127

WORKING PAPERS IER SAS
Pracovné listy EÚ SAV

Sergei Kharin - Ivan Lichner - Ján Výbošťok

Residential property price developments clustering – case of Slovakia

eúsav
Ekonomický ústav SAV
Institute of Economic Research SAS

AUTHORS
Sergei Kharin, Institute of Economic Research SAS, Bratislava, Slovakia
Ivan Lichner, Institute of Economic Research SAS, Bratislava, Slovakia
Ján Výbošťok, Institute of Geography SAS, Bratislava, Slovakia

Ivan Lichner and Sergei Kharin are acknowledging the support from grant no. VEGA 2/0158/24. Ján Výbošťok is grateful for support grant no. VEGA 2/0006/24.

Residential property price developments clustering – case of Slovakia

ABSTRACT
In this paper we utilize the rich database on the properties sales offers from online advertisements providers across period of 2012–2021 to identify the patterns in the price formation across the geographical units in Slovakia. The results indicate that the real estate markets do not follow the administrative territorial setup and utilization of the commuting regions is valuable option. Using integrated clustering mechanisms co-movements of the prices are identified. The results are line with previous research on factors forming the real estate prices, such as economic activity and transport infrastructure connectivity. Results provide arguments for further discussions on the policies relevant to the tackle of the housing affordability and its nuanced regional focusing.

KEYWORDS: regional prices, clustering, real estate
JEL LASSIFICATION: D31; E32; O47

Regionálne zhlukovanie vývoja cien slovenských nehnuteľností na bývanie

ABSTRAKT
V tomto článku je využitá detailná databáza ponúk na predaj nehnuteľností z online inzerátov za obdobie rokov 2012–2021. Tieto údaje umožnili identifikovať vzorce vývoja cien v rôznych geografických jednotkách na Slovensku. Výsledky naznačujú, že realitné trhy nesledujú administratívne územné členenie a využitie štruktúry regiónov podľa dochádzky za prácou, je cennou alternatívou. Pomocou integrovaných mechanizmov zhlukovej analýzy sme identifikovali spoločné pohyby cien. Výsledky sú v súlade s predchádzajúcim výskumom faktorov, ktoré ovplyvňujú ceny nehnuteľností, ako je hospodárska aktivita a dostupnosť dopravnej infraštruktúry. Výsledky poskytujú argumenty pre ďalšiu diskusiu o politikách súvisiacich s riešením dostupnosti bývania a jeho regionálnym zameraním.

KĽÚČOVÉ SLOVÁ: regionálne ceny, zhluková analýza, nehnuteľnosti
JEL KLASIFIKÁCIA: D31; E32; O47

The WORKING PAPER SERIES is intended to convey preliminary, partial results of ongoing research achieved by fellows or research groups of the Institute of Economic Research which can be prepared for later publications.

The views expressed in the WP and the language revision is those of the authors.

LAYOUT BY: Sergei Kharin - Ivan Lichner - Ján Výbošťok

Institute of Economic Research SAS
Ekonomický ústav SAV, v.v.i.
Šancová 56, 811 05 Bratislava
www.ekonom.sav.sk

CONTACT / KONTAKT: ekonedra@savba.sk

1. Introduction 5
2. Literature review 5
3. Data and Methodology 7
4. Results 10
5. Conclusions and discussion 13
Acknowledgements 14
References 15
Appendix 18

1. Introduction

Affordability of housing are of growing concern across the Europe and are more profound in Central-Eastern European countries (Deloitte, 2025). Recent developments in the real-estate markets across Europe has triggered lively discussions on the polices that should help to mitigate the situation. According to the Eurofound (2023) there was significant divide in the real estate prices between capital cities, localities with job opportunities and other parts of the Europe. Recent study for EESC (2025) turns focus on the housing affordability deterioration in the period after year 2010. Interplay of the various factors forming this development including social, economic, environmental, and institutional are stressed in this study. The findings of Gabrieli et al. (2024) highlights growing importance of utilizing the new sources of data to support policy making in context of territorial impact assessment. To shed some light on the matter of real estate market developments in the pre-COVID era this paper provides additional perspective. In our work we utilize rich online based database of real-estate listing prices and apply advanced clustering methods to yield information on similarities on the price formation process.

The rest of paper is organized as follow: the next section provides overview of recent advances in the field of regional research of real estate market. Third section describe the data used to perform the empirical analysis of flat prices developments in Slovakia and methodological grounds of such analysis. Following part focuses on the results of the analysis and final sub-chapter conclude and hints on the possible avenues for the further research.

2. Literature review

Various methodologies are utilized to investigate the regional patterns in the real estate markets including a panel data analysis approach (Marinković, Džunić and Marjanović, 2024), hierarchical linear models (Huang, Qiao and Yeh, 2024) and interpretable machine learning algorithms (Liu et al., 2024). In the work of Valadkhani, Smyth and Worthington (2017) calendar effects for appartement and house market are discussed. Results provide reasoning for spatial and seasonal differences and identify changes in patterns of the seasons with relatively higher returns for sellers. Housing prices are result of complex relationship of various factors. Apart from size, condition, and amenities of a property, location is the most significant factor (Wittowsky et al., 2020). Location of property reflets in the context of real-estate prices reflects multiple dimensions in terms of spatial distance to services, amenities, and economic activity. In this sense the real-estate prices refers to balance between pros and cons related to the property. This resonates with the results of Skaburskis (2004) which argue that after accounting for socio-economic factors differences in real-estate prices become marginal. However, Wen, Xiao and Hui (2019) using spatial quantile regression models identified that after neglecting spatial relationships property capitalization become overestimated. Analysis of Li and Janet Ge (2023) investigate factors on the housing prices increases in the regional areas during the COVID-19 period. In the Gray (2018) author using clustering and common trend methods investigate the convergence of prices in the UK. The results suggest the existence of hegemon and potential regional convergence outside London.

Results of variability of homeownership analysed in Lerbs and Oberst (2014) suggest that the policy measures should be more focused on the supply side of the real estate market rather than on the demand. Panel data based approach of Deng, Nanda and Ong (2019) investigate the effects of infrastructure investments in India on real estate prices developments. Results suggest the increased property prices, but on the other hand deteriorating effects on rental market. Wolniak et al. (2020) comparing transactions on primary and secondary markets in Central Pomerania found that apartment price formation is affected by market structure developments. The research (e.g. Helderman, Van Ham, & Mulder, 2006; Palomares-Linares & Van Ham, 2020) suggests that homeownership lock-in effect is stronger for less developed regions. This might in turn result in less housing market transactions.

Meen (1999) has devised the ripple effect in the housing market price developments. From regional perspective three main types of components are suggested. In our paper we focus on the structural differences in regional real estate markets and try to investigate potential co-movements at level of clusters. Propagation of the house price shocks from centre to the regions across time and space (and dependence of centre on development in its global counter-part) is analysed in the Holly, Pesaran and Yamagata (2011). The results suggest slower diffusion across space compared to the time dimension. Kaas, Kocharkov and Syrichas (2024) argue that the most important factor driving the property price increases is housing demand. The authors utilize labour market regions as spatial unit of analysis.

Hromada, Čermáková and Piecha (2022) utilizing the web-scraped real estate market advertising data investigate the connection between the property prices and inequality. Intergenerational, interclass and homeowner status polarizations are exclaimed as possible future source of social tensions. Combining the online ask prices and transaction prices Ardila, Ahmed and Sornette (2021) concluded long-term co-integrated relation. The authors argue that online real-estate listing prices are good representation of market dynamics. Using the novel tool combining machine learning techniques Adolfsen et al. (2022) segment he internet data on real estate prices to spatial units differing from administrative ones. This new segmentation improve monitoring performance on the micro-level and allow to identify temporal change of clusters characteristics or their spatial composition.

3. Data and Methodology

3.1. Data

In this work we utilize dataset comprising of the real estate sales offers advertised online at reality.sk property listing portal (Datalan, 2022) during the period of 2012-2021 across the all Slovak regions. Total sample of the sales from online advert data comprise of more 1.28 millions of unique entries. The coverage reflects differences in the internet usage literacy and availability and the local specifics of some (remote) territories of Slovakia. Thus in some regions there is significantly low occurrence of the data as result of prevailing utilization of other (more traditional) sales channels for the execution of the property sale process. The Table 1 presents the descriptive stats of data used. We split the original dataset to differentiate the real estate market between the all properties and flats as the second type of the dwellings (family houses) is showing different pattern in terms of the geographical availability and price formation process. The main focus of the analysis presented in this paper is on the spatial-temporal price developments of the flats. The all property results serve as robustness check in current research strategy.

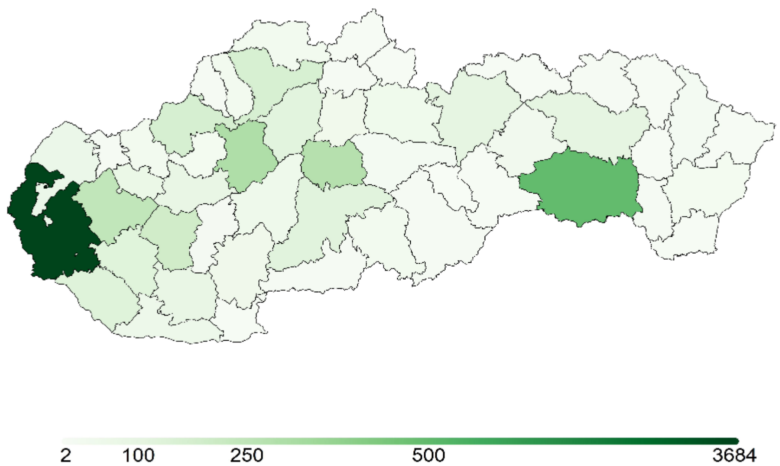
Table 1: Descriptive statistics

		Area / m^2			**Offer price / EUR**			**Price per sq. meter EUR/m^2**
	N. obs.	*median*	*mean*	*max*	*median*	*mean*	*max*	*mean*
All property	1,276,364	74	110.8	3,060	87,500	108,955	899,000	1,288.6
Flats	877,201	65	66.4	297	82,000	96,897	620,189	1,478.1

Source: Authors

To reflect recent findings of Halás and Klapka (2024) about the functional commuting regions (FCRs) in Slovakia we aggregate the original data from municipal level of the 50 FCRs1. This data manipulation removes the potential administrative bias so the results should be more robust to the current administrative regional setup which do not reflect the natural flows and interconnections of the regions. Additionally, temporal dimension of data was aggregated from daily updates to monthly values to allow utilization of the clustering approach defined in the following sub-chapter. To remove outliers and potential inconsistencies in the manual data inputs upper 0.1% of observations were eliminated for following characteristics: area size, offer price and price per square meter.

Map 1: Average monthly number of listings per FCRs, flats



Source: Authors

The data on the sales offers of the flats are significantly concentrated in area of the capital city – Bratislava – with on average more than 3,600 active sales offers per month. Other FCRs are lagging behind tremendously, second most active real estate market by online listings of offers is in Košice (the second largest city in country) with approximately 500 sales offers per month. The lowest utilization of online offers for sales is most visible in the bordering regions on east, north and central-south parts of Slovakia. In those regions only up to 10 offers of flats sales per month were listed on average during the period analysed.

[1] Regional system FRD-B assessed in Halás and Klapka (2024) as most suitable observation structure for spatial analysis.

3.2. Methodology

Residential property price series exhibit temporal dynamics characterized by time shifts and non-linear patterns that traditional Euclidean based clustering algorithms often do not capture adequately. This study employs a clustering approach that combines the K-means algorithm by MacQueen (1967) with Dynamic Time Warping (DTW) distance to identify patterns in price series data. We use traditional K-means approach to accommodate the temporal nature of the data by replacing Euclidean distance with DTW by Sakoe and Chiba (1978), that takes temporal distortions and phase shifts between sequences into account. DTW is known to be more accurate than Euclidean distance for time series analysis, particularly in the case of small datasets (Wang et al., 2013). Indeed, DTW distance is suitable for clustering, producing more meaningful results where clusters gather time series of similar shape (Petitjean et al., 2011).

The integration of K-means with DTW distance has proven valuable in the empirical studies, where understanding price dynamics is crucial for investment strategy and market segmentation (Aqsari et al., 2022; Nakagawa et al., 2019). However, applications of K-means with DTW to real estate market clustering remain limited in the empirical literature that is motivated to fill this gap.

We collected data on the monthly residence real estate prices over the period of Jan 2012-Dec 2021 across 50 Slovak administrative units. Missing values were handled using imputation for some regions with exponential weighted moving averages.

DTW computes the optimal alignment between two time series by finding the minimum-cost warping path through a distance matrix.

Given two time series $X = \{x_1, x_2, ..., x_n\}$ and $Y = \{y_1, y_2, ..., y_m\}$, the DTW distance is calculated in few steps:

1. Distance Matrix Construction. A cost matrix $C(i,j)$ is constructed where each element represents the Euclidean distance between points x_i and y_j: $C(i,j) = |x_i - y_j|$
2. Accumulated Cost Matrix. The cumulative distance matrix $D(i,j)$ is computed recursively:

$$D(i,j) = C(i,j) + min\{D(i-1,j), D(i,j-1), D(i-1,j-1)\}$$

3. Optimal Warping Path. The DTW distance is the value at D(n,m), representing the minimum cumulative cost of aligning the two sequences.

The traditional K-means algorithm was adapted to incorporate DTW distance.

First, we select k initial centroids randomly from the dataset (in our study, k=8 in accordance with the number of Slovak self-governing regions).

Second, each time series was assigned to the nearest cluster by computing DTW distances between the time series and all k centroids. Time series x_i was assigned to cluster C_j if:

$$DTW(x_i, \mu_j) = min\_\{k=1,...,K\}\ DTW(x_i, \mu_k)$$

where μ_j represents the centroid of cluster j.

Third, new cluster centroids were computed as the arithmetic mean of all-time series assigned to each cluster. For cluster C_j containing n_j time series, the centroid μ_j was calculated as:

$$\mu_j(t) = (1/n_j)\ \Sigma\ x_i(t) \text{ for all } x_i \in C_j$$

Fourth, steps 2 and 3 were repeated until convergence criteria were met, namely no reassignment of time series between clusters.

The optimal number of clusters k was determined using the elbow and Silhouette methods, as well as Calinski-Harabasz index, Dunn index and Davies-Bouldin score.

The clustering algorithm was implemented using R software (version 4.5.0) with the specific libraries: dtwclust, NbClust, factoextra.

4. Results

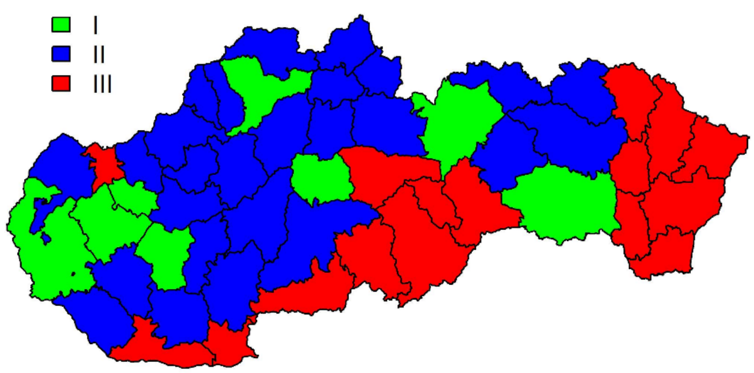
Data used for the analysis were relatively rich in terms of granularity and required significant efforts to perform the data cleaning tasks. After this process we elaborated them to the structure necessary to conduct the clustering analysis as detailed in the methodological section. As the main results we describe the price developments of the sub-group of the flats asking prices which comprises approximately 69% of the total entries in the cleaned version of underlying database. Results on the full-sample data are provided in the Appendix as robustness check and to illustrate the change induced by the inclusion of houses advertised on the online real estate portals[2].

In the first step we tested our clustering approach on the price levels of the flats marketed during the period of years 2012–2021. The resulting optimal number of clusters for this instance of analysis was three and their spatial distribution significantly overlap with the economic performance of the respective regions, demographic processes and migration patterns. There results mimic the similarities in various aspects within the clusters that indicates that also

[2] Within the category of houses following types of the properties were listed: Family house; Former farm, homestead; Cottage, holiday home; Other residential and recreational facility; Residential and recreational facilities; Family villa; Country house and Garden cottage.

process of real estate price formation is correlated with those variables. In case of cluster I FCRs of capital city and local centres are dominantly covered. Those localities included represent regional capitals and their hinterlands with relatively strong economic performance and representing the within-country migration destinations. As the regions included are more urbanized higher services concentration in economic mix is present. Member regions of the cluster II represent an semi-peripheral type of areas with higher concentration of industrial and agricultural production activities. Those regions are in relative vicinity to centres with stabilized population. Additionally, the results of this cluster corroborate findings of Deng, Nanda and Ong (2019) of higher price levels in regions with connections to highways. Last cluster III comprises of the economic periphery of the Slovakia and covers majority of disadvantaged regions as defined in the law[3]. The results of our clustering resonates with the findings of Výbošťok and Paur (2025) that concluded concentration of less-developed regions in the south-east parts of the country. Among the common characteristics of those localities on-going depopulation, overrepresentation of marginalized communities and long-term economic low performance are present.

Map 2: Clustering of flat prices, levels



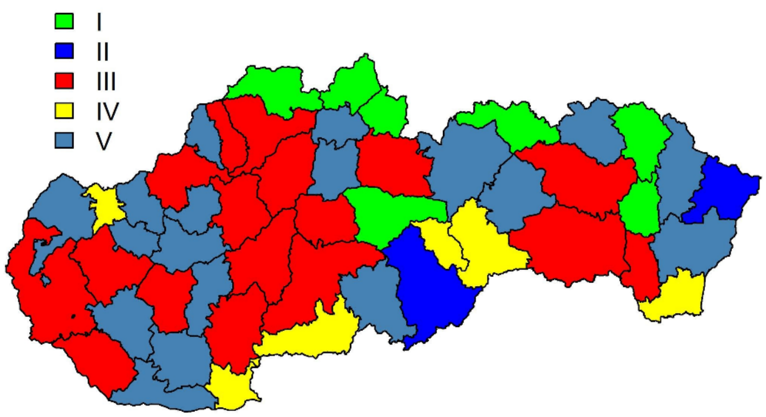
Source: Authors

Analysis of the price dynamics and its co-movements was performed on the data transformed from levels to the year-on-year differences for each month and the comparison of the 12-months moving averages. The first indicator represent seasonally unadjusted view on the price developments and provides a slightly different picture as the analysis on the price levels. Clustering approach resulted in five optimal clusters. There is significant overlap of clusters V and

[3] 336/2015 Coll. on support for the least developed districts.

III with clusters I and II from price levels clustering. However, both are telling a slightly different story. While in case of cluster V the year-on-year price dynamics variability are relatively stable over period analysed. On the other hand, price dynamics in regions of cluster III recorded steadily increasing trend (Figure 2 in Appendix). There were different price dynamics in case of peripheral regions. The cluster I regions price dynamics shown significant volatility peaking around year 2017 and 2018. Cluster II comprise of only two regions and indicate high dynamics of year-on-year flat prices which is resulting from low volumes of transactions. The regions covered in the cluster IV are from peripheral localities identified price levels clustering and has two peaks in volatility around the period of economic system heights and COVID-19 era.

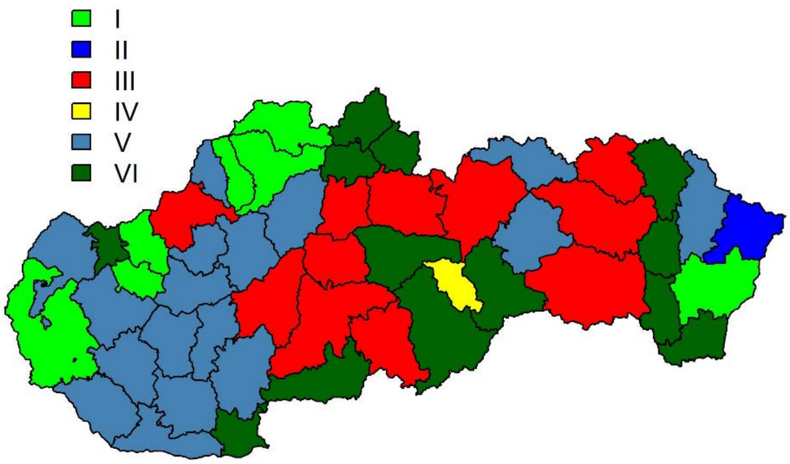
Map 3: Clustering of flat prices, y-o-y changes



Source: Authors

The final clustering analysis represent robustness check for year-on-year price dynamics by smoothing the seasonality using the 12-months moving average procedure. This clustering resulted in six groups of regions with similar developments. The results suggests significant effects of highway connection effects on the co-movements in clusters III and V. Comparing the price dynamics between those two clusters (Figure 3 in Appendix) indicate accelerating prices in cluster III and peak in price dynamics in years 2017–2019 for cluster V. The price developments of Revúca (cluster IV) and Snina (cluster II) were so unique that represent outliers and are forming a single region clusters. The prices in the cluster I were growing over the period analysed in relatively moderate pace. On the other hand there were significantly more dynamic price changes in the regions covered within cluster VI.

Map 4: Clustering of flat prices, 12-months moving averages



Source: Authors

5. Conclusions and discussion

The study present utilization of the clustering approach to analyse unique dataset of the online real estate price adverts in Slovakia for period of years 2012–2021. To overcome the issue identified in the empirical work of Adolfsen et al. (2022) we utilized FCRs as spatial units as more appropriate replacements of administrative structure of Slovakia. However, this remain one of the limitations of our research. Application of spatial aggregation on the level of FCRs defined by Halás and Klapka (2024) based on the daily movement of the population still not necessarily reflect the local real estate markets borders. The results of the performed price levels analysis segmented the housing market in Slovakia to three distinct groups of regions: central, semi-peripheral and peripheral. Those findings are in line with expectations and resonate also with the views of Výbošťok and Paur (2025). From the policy perspective important view is provided by the clustering over the price dynamics which highlights the main differences and should be crucial in area of housing support.

Further research in the area should be conducted on the enlarged dataset including the post-COVID-19 periods and focusing on the analysis of housing prices spill-over effects and underlying factors. Additionally alternative spatial structures need to applied to verify the robustness of the results and to possibly identify new regional functional structure aligning with the real estate market.

Acknowledgements

Ivan Lichner and Sergei Kharin are acknowledging the support from grant no. VEGA 2/0158/24. Ján Výbošťok is grateful for support grant no. VEGA 2/0006/24.

References

Adolfsen, J. F., Mønsted, B. M., Bay Schmith, A. M., Tang-Andersen Martinello, A., Gudiksen, S., & Sonberg, K. F. (2022). Segmentation of the housing market with internet data: Evidence from Denmark (No. 188). Danmarks Nationalbank Working Papers.

Ardila, D., Ahmed, A., & Sornette, D. (2021). Comparing ask and transaction prices in the Swiss housing market. Quantitative Finance and Economics, 5(1), 67-93.

Aqsari, H. W., Prastyo, D. D., & Rahayu, S. P. (2022). Clustering Stock Prices of Financial Sector Using K-Means Clustering With Dynamic Time Warping. In 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (pp. 503-507). IEEE.

Datalan. (2022). Property prices 2012–2019 [Data set]. Cenová mapa nehnuteľností.

Deliotte. (2025). Property index. Overview of European Residential Markets. 14th edition, August 2025. https://www.deloitte.com/content/dam/assets-zone2/cz-sk/cs/docs/services/financial-advisory/real-estate/property-index/2025_ADV_Property-index_online_22082025.pdf

Deng, X., Nanda, A., & Ong, S. E. (2019). Does infrastructure spending lead to price effects in the property market? Evidence from major cities across India. Regional Studies, 53(12), 1747-1760.

Eurofound (2023). Unaffordable and inadequate housing in Europe. https://www.eurofound.europa.eu/system/files/2023-05/ef22024en.pdf.

Gabrielli, L., Sulis, P., Fontana, M., Signorelli, S., Vespe, M., & Lavalle, C. (2024). Computational social science in regional analysis and the European real estate market. Regional Studies, 58(8), 1583-1602.

Gray, D. (2018). Convergence and divergence in British housing space. Regional Studies, 52(7), 901-910.

Halás, M., & Klapka, P. (2024). Aktualizácia vymedzenia funkčných regiónov Slovenska: hierarchia, neurčitosť a využiteľnosť. Geografický časopis, 76, 141-163.

Helderman, A. C., Van Ham, M., & Mulder, C. H. (2006). Migration and home ownership. Tijdschrift voor Economische en Sociale Geografie, 97(2), 111-125.

Holly, S., Pesaran, M. H., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. Journal of urban economics, 69(1), 2-23.

Hromada, E., Čermáková, K., & Piecha, M. (2022). Determinants of house prices and housing affordability dynamics in the Czech Republic. European Journal of Interdisciplinary Studies, 14(2), 119-132.

Huang, G., Qiao, S., & Yeh, A. G. O. (2024). Multilevel effects of urban form and urban functional zones on housing prices: evidence from open-source big data. Journal of Housing and the Built Environment, 39(2), 987-1011.

Lerbs, O. W., & Oberst, C. A. (2014). Explaining the spatial variation in homeownership rates: Results for German regions. Regional Studies, 48(5), 844-865.

Li, D., & Janet Ge, X. (2023). Factors of regional spillover effects on housing prices: A literature review. The Australasian Journal of Regional Studies, 29(1), 49-72.

Liu, X., Chen, X., Orford, S., Tian, M., & Zou, G. (2024). Does better accessibility always mean higher house prices?. Environment and Planning B: Urban Analytics and City Science, 51(9), 2179-2195.

Marinković, S., Džunić, M., & Marjanović, I. (2024). Determinants of housing prices: Serbian Cities' perspective. Journal of housing and the built environment, 39(3), 1601-1626.

McQueen, J. B. (1967). Some methods of classification and analysis of multivariate observations. In Proc. of 5th Berkeley Symposium on Math. Stat. and Prob. (pp. 281-297).

Meen, G. (1999). Regional house prices and the ripple effect: a new interpretation. Housing studies, 14(6), 733-753.

Palomares-Linares, I., & Van Ham, M. (2020). Understanding the effects of homeownership and regional unemployment levels on internal migration during the economic crisis in Spain. Regional Studies, 54(4), 515-526.

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. Pattern recognition, 44(3), 678-693.

Sakoe, H., & Chiba, S. (2003). Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing, 26(1), 43-49.

Skaburskis, A. (2004). Decomposing Canada's Growing Housing Affordability Problem: DoCity Differences Matter? Urban Studies, 41(1), 117–149.https://doi.org/10.1080/0042098032000155713

Valadkhani, A., Smyth, R., & Worthington, A. (2017). Regional seasonality in Australian house and apartment price returns. Regional Studies, 51(10), 1553-1567.

Výbošťok, J., & Paur, D. (2025). Housing in Slovakia: COVID-19 pandemics impacts on regions, affordability and residential development. Regional Studies, Regional Science, 12(1), 881-895.

Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., & Keogh, E. (2013). Experimental comparison of representation methods and distance measures for time series data. Data Mining and Knowledge Discovery, 26(2), 275-309.

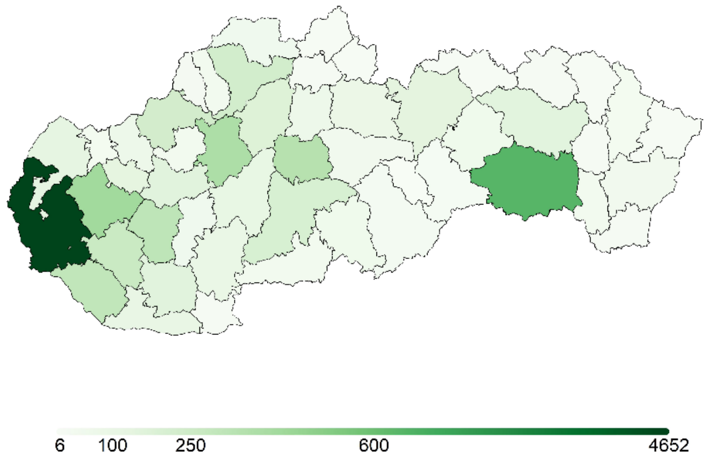
Wen, H., Xiao, Y., & Hui, E. C. M. (2019). Quantile effect of educational facilities on housingprice: Do homebuyers of higher-priced housing pay more for educational resources?Cities, 90, 100–112. https://doi.org/10.1016/j.cities.2019.01.019

Wittowsky, D., Hoekveld, J., Welsch, J., Steier, M. (2020). Residential housing prices: impactof housing characteristics, accessibility and neighbouring apartments – a case study ofDortmund, Germany. Urban, Planning and Transport Research, 8(1), 44–70.https://doi.org/10.1080/21650020.2019.1704429

Wolniak, R., Olkiewicz, M., Szymczewska, M., & Olkiewicz, A. (2020). The functioning of the real estate market: Dynamics of price formation and the sale of apartments.

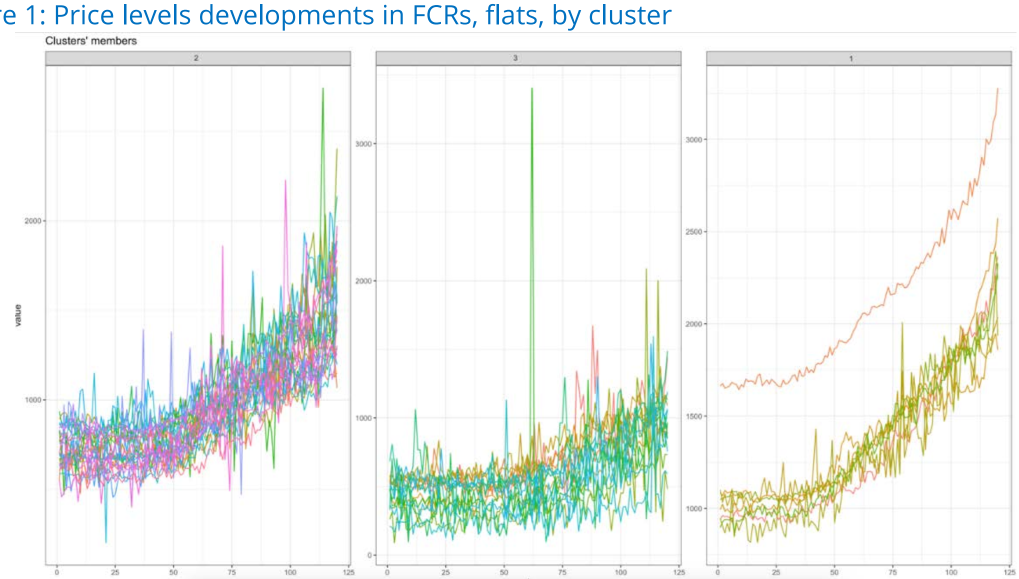
Appendix

Map 5: Average monthly number of listings per FCR, all property



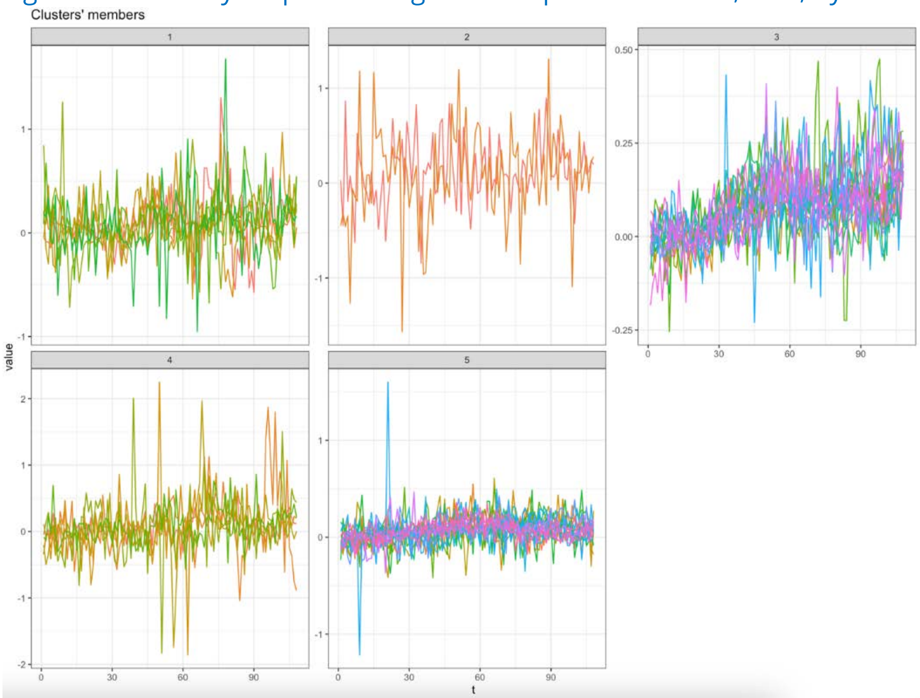
Source: Authors

Figure 1: Price levels developments in FCRs, flats, by cluster



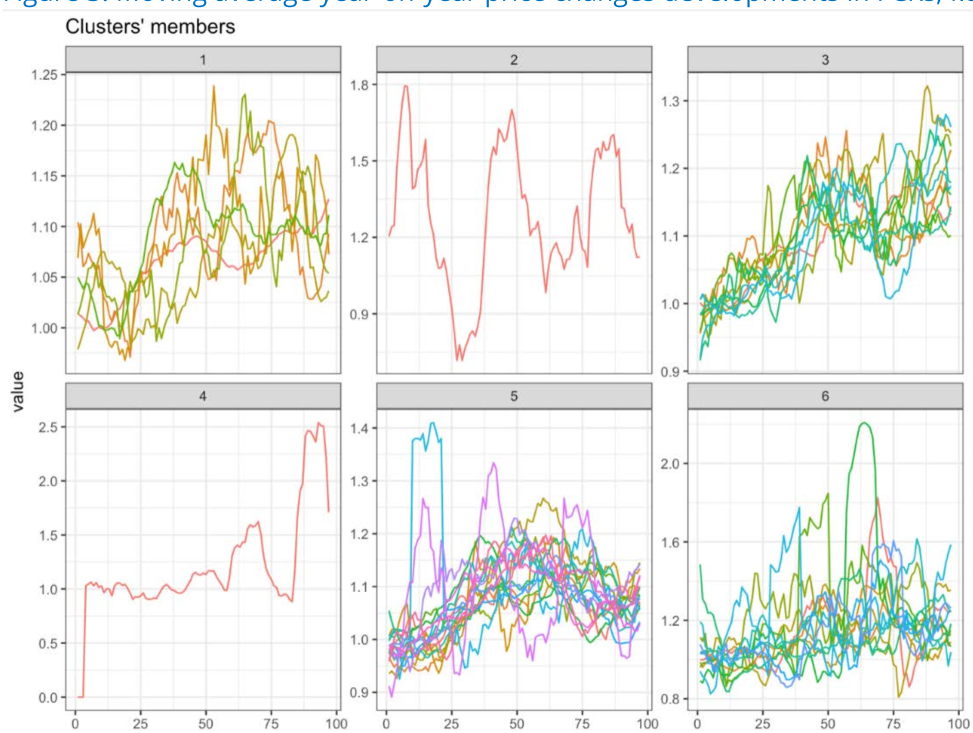
Source: Authors

Figure 2: Year-on-year price changes developments in FCRs, flats, by cluster



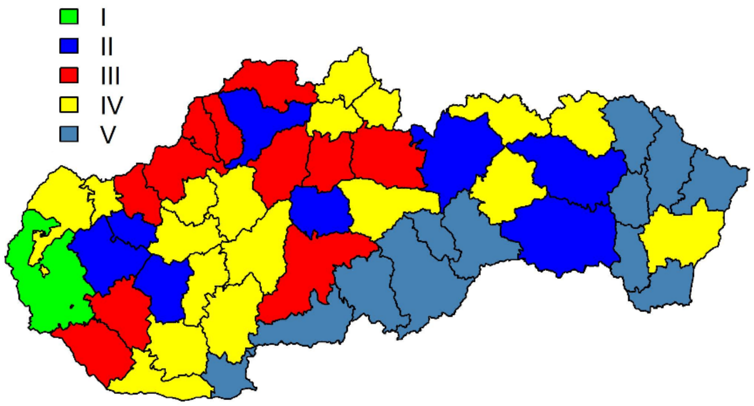
Source: Authors

Figure 3: Moving average year-on-year price changes developments in FCRs, flats, by cluster



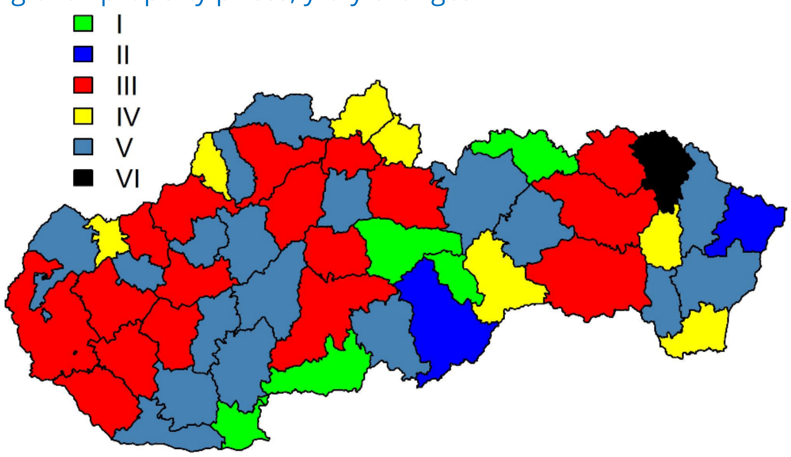
Source: Authors

Map 6: Clustering of all property prices, levels



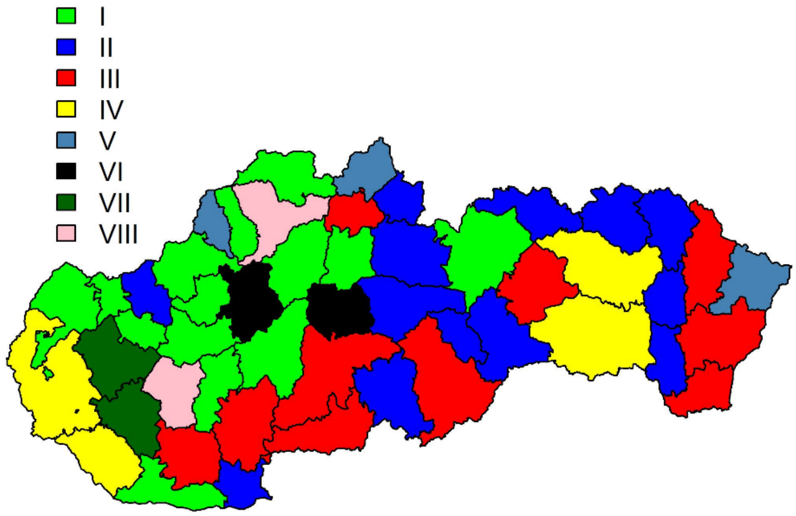
Source: Authors

Map 7: Clustering of all property prices, y-o-y changes



Source: Authors

Map 8: Clustering of all property prices, 12-month moving averages



Source: Authors