustration of the procedure used, Figure 10 shows the outcome of Brown's linear exponential smoothing which was evaluated as the most suitable for average wage time series. The autocorrelation function is a tool for assessing dependence between time series values, estimating correlation coefficients for each pair of equally distant time series values. Figure 11 provides a graphic image of a sample residual autocorrelation function (related to Brown's linear exponential smoothing of average wage time series for Czechia), individual bars representing the estimated correlation coefficients. In this figure, there is a dashed line at a distance of about twice the standard error estimate from zero for each estimated correlation coefficient, this limit being important for determining the significance of particular coefficients. Generally, the correlation coefficients are slightly different from zero if the bars do not overlap the dashed areas (top or bottom). On the other hand, the bars extending into these areas indicate the significance of the respective correlation coefficients. A partial autocorrelation function serves as a tool for assessing the dependence between the values of individual time series, estimating, however, partial correlation coefficients for each pair of values in the time series. Figure 12 displays the sample residual partial autocorrelation



function (related to Brown's linear exponential smoothing of time series of the Czech average wage), the bars representing the estimated partial correlation coefficients. Again, the dashed line plots the distance of approx. twice the standard error estimate from zero for each estimated partial correlation coefficient. This line is important for determining the significance of individual coefficients. Partial correlation coefficients are insignificantly different from zero if the bars do not go beyond the dashed boundaries. Bars which exceed them point to the significance of the respective partial correlation coefficients. It is clear from Figures 11 and 12 that the non-systematic component does not show autocorrelation and so Brown's linear exponential smoothing is satisfactory. The Durbin-Watson statistic value is close to 2, in the range 1.6–2.4. Random faults can be therefore considered as independent.

**Tab. 7:** Prediction errors in time series for Czechia

Year	Reality	Model	Estimate	Error
2000	15,569	14,789.0		
2001	16,349	15,568.8		
2002	17,335	17,128.8		
2003	18,659	18,321.0		
2004	19,592	19,982.9		
2005	20,105	20,525.1		
2006	20,939	20,618.1		
2007	21,584	21,772.9		
2008	21,535	22,229.0		
2009	21,496	21,486.1		
2010	21,996	21,457.0		
2011	22,314	22,495.9		
2012	22,266	22,632.0	22,995.8	729.8
2013	22,028	22,218.1	23,495.7	1467.7
2014	22,495	21,790.0	23,995.6	1500.6
2015	23,003	22,961.9	24,495.5	1492.5
2016	23,722	23,511.0	24,995.4	1273.4
	$T_H^2$		0.00340752	
	$T_H$		0.05837393	

Source: Own research

An illustrative example of the model quality evaluation procedure was performed for the average wage in Czechia (cf. Table 7) and subsequently applied to other countries. Based on these models, the average wage predictions for the next four years were made. Annual time series for the period between 2000 and 2016 were shortened to  $m = 5$  observations, i.e. the period 2012–2016, forecasts for these five years being made using Brown's linear exponential smoothing. The deviations between the predicted and actual values were calculated as

$$\Delta_t(i) = P_t(i) - y_{t+i}, \quad (19)$$

where  $P_t(i)$  is the prediction of the monitored indicator at time  $t$  of  $i$  time units (prediction horizon) shifted forward, and  $y_{t+i}$  is the real value of the predicted indicator at time  $t + i$ . These deviations are called predictive errors for a given time  $t$  and the prediction horizon  $i$ ; see Table 7. If  $\Delta_t(i) < 0$ , it is the so-called undervalued prediction, and if, on the other hand,  $\Delta_t(i) > 0$ , an overestimated prediction occurs.

The Theil index is a mismatch coefficient commonly measuring the variability of relative prediction errors

$$T_H^2 = \frac{\sum_{t=1}^m [P_t(i) - y_{t+i}]^2}{\sum_{t=1}^m y_{t+i}^2}. \quad (20)$$

It can only acquire non-negative values, hitting the lower zero boundary just in the case of a flawless prognosis when  $P_t(i) = y_{t+i}$ . The more the Theil mismatch coefficient deviates from zero, the more the prediction differs from an ideal flawless forecast. The square root of the Theil index can be interpreted as a relative prediction error.

In the construction of extrapolation forecasts of the average wage, we made an error averaging at 5.84 percent; see Table 7. The Theil mismatch coefficient and relative prediction error values indicate a high quality of the selected model of exponential smoothing. A similar verification of the suitability of the exponential smoothing type models was carried out for the other countries analysed.

## 2 Results and Discussion

The OECD brings together countries with the most advanced economies which produce more than two-thirds of the world's goods and services, meeting the challenges of economic globalization. Therefore, grouping of OECD countries by selected economic indicators, using the three distance matrices, offers useful insights; see Table 8.

The most advanced countries such as Australia, Belgium, Canada, Germany, the Netherlands, New Zealand, Great Britain and the United States always appear in the same group according to the twelve variables analysed. Scandinavian and other highly developed European countries (Austria and Switzerland) create another cluster. A separate group consists of three non-European countries, namely Chile, South Korea and Mexico. The Czech Republic is always part of a group of other post-communist countries except the more advanced Slovenia. South European countries facing economic problems in recent years constitute another group. Ireland, having also experienced a debt crisis, and Luxembourg always form “groups” of their own. The position of the latter is exceptional, its high GDP and, consequently, the level of wages relating to the small size of the country and the fact that about a third of the labour force commutes for work to Luxembourg from the neighbouring countries, thus not being included in the total population.

In terms of cartograms, this is clear that the Southern European countries (Greece, Italy, Portugal and Spain, designated often by their initial letters GIPS) always appear in the same cluster, as mentioned above. All these countries were primarily agrarian economies until the mid-20<sup>th</sup> century, and industrialization affected them only marginally. The vast majority of the population lived in rural areas and engaged in cultivation, breeding, fishing and related with this small-scale production.

The Scandinavian countries (Denmark, Finland, Norway, Sweden and Iceland) feature in the same cluster, too. These countries boast the most specific feature in the global context, which is relatively high spending on social systems. This mainly concerns the financing of the health system and old-age pensions. These countries have extraordinary results in a number of areas, such as economics, technology and sports. The economies of the Scandinavian countries are typical examples of a successful small open economy. The inevitable consequence of a strong economy is the high price level, which is extremely high in the Scandinavian countries compared to the rest of Europe. Agriculture, forestry and fisheries form the basis for a substantial part of the Scandinavian countries’ production. These countries export a large part of their agricultural products abroad: Norwegian salmon, Swedish wood or Danish salami and cheese are world-famous for their quality.

**Tab. 8:** Cluster analysis results

Euclidean distance			Squared Euclidean distance			City-block distance					
Cluster		Country	Cluster		Country	Cluster		Country			
1 <sup>st</sup>	1 <sup>st</sup>	Australia	1 <sup>st</sup>	1 <sup>st</sup>	Australia	1 <sup>st</sup>	1 <sup>st</sup>	Australia			
	2 <sup>nd</sup>	Belgium		2 <sup>nd</sup>	Belgium		2 <sup>nd</sup>	Belgium			
	3 <sup>rd</sup>	Canada		3 <sup>rd</sup>	Canada		3 <sup>rd</sup>	Canada			
	4 <sup>th</sup>	France		4 <sup>th</sup>	France		4 <sup>th</sup>	Germany			
	5 <sup>th</sup>	Germany		5 <sup>th</sup>	Germany		5 <sup>th</sup>	Luxembourg			
	6 <sup>th</sup>	Israel		6 <sup>th</sup>	Japan		6 <sup>th</sup>	Netherlands			
	7 <sup>th</sup>	Japan		7 <sup>th</sup>	Netherlands		7 <sup>th</sup>	New Zealand			
	8 <sup>th</sup>	Netherlands		8 <sup>th</sup>	New Zealand		8 <sup>th</sup>	Great Britain			
	9 <sup>th</sup>	New Zealand		9 <sup>th</sup>	Great Britain		9 <sup>th</sup>	United States			
	10 <sup>th</sup>	Great Britain		2 <sup>nd</sup>	10 <sup>th</sup>		United States	2 <sup>nd</sup>	1 <sup>st</sup>	Austria	
	11 <sup>th</sup>	United States			2 <sup>nd</sup>		Austria		2 <sup>nd</sup>	Denmark	
2 <sup>nd</sup>	1 <sup>st</sup>	Austria	2 <sup>nd</sup>	2 <sup>nd</sup>	Denmark	3 <sup>rd</sup>	3 <sup>rd</sup>		Finland		
	2 <sup>nd</sup>	Denmark		3 <sup>rd</sup>	Finland		4 <sup>th</sup>		Iceland		
	3 <sup>rd</sup>	Finland		4 <sup>th</sup>	Iceland		5 <sup>th</sup>		Norway		
	4 <sup>th</sup>	Iceland		5 <sup>th</sup>	Norway		6 <sup>th</sup>		Sweden		
	5 <sup>th</sup>	Norway		6 <sup>th</sup>	Sweden		7 <sup>th</sup>		Switzerland		
	6 <sup>th</sup>	Sweden		3 <sup>rd</sup>	1 <sup>st</sup>		Chile		4 <sup>th</sup>	1 <sup>st</sup>	Chile
	7 <sup>th</sup>	Switzerland			2 <sup>nd</sup>		South Korea			2 <sup>nd</sup>	South Korea
3 <sup>rd</sup>	1 <sup>st</sup>	Chile	3 <sup>rd</sup>	3 <sup>rd</sup>	Mexico	5 <sup>th</sup>	3 <sup>rd</sup>		Mexico		
	2 <sup>nd</sup>	South Korea		4 <sup>th</sup>	1 <sup>st</sup>		Czechia		6 <sup>th</sup>	1 <sup>st</sup>	Czechia
	3 <sup>rd</sup>	Mexico			2 <sup>nd</sup>		Estonia	2 <sup>nd</sup>		Estonia	
4 <sup>th</sup>	1 <sup>st</sup>	Czechia	4 <sup>th</sup>	3 <sup>rd</sup>	Hungary	6 <sup>th</sup>	3 <sup>rd</sup>	Hungary			
	2 <sup>nd</sup>	Estonia		4 <sup>th</sup>	Israel		4 <sup>th</sup>	Latvia			
	3 <sup>rd</sup>	Hungary		5 <sup>th</sup>	Latvia		5 <sup>th</sup>	Poland			
	4 <sup>th</sup>	Latvia		6 <sup>th</sup>	Poland		5 <sup>th</sup>	1 <sup>st</sup>	France		
	5 <sup>th</sup>	Poland		7 <sup>th</sup>	Slovakia			2 <sup>nd</sup>	Israel		
	6 <sup>th</sup>	Slovakia		5 <sup>th</sup>	1 <sup>st</sup>		Greece	6 <sup>th</sup>	3 <sup>rd</sup>	Japan	
5 <sup>th</sup>	1 <sup>st</sup>	Greece	2 <sup>nd</sup>		Italy	4 <sup>th</sup>	Slovenia				
	2 <sup>nd</sup>	Italy	3 <sup>rd</sup>		Portugal	1 <sup>st</sup>	Greece				
	3 <sup>rd</sup>	Portugal	4 <sup>th</sup>		Slovenia		2 <sup>nd</sup>		Italy		
	4 <sup>th</sup>	Slovenia	5 <sup>th</sup>		Spain		3 <sup>rd</sup>		Portugal		
	5 <sup>th</sup>	Spain	6 <sup>th</sup>	1 <sup>st</sup>	Ireland		4 <sup>th</sup>	Spain			
6 <sup>th</sup>	Ireland	7 <sup>th</sup>		Luxembourg	7 <sup>th</sup>		1 <sup>st</sup>	Ireland			
7 <sup>th</sup>	Luxembourg	7 <sup>th</sup>	1 <sup>st</sup>	Luxembourg							

Source: Own research

The Baltic countries, i.e. Estonia and Latvia as member countries of the former Soviet Union (Lithuania was not included in the research due to insufficient data), together with other Central European post-communist countries (Czechia, Slovakia, Poland and Hungary) are always in the same cluster. There are the countries where the transformation of the economy from a centrally planned economy to a market economy took place at the end of the 20<sup>th</sup> century.

American countries like Chile and Mexico, together with South Korea, always form a separate cluster, too. Mexico is the third largest economy on the American continent (after the United States and Brazil). This includes a combination of services, industry and agricultural production and it is based primarily on the export of raw materials and products. The Chilean

economy is one of the most successful in South America. Less labour productivity remains a challenge for the Chilean economy. The government seeks to solve it through reforms aimed at education and modernization, labour law reforms and cutting bureaucracy. An important part of the Chilean economy is the extraction of minerals, predominantly copper. South Korea is the 11<sup>th</sup> largest economy in the world and the 4th largest economy in Asia. There is a post-industrial economy, the service sector occupies 57.5% of the economy, industry 39.8% and agriculture 2.7%. The South Korean science and technology industry is one of the most advanced in the world. South Korea has a highly digitized economy and is developing a national cyber security strategy in the context of hacker attacks.

Australia, Belgium, Canada, Germany, Netherlands, New Zealand, Great Britain and United States remain in the same cluster, too. Only in terms of City-block distance, Japan and France separate from this group and form a separate cluster with Israel and Slovenia. All they are the most advanced countries in the world.

Only four out of the eleven explanatory variables considered (GDP per capita, income tax, manufacture and retail trade) affect significantly the explained variable (average wage) at a five percent significance level – three of them positively and one (manufacture) in a negative way. Sample regression coefficients (in the “Estimate” column, cf. Tab. 3) indicate the change in the average wage if the value of the corresponding explanatory variable increases by one unit, provided the values of the other explanatory variables remain unchanged. When examining the double dependence of the average wage on GDP per capita, the concave parabola with the peak at 77,252 USD PPP of per capita GDP best describes the dependence of the average wage on GDP per capita. This means that the average wage shows on average a parabolic growth along with GDP per capita rising to 77,252 USD PPP of GDP per capita, and then it starts falling parabolically with rising GDP per capita.

Regarding the predictions of the average wage, Table 9 provides the predicted values for the seven selected countries representing each cluster. It is clear from the table that we can expect a marked rise in Slovenia and the countries of the same cluster by 2020. Only gradual wage growth is likely to appear in other OECD countries, especially in Mexico, Chile and South Korea. Wage level decline is not predicted in any OECD member country over the next years.

**Tab. 9:** Average wage prediction for selected countries from each cluster

Year	Country						
	United States	Norway	Mexico	Czechia	Slovenia	Ireland	Luxembourg
2017	60,207	54,143	15,423	24,019	35,776	51,411	63,042
2018	60,678	54,642	15,504	24,527	36,587	51,732	63,502
2019	61,150	55,142	15,585	25,035	37,398	52,053	63,963
2020	61,622	55,641	15,665	25,543	38,210	52,375	64,423

Source: Own research

## Conclusion

In terms of the twelve variables considered, Luxembourg and Ireland have a unique position. The former country (a small inland one, contrary to the latter) reports the highest annual PPP-based GDP per capita. The reason is that about a third of the workforce consists of foreign nationals commuting from neighbouring countries who are not included in the total population of Luxembourg. For example, the authors [2] show that Luxembourg stands apart as one of the richest regions in Europe and that disposable incomes in Luxembourg are some 70% higher than in the neighbouring regions. The cluster analysis shows that Czechia is always grouped along with other post-communist OECD countries, except for Slovenia. The

study [14] shows that the most important economic growth determinants in the Central and Eastern European countries are investment rate, human capital measured by the education level of the labour force, financial sector development, good fiscal stance (low budget deficit and low public debt), economic structure (high services share in GDP), low interest rates and low inflation, population structure (high share of working-age population), development of information technology and communications, high private sector share in GDP and favourable institutional environment (economic freedom, progress in market and structural reforms). Greece and Spain, for example, are clustered together in groups of countries encountering economic difficulties. The research [6] examines the political economies of Greece and Spain and how they affected the industrial policies adopted in either country during the 1980s and 1990s. This analysis suggests that historical legacies, political institutions and industry structure go a long way towards explaining the different industrial policies pursued by either country during the 1980s.

The analysis shows that a gradual, mostly modest increase in the average annual wage in almost all OECD countries can be expected by 2020. This is in line with OECD economic forecasts. Trade and private investment extension has restarted job creation and inflation will grow only slightly. However, new tensions and threats may derail the recovery. The economic outlook highlights a range of policies that can help maintain medium-term growth and ensure that its benefits are widely shared. Wage growth is expected to support household consumption, relatively low interest rates allowing for further capital investment. Overall economic growth will also alleviate the labour shortage.

In terms of indication for orientation of the future research, differences in the quantitative aspect of the standard of living between the G7 countries (the seven most economically advanced countries in the world) on the one hand and the V4 countries (countries of the Viszegrad Four: Czechia, Slovakia, Hungary and Poland) on the other hand will be examined.

### **Acknowledgements**

This article was subsidized by the funds of institutional support of a long-term conceptual advancement of science and research number IP400040 at the Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic.

### **Literature**

- [1] AKKEMIK, K. A.: The Response of Employment to GDP Growth in Turkey: An Econometric Estimation. *Applied Econometrics and International Development*. 2007, Vol. 7, Issue 1, pp. 65–74. ISSN 1578-4487.
- [2] ALLEGREZZA, S.; HEINRICH, G.; JESUIT, D.: Poverty and Income Inequality in Luxembourg and the Grande Région in Comparative Perspective. *Socio-Economic Review*. 2004, Vol. 2, Issue 2, pp. 263–283. ISSN 1475-1461. DOI: [10.1093/soceco/2.2.263](https://doi.org/10.1093/soceco/2.2.263)
- [3] ANGELES, L.: GDP per Capita or Real Wages? Making Sense of Conflicting Views on Pre-industrial Europe. *Explorations in Economic History*. 2008, Vol. 45, Issue 2, pp. 147–163. ISSN 0014-4983.
- [4] BROCKWELL, P. J.; DAVIS, R. A.: *Introduction to Time Series and Forecasting*. Springer, New York, 2016. ISBN 978-3-319-29852-8.
- [5] BRUNO, G. S. F.; CRINÒ, R.; FALZONI, A. M.: Foreign Direct Investment, Trade, and Skilled Labour Demand in Eastern Europe. *Labour*. 2012, Vol. 26, Issue 4, pp. 492–513. ISSN 1467-9914. DOI: [10.1111/labr.12001](https://doi.org/10.1111/labr.12001)

- [6] CALOGHIROU, Y.; VOULGARIS, Y.; ZAMBARLOUKOS, S: The Political Economy of Industrial Restructuring: Comparing Greece and Spain. *South European Society and Politics*. 2000, Vol. 5, Issue 1, pp. 73–96. ISSN 1360-8746.
- [7] DARLINGTON, R. B.; HAYES, A F.: *Regression Analysis and Linear Models: Concepts, Applications, and Implementation*. The Guilford Press, New York, 2016. ISBN 978-1-4625-2113-5.
- [8] HÖLSCHER, J.; PERUGINI, C.; POMPEI, F.: Wage Inequality, Labour Market Flexibility and Duality in Eastern and Western Europe. *Post-Communist Economies*. 2011, Vol. 23, Issue 3, pp. 271–310. ISSN 1463-1377. DOI: [10.1080/14631377.2011.595119](https://doi.org/10.1080/14631377.2011.595119)
- [9] GABRIELIAN, D.: Forecasting Inflation Using the Phillips Curve in Inflation Targeting Countries. *International Review of Applied Economics*. 2019, Vol. 33, Issue 5, pp. 601–623. ISSN 0269-2171. DOI: [10.1080/02692171.2018.1516740](https://doi.org/10.1080/02692171.2018.1516740)
- [10] KASPEROWICZ, R.; ŠTREIMIKIENĖ, D.: Economic growth and energy consumption: a comparison of comparative analysis of V4 and “old” EU countries. *Journal of International Studies*. 2016, Vol. 9, Issue 2, pp. 181–194. ISSN 2071-8330. DOI: [10.14254/2071-8330.2016/9-2/14](https://doi.org/10.14254/2071-8330.2016/9-2/14)
- [11] KLIBER, A.; PŁUCIENNIK, P.: Euro or Not? Vulnerability of Czech and Slovak Economies to Regional and International Turmoil. *Economic Modelling*. 2017, Vol. 60, Issue 1, pp. 313–323. ISSN 0264-9993. DOI: [10.1016/j.econmod.2016.09.019](https://doi.org/10.1016/j.econmod.2016.09.019)
- [12] MELIKHOVA, Y.; BAŽÓ, L.; HOLUBCOVA, I.; CAMACHO, J. A.: Trade in Services and Tertiarisation of the Visegrád Four Economies. *Post-Communist Economies*. 2015, Vol. 27, Issue 1, pp. 1–22. ISSN 1463-1377. DOI: [10.1080/14631377.2015.992219](https://doi.org/10.1080/14631377.2015.992219)
- [13] PAVELKOVÁ, D.; BIALIC-DAVENDRA, M.; JIRCIKOVÁ, E.; HOMOLKA, L.: Clusters’ Activities and Economy Stage of Development: Evidence from V4 and Advanced Economies. *Ekonomický časopis (Journal of Economics)*. 2013, Vol. 61, Issue 2, pp. 187–205. ISSN 0013-3035.
- [14] PRÓCHNIAK, M.: Determinants of Economic Growth in Central and Eastern Europe: The Global Crisis Perspective. *Post-Communist Economies*. 2011, Vol. 23, Issue 4, pp. 449–468. ISSN 1463-1377. DOI: [10.1080/14631377.2011.622566](https://doi.org/10.1080/14631377.2011.622566)
- [15] PRÓCHNIAK, M.: The Impact of Product Market Competition on GDP per Capita Growth in the EU Countries: Does the Model of Capitalism Matter? *Post-Communist Economies*. 2018, Vol. 30, Issue 2, pp. 131–155. ISSN 1463-1377. DOI: [10.1080/14631377.2017.1362098](https://doi.org/10.1080/14631377.2017.1362098)
- [16] RENCHER, A. C.; CHRISTENSEN, W. F.: *Methods of Multivariate Analysis*. John Wiley & Sons, New Jersey, 2012. ISBN 978-0-470-17896-6.
- [17] SIMIONESCU, M.; LAZÁNYI, K.; SOPKOVÁ, G.; DOBEŠ, K.; BALCERZAK, A. P.: Determinants of Economic Growth in V4 Countries and Romania. *Journal of Competitiveness*. 2017, Vol. 9, Issue 1, pp. 103–116. ISSN 1804-1728. DOI: [10.7441/joc.2017.01.07](https://doi.org/10.7441/joc.2017.01.07)
- [18] TSIAPA, M.; BATSIOLAS, I.: Firm resilience in regions of Eastern Europe during the period 2007–2011. *Post-Communist Economies*. 2018, Vol. 31, Issue 1, pp. 19–35. ISSN 1463-1377. DOI: [10.1080/14631377.2018.1443250](https://doi.org/10.1080/14631377.2018.1443250)
- [19] OECD: *OECD.Stat*. [online]. Available from WWW: <https://stats.oecd.org>



## ZHDNOCENÍ UKAZATELŮ ŽIVOTNÍ ÚROVNĚ V ČLENSKÝCH ZEMÍCH OECD

Výzkumná databáze této práce se skládá ze zemí OECD (s výjimkou Turecka a Litvy). Primárním cílem studie je rozdělit země do skupin podle dvanácti proměnných – průměrná mzda, minimální mzda, HDP na obyvatele, míra nezaměstnanosti, směnná a inflační sazba, daň z příjmu, HDP za odpracovanou hodinu, indexy průmyslu, stavebnictví a zpracovatelského průmyslu, index produkce a maloobchodu, a určit, která z nich významně ovlivňuje průměrnou mzdu, a definuje typ a sílu takového vztahu. Průměrná mzda, minimální mzda a HDP na obyvatele se počítají po jejich převodu na paritu kupní síly, což umožňuje srovnání cenových hladin a PPP v různých zemích. Dalším důležitým cílem je vypracovat prognózy úrovně mezd pro země OECD do roku 2020. S ohledem na seskupení zemí podle výše uvedených kritérií je Česko vždy na úrovni ostatních postkomunistických zemí (kromě Slovinska). Jedinými vysvětlujícími proměnnými, které významně ovlivňují průměrnou mzdu, jsou HDP na obyvatele, daně z příjmu a indexy zpracovatelského průmyslu a maloobchodu, přičemž HDP má hlavní vliv. Jednoduchá regresní analýza závislosti mezi průměrnou mzdou a HDP na obyvatele naznačuje, že její průběh nejlépe zachycuje konkávní parabola s maximem 77 252 PPP USD. Vybraná polynomiální regrese druhého řádu vysvětluje cca 89% variability pozorovaných hodnot průměrné roční mzdy. Růst mezd do roku 2020 se očekává prakticky ve všech zemích OECD.

## BEWERTUNG DER INDIKATOREN DES LEBENSNIVEAUS IN DEN OECD MITGLIEDSSTAATEN

Die Forschungsdatenbank dieser Arbeit besteht aus den Ländern der OECD (mit Ausnahme der Türkei und Litauens). Das primäre Ziel der Studie besteht in der Unterteilung der Länder in Gruppen gemäß bestimmter Variablen: Durchschnittslohn, Mindestlohn, Bruttoinlandsprodukt pro Einwohner, Arbeitslosenrate, Wechsel- und Inflationsrate, Einkommenssteuer, Bruttoinlandsprodukt für die abgeleiteten Stunden, Indexe von Industrie, Bauwesen und verarbeitender Industrie, Indexe von Produktion und Einzelhandel. Ein weiteres Ziel dieser Studie besteht in der Bestimmung, welcher dieser Faktoren den Durchschnittslohn beträchtlich beeinflusst. Sie definiert den Typ und die Kraft einer solchen Beziehung. Der Durchschnittslohn, der Mindestlohn und das Bruttoinlandsprodukt werden nach deren Überführung in die Parität der Kaufkraft berechnet, was einen Vergleich zwischen den Preisspiegeln und der Purchasing Power Parity (PPP) in den verschiedenen Ländern ermöglicht. Ein weiteres wichtiges Ziel besteht in der Ausarbeitung von Prognosen des Lohnniveaus für die Länder der OECD bis zum Jahr 2020. Unter Berücksichtigung der Gruppierungen der Länder gemäß den oben angeführten Kriterien befindet sich Tschechien immer auf dem Niveau der übrigen postkommunistischen Länder (außer Slowenien). Die einzigen erklärenden Variablen, welche einen bedeutenden Einfluss auf den Durchschnittslohn ausüben, sind das Bruttoinlandsprodukt pro Einwohner, die Einkommenssteuer und die Indexe der verarbeitenden Industrie und des Einzelhandels, wobei das Bruttoinlandsprodukt den Haupteinfluss ausübt. Die einfache Regressanalyse der Abhängigkeit zwischen dem Durchschnittslohn und dem Bruttoinlandsprodukt pro Einwohner deutet an, dass deren Verlauf am besten mit einer konkaven Parabel mit einem Maximum von 77.252 PPP USD dargestellt wird. Die ausgewählte Polynomialregression der zweiten Ordnung erklärt etwa 89 % der Variabilität der beobachteten Werte des Durchschnittsjahreslohns. Ein Wachstum der Löhne bis zum Jahr 2020 wird praktisch in allen Ländern des OECD erwartet.



## OCENA WSKAŹNIKÓW POZIOMU ŻYCIA W KRAJACH CZŁONKOWSKICH OECD

Badawcza baza danych niniejszego opracowania obejmuje kraje OECD (z wyjątkiem Turcji i Litwy). Głównym celem badań jest podzielenie krajów do grup według dwunastu zmiennych – przeciętne wynagrodzenie, wynagrodzenie minimalne, PKB na mieszkańca, stopa bezrobocia, stopa zmienna i stopa inflacji, podatek dochodowy, PKB na godzinę pracy, wskaźniki przemysłu, budownictwa i przemysłu przetwórczego, wskaźnik produkcji i handlu detalicznego, oraz wskazanie, która z nich znacznie wpływa na przeciętne wynagrodzenie oraz określa typ i siłę takiej relacji. Przeciętne wynagrodzenie, wynagrodzenie minimalne oraz PKB na mieszkańca uwzględniane są po ich przeliczeniu na parytet siły roboczej (PPP), co umożliwia porównanie cen i parytetów siły nabywczej w różnych krajach. Kolejnym ważnym celem jest opracowanie prognozy poziomu wynagrodzeń dla krajów OECD do 2020 roku. W wyniku pogrupowania krajów według ww. kryteriów Republika Czeska jest każdorazowo na poziomie pozostałych krajów postkomunistycznych (za wyjątkiem Słowenii). Jedynymi wyjaśniającymi zmiennymi, które mają znaczny wpływ na przeciętne wynagrodzenie, są PKB na mieszkańca, podatki dochodowe oraz wskaźnik przemysłu przetwórczego i handlu detalicznego, przy czym PKB ma wpływ decydujący. Prosta analiza regresji zależności pomiędzy przeciętnym wynagrodzeniem a PKB na mieszkańca wskazuje, że jej przebieg najlepiej odzwierciedla parabola wklęsła z maksimum 77 252 PPP USD. Wybrana regresja wielomianowa drugiego rzędu wyjaśnia ok. 89% zmienności badanych wartości przeciętnego rocznego wynagrodzenia. Wzrost wynagrodzeń do 2020 roku spodziewany jest praktycznie we wszystkich państwach OECD.