QUANTITATIVE ASPECTS OF THE LIVING STANDARD OF CZECH PENSIONERS DURING THE ENERGY CRISIS: CLUSTER ANALYSIS OF CZECH REGIONS

Diana Bílková

Abstract

In 2022, energy prices reached their all-time high, with the wholesale price of electricity in the European Union internal market is directly linked to the price of gas, which is mostly imported. This paper is focused on the standard of living of Czech pensioners in 2022 in individual regions of the Czech Republic. The data set for this research comes from the Czech Statistical Office and includes pensioners of the Czech Republic aged 65+. Average monthly pensioner's household income per capita, property of the pensioner's family and flat size by number of living rooms are chosen as variables characterizing the standard of living. Ward's method and Euclidean distance, which are the most widely used techniques, were used for cluster analysis. The Czech regions are clustered into four, five or six clusters. The results show that the region Hlavní město Praha always forms a separate cluster. Similarly, the Kraj Vysočina always forms a separate cluster, too.

Key words: Energy crisis, living standard, pensioner, cluster analysis, Ward's method.

JEL Code: H75, H55, E24

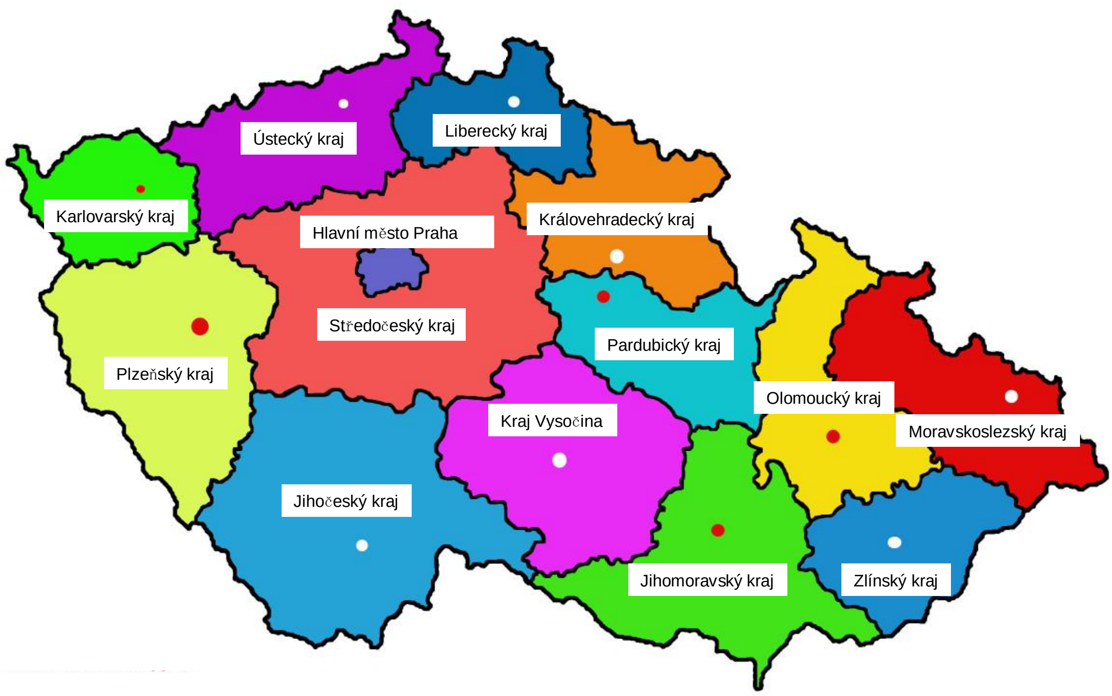
Introduction

The population of the most developed countries is growing older. The phenomenon of population aging thus becomes a problem for almost all developed countries, and the reform of the pension system is a frequently discussed topic in professional literature, too. For example, social security, pensions and retirement behaviour within the family is the subject of research of a pair of authors Gustman & Steinmeier (2024) or the public pension reform with ill-informed individuals is the subject of research of the Gustafsson (2023). A group of authors Porisky, Mohamed & Muthui (2023) deals with Kenya universal social pension linked to registration policy in Marsabit district. The pair of authors Prisiazhniuk & Sokhey (2023). compare pension politics in post-communist Russia and Hungary. Valls Martínez, Santos-

Jaén, Amin, F & Martín-Cervantes (2021) examines old-age pensions and social security from the perspective of literature review and global trends. Political representation and governance based on evidence from the investment decisions of public pension funds is the purpose of the research Andonov, Hochberg & Rauh (2018) or pension fund asset allocation and liability discount rates are addressed in the study Andonov, Bauer & Cremers (2017). The pair of authors Engels, Geyer & Haan (2017) deal with pension incentives and early retirement, the study Hagen (2017) deals with pension principles in the Swedish pension system. The group of authors Carone, Eckefeldt, Giamboni, Laine & Pamies (2016) examines pension reforms in the European Union since the early 2000's. Pension topics are the subject of interest for other publications: Jackwerth, & Slavutskaya (2016), Liu & Sun (2016), Mesa-Lago & Bertranou (2016), Atalay & Barrett (2015) or Möhring (2015).

This paper is focused on the living situation of Czech pensioners (in the 65+ age category) during the energy crisis, which hit them with the main force in 2022.

Fig. 1: Map of regions of the Czech Republic[1]



Source: Own construction

[1] Dots on the map of individual regions show the location of the regional capital.

The data set comes from the Czech Statistical Office and is for the year 2022. The following quantitative variables enter the analysis: average monthly household income per capita, total family property, flat size according to the number of rooms. The main objective of the research is to create clusters of regions of the Czech Republic that are as similar as possible in terms of the mentioned variables characterizing the quantitative aspect of the living standards of Czech pensioners. Cluster analysis is a tool used for this purpose. In cluster analysis, Ward's method and Euclidean distance are chosen as analysis tools. Considering the fourteen regions of the Czech Republic, the regions of the Czech Republic are clustered into four, five and six clusters. Figure 1 shows a map of the regions of the Czech Republic including the location of the respective regional capitals. The data were processed using the statistical program Statgraphics and the spreadsheet calculator MS Excel.

1 Theory and Methodology

One of the possibilities of using the information contained in the input data is the classification of a set of objects into a certain number of relatively homogeneous clusters. Using appropriate algorithms, it is possible to reveal the structure of the data file and classify individual objects. We will denote the number of clusters with the symbol k.

An important step is to assess to what extent the aim of cluster analysis has been achieved using a particular algorithm. Several decomposition quality functionals have been proposed for this purpose. There are intragroup variability matrix

$$\mathbf{E} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)(\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)^{\mathrm{T}} \tag{1}$$

and intergroup variability matrix

$$\mathbf{B} = \sum_{h=1}^{k} n_h (\bar{\mathbf{x}}_h - \bar{\mathbf{x}})(\bar{\mathbf{x}}_h - \bar{\mathbf{x}})^{\mathrm{T}}, \tag{2}$$

which together give the total variability matrix

$$\mathbf{T} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}})(\mathbf{x}_{hi} - \bar{\mathbf{x}})^{\mathrm{T}}, \quad \mathbf{T} = \mathbf{E} + \mathbf{B}. \tag{3}$$

In relations (1)–(3) appear:

$\mathbf{x}_{hi}$ … vector of observations for the i^{th} object in the h^{th} cluster;

$\bar{\mathbf{x}}_h$ … vector of averages for the h^{th} cluster;

$\bar{\mathbf{x}}$ … vector of averages for the entire file.

If we use p variables, these are p-member vectors, and **E**, **B**, and **T** represent symmetric square matrices of p^{th} order.

The basic objective is to construct maximally distant compact clusters. We will achieve this aim if the minimum of the total sum of squares of the deviations of all values from the respective cluster averages is reached

$$G_1 = \operatorname{st} \mathbf{E} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} \sum_{j=1}^{p} (x_{hij} - \bar{x}_{hj})^2. \tag{4}$$

The criterion (4) is called Ward's criterion. In addition to this criterion, it is also possible to use its simple monotonic functions. Since st **T** is the same for all decompositions, minimizing st **E** is the same as maximizing st **B**.

After the choice of variables to characterize the properties of clustered objects and their values have been determined, it is necessary to decide on the method of evaluating the distance or similarity of the objects. Often, the first stage of implementing a clustering algorithm is the calculation of the relevant measures for all pairs of objects. This creates a symmetric square matrix of the type $n \times n$, whose diagonals are zeros in the case of the distance matrix **D**, or ones in the case of the similarity matrix **A**.

If the individual variables are roughly at the same level or at least expressed in the same measurement units, it is possible to use the Hemming's distance, which can also be found in the literature under the name Manhattan or city-block distance

$$D_H(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^{p} \left| x_{ij} - x_{i'j} \right|, \tag{5}$$

Euclidean distance

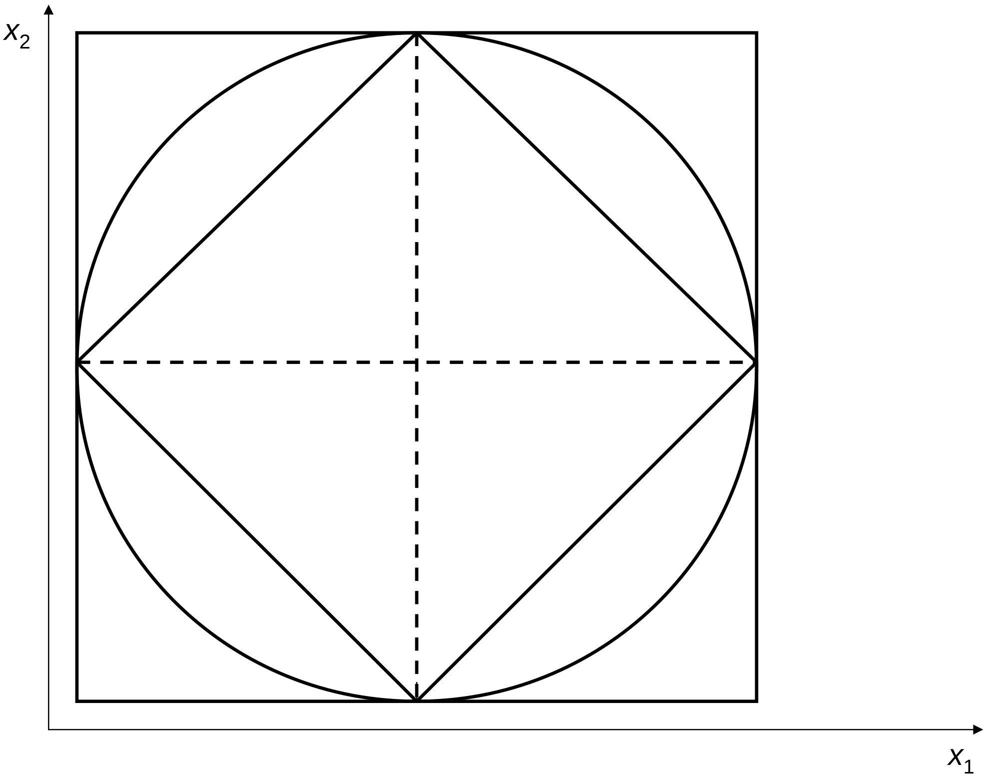
$$D_E(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2}, \tag{6}$$

or the Chebyshev's distance

$$D_C(\mathbf{x}_i, \mathbf{x}_{i'}) = \max_j \left| x_{ij} - x_{i'j} \right|. \tag{7}$$

We choose between these distances according to how we need to strengthen the influence of the variables for which we observe an exceptionally large difference on the total sum. Figure 2 provides a certain idea in this direction. For a given measure, objects on the perimeter of the respective figure have the same distance from its center: the inner square characterizes the Hemming's distance, the circle the Euclidean distance and the outer square the Chebyshev's distance.

Fig. 2: Representation of object distances in two-dimensional space



Source: Own construction

In terms of agglomerative procedures, a number of procedures exists. Nearest neighbour method (single bond, simple connection) is historically the oldest method, within which both clusters, the connection of which we are considering, are represented by the objects that are the closest to each other. The distance between h^{th} a h'^{th} cluster $D_{hh'}$ is understood as the minimum of all $q = n_h n_{h'}$ distances between their objects. In the next step, we replace h^{th} a h'^{th} row and column in the distance matrix with the row and column of the distances of the new g^{th} cluster from the other clusters (denoted by g'). In the v^{th} cycle we write a total of $n - v - 1$ distances determined on the basis

$$D_{gg'} = \min(D_{g'h}, D_{g'h'}). \quad (8)$$

When using this method, a situation occurs that often even very distant objects can meet in the same cluster if a larger number of other objects establish a kind of bridge between them. The characteristic chaining described is considered a disadvantage, especially if we have reason to require the clusters to have a regular elliptical shape with a compacted core.

Farthest neighbour method (full bond, full connection) is based on the opposite principle. The criterion for connecting clusters is the maximum from $q = n_h n_{h'}$ of the possible inter-cluster distances of objects. As part of the adjustment of the distance matrix, we proceed with the use

$$D_{gg'} = \max\left(D_{g'h}, D_{g'h'}\right). \qquad (9)$$

The unwanted chain effect is eliminated here, on the contrary, there is a tendency to create compact clusters, not exceptionally large ones.

The average bond method (Sokal-Sneaths' method) uses the average from $q = n_h n_{h'}$ of the possible inter-cluster distances of objects as a criterion for connecting clusters. As part of the recalculation of the distance matrix, we use

$$D_{gg'} = \frac{n_h D_{g'h} + n_{h'} D_{g'h'}}{n_h + n_{h'}}. \qquad (10)$$

The mentioned method often leads to analogous results as the farthest neighbour method.

Centroid method (Gower's method) is no longer based on the summation of information on the inter-cluster distances of objects, but the criterion is the Euclidean distance of the centroids

$$D_E\left(\bar{\mathbf{x}}_h, \bar{\mathbf{x}}_{h'}\right) = \sum_{j=1}^{p} \left(\bar{x}_{hj} - \bar{x}_{h'j}\right)^2, \qquad (11)$$

the recalculation of the distance matrix is done using

$$D_{gg'} = \frac{1}{n_h + n_{h'}}\left(n_h D_{g'h} + n_{h'} D_{g'h'} - \frac{n_h n_{h'}}{n_h + n_{h'}} D_{hh'}\right). \qquad (12)$$

<u>Ward's method</u> applies the decomposition quality functional G_1 in relation (4), the criterion for joining clusters is the increment of the total intra-group sum of squares of observation deviations from the cluster average

$$\Delta G_1 = \sum_{i=1}^{g} \sum_{j=1}^{p} (x_{gij} - \bar{x}_{gj})^2 - \sum_{i=1}^{h} \sum_{j=1}^{p} (x_{hij} - \bar{x}_{hj})^2 - \sum_{i=1}^{h'} \sum_{j=1}^{p} (x_{h'ij} - \bar{x}_{h'j})^2. \qquad (13)$$

The increment is expressed as the sum of squares in the emerging cluster reduced by the sums of squares in both disappearing clusters. Using simple arithmetic adjustments, the relation (13) can be simplified to the form

$$\Delta G_1 = \frac{n_h \, n_{h'}}{n_h + n_{h'}} \sum_{j=1}^{p} (\bar{x}_{h'j} - \bar{x}_{h'j})^2, \qquad (14)$$

as the product of the Euclidean distance between the centroids of the clusters considered for connection and a coefficient that presents the dependence on the size of the clusters, i.e. the value of the given coefficient increases with the increasing size of the clusters and for fixed $n_h + n_{h'}$ it is maximal if we have clusters of the same size $n_h = n_{h'}$. Considering the fact that we perform a connection that ensures minimization of the Δ G1 criterion, Ward's method has the often welcome property of tending to remove small clusters and thus to produce clusters of roughly the same size.

Ward's method and Euclidean distance are used in this research. Considering the fourteen regions of the Czech Republic, these regions are divided into four, five and six clusters.

2 Results

Table 1 shows the results of the cluster analysis. We can see from Table 1 that the Hlavní město Praha always forms a separate cluster. The Kraj Vysočina also forms a separate cluster in each case. Středočeský kraj, Královehradecký kraj, Pardubický kraj, Olomoucký kraj and Zlínský kraj represent a group of regions always included in the same cluster. Jihočeský kraj, Plzeňský kraj and Ústecký kraj also represent another group of regions always included in the same cluster. The two Moravian regions, namely Jihomoravský kraj and Moravskoslezský kraj are always included in the same cluster. An interesting fact is that Karlovarský kraj and Liberecký kraj always belong to the same cluster.

Conclusion

Clusters of Czech regions that are as similar as possible in terms of selected indicators of the standard of living of pensioners very closely correspond to clusters formed on the basis of quantitative indicators of the standard of living of the working-age population, where the Hlavní město Praha is also always an independent cluster. The exception is the Karlovarský kraj, which has been long the poorest region in quantitative terms of the population's standard of living.

Tab. 1: Cluster analysis calculated using variables reflecting the quantitative aspect of a pensioner's standard of living: average monthly household income per capita, total family property, flat size – four, five and six clusters

Four clusters	Five clusters	Six clusters
1st cluster: Hlavní město Praha	**1st cluster:** Hlavní město Praha	**1st cluster:** Hlavní město Praha
2nd cluster: Středočeský kraj Karlovarský kraj Liberecký kraj Královehradecký kraj Pardubický kraj Olomoucký kraj Zlínský kraj	**2nd cluster:** Středočeský kraj Královehradecký kraj Pardubický kraj Olomoucký kraj Zlínský kraj	**2nd cluster:** Středočeský kraj Královehradecký kraj Pardubický kraj Olomoucký kraj Zlínský kraj
3rd cluster: Jihočeský kraj Plzeňský kraj Ústecký kraj Jihomoravský kraj Moravskoslezský kraj	**3rd cluster:** Jihočeský kraj Plzeňský kraj Ústecký kraj Jihomoravský kraj Moravskoslezský kraj	**3rd cluster:** Jihočeský kraj Plzeňský kraj Ústecký kraj
4th cluster: Kraj Vysočina	**4th cluster:** Karlovarský kraj Liberecký kraj	**4th cluster:** Karlovarský kraj Liberecký kraj
	5th cluster: Kraj Vysočina	**5th cluster:** Kraj Vysočina
		6th cluster: Jihomoravský kraj Moravskoslezský kraj

Source: Own construction

The amount of the pension depends on the number of years of service and lifetime earnings of employees (since 1986), from which they paid contributions to the pension insurance. The amount of earnings is different in individual regions. The highest earnings are on average in Prague, however, seniors in Prague do not necessarily have a higher standard of living than pensioners with a lower pension, as housing expenses play a decisive role.

Acknowledgment

This paper 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, Prague University of Economics and Business, Czech Republic.

References

Andonov, A., Bauer, R. M. M. J., & Cremers, K. J. M. (2017). Pension Fund Asset Allocation and Liability Discount Rates. *The Review of Financial Studies*, 30(8), 2,555–2,595. DOI: https://doi.org/10.1093/rfs/hhx020

Andonov, A., Hochberg, Y. V., & Rauh, J. D. (2018). Political Representation and Governance: Evidence from the Investment Decisions of Public pension Funds. *The Journal of Finance*, 73(5), 2,041–2,086. DOI: https://doi.org/10.1111/jofi.12706

Atalay, K., & Barrett, G. F. (2015). The Impact of Age Pension Eligibility Age on Retirement and Program Dependence: Evidence from an Australian Experiment. *Review of Economics and Statistics*, 97(1), 71–87. DOI: https://doi.org/10.1162/REST_a_00443

Carone, G., Eckefeldt, P., Giamboni, L., Laine, V., & Pamies, S. (2016). Pension Reforms in the EU since the Early 2000's: Achievements and Challenges Ahead. *European Economy Discussion*, Paper No. 042, Available at SSRN: https://ssrn.com/abstract=2964933 or http://dx.doi.org/10.2139/ssrn.2964933

Engels, B., Geyer, J., & Haan, P. (2017). Pension Incentives and Early Retirement. *Labour Economics*, 47, 216–231. DOI: https://doi.org/10.1016/j.labeco.2017.05.006

Hagen, J. (2017). Pension Principles in the Swedish Pension System. *Scandinavian Economic History Review*, 65(1), 28–51. DOI: https://doi.org/10.1080/03585522.2016.1269670

Gustafsson, J. (2023). Public Pension Reform with Ill-Informed Individuals. *Economic Modelling*, 121. DOI: https://doi.org/10.1016/j.econmod.2023.106219

Gustman, A. L., & Steinmeier, T. L. (2024). Social Security, Pensions and Retirement Behaviour within the Family. *Journal of Applied Econometrics*, 19(6), 723–737. DOI: https://doi.org/10.1002/jae.753

Jackwerth, J. C., & Slavutskaya, A. (2016). The Total Benefit of Alternative Assets to Pension Fund Portfolios. *Journal of Financial Markets*, 31, 25–42. DOI: https://doi.org/10.1016/j.finmar.2016.06.002

Liu, T., & Sun, L. (2016). Pension Reform in China. *Journal of Aging & Social Policy*, 28(1), 15–28. DOI: https://doi.org/10.1080/08959420.2016.1111725

Mesa-Lago, C., & Bertranou, F. (2016). Pension Reforms in Chile and Social Security Principles, 1981–2015. *International Social Security Review*, 69(1), 25–45. DOI: https://doi.org/10.1111/issr.12093

Möhring, K. (2015). Employment Histories and Pension Incomes in Europe: A Multilevel Analysis of the Role of Institutional Factors. *European Societies*, 17(1), 3–26. DOI: https://doi.org/10.1080/14616696.2014.934874

Porisky, A., Mohamed, T. S., & Muthui, P. M. (2023). Kenya's 'Universal' Social Pension: The Politics of Registration in Marsabit County. *World Development*, 164. DOI: https://doi.org/10.1016/j.worlddev.2022.106164

Prisiazhniuk, D., & Sokhey, S. W. (2023). Cracking the Nest Egg: Comparing Pension Politics in Post-Communist Russia and Hungary. *Social Policy and Society*, 22(2), 338–354. DOI: https://doi.org/10.1017/S1474746422000653

Valls Martínez, M. C., Santos-Jaén, J. M., Amin, F., & Martín-Cervantes (2021). Pensions, Ageing and Social Security research: Literature Review and Global Trends. *Mathematics*, 9(24), 1–26. DOI: https://doi.org/10.3390/math9243258

Contact

Diana Bílková
Prague University of Economics and Business
Faculty of Informatics and Statistics
Department of Statistics and Probability
Sq. W. Churchill 1938/4, 130 67 Prague 3
Czechia
E-mail: bilkova@vse.cz