STU SvF
SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA
FACULTY OF CIVIL ENGINEERING

Slovak Journal of Civil Engineering

Vol. 33, 2025, No. 4, 42 –53
DOI: 10.2478/sjce-2025-0026

WHAT DO MACHINE LEARNING MODELS REVEAL ABOUT HOUSING UNAFFORDABILITY IN EUROPE? IMPLICATIONS FOR HOUSING POLICY

Janka ZAJACOVA [1*], Beata STEHLIKOVA[2], Maria ZUBKOVA[1]

Abstract

The paper aims to investigate the main factors affecting stress from the housing costs of household in Europe by applying machine learning methods to a set of structural, demographic, market and policy indicators. The research design combines variable selection using the VSURF algorithm with modelling via Random Forest and a subsequent interpretation of the results using SHAP analysis, including the identification of non-linear relationships and interactions between the variables. The data are drawn from the publicly available databases Eurostat, Housing Europe, Numbeo, Horwath HTL, and AllTheRooms. The analysis encompasses all 27 EU member states. It identifies threshold values for individual factors at which their effect on the household cost burden becomes disruptive. In addition, it quantifies the interactions among these variables. The findings thereby underscore the necessity of a comprehensively integrated housing policy. The contribution to knowledge lies in the methodological integration of machine learning with a public policy framework, which allows for a more precise identification of the determinants of housing unaffordability and better targeting of policy interventions.

Address

[1] Dept. of the Humane Sciences, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Bratislava, Slovak Republic

[2] Dept. of Economics, Faculty of Economics and Entrepreneurship, Pan-European University in Bratislava, Bratislava, Slovak Republic

[*] ***Corresponding author:*** janka.zajacova@stuba.sk

Key words

- Housing Cost Overburden Rate,
- Random Forest,
- SHAP Analysis,
- Thresholds,
- Housing Policy.

1 INTRODUCTION

As a part of the construction sector, residential construction represents a key component of national economies with a significant multiplier effect that goes beyond the construction sector itself. International studies (e.g., the OECD or Euroconstruct) estimate this multiplier to be in a range of 1.5 to 2.5, depending on the national context, the type of construction, and the investment regime. It is one of the most interconnected segments of an economy, the activity of which also influences the broader real estate market, financial flows, consumption, employment and regional development. In this context, housing construction is not only a response to demographic and market impulses, but is also an important factor in retrospectively shaping housing affordability and stress caused by household costs. The construction sector provides the physical infrastructure for housing, whereas innovation and productivity in the construction sector also effect the availability, quality, and price of housing.

The deterioration in the availability of housing represents a major socio-economic challenge for contemporary Europe. A confluence of factors, including escalating property prices, stagnant real wages, mounting energy costs, and mounting pressures due to urbanisation, has resulted in a marked increase in the difficulties experienced by individuals in accessing adequate and affordable housing. The impact of these processes is particularly pronounced on young households,

© 2025 The Author(s). This is an open access article licensed under the Creative Commons Attribution 4 .0 License (https://creativecommons.org/licenses/by/4.0/)

families with children, single people, and socially vulnerable groups. In response to this situation, there is an urgent need in European political and research discourse to identify and gain a deeper understanding of the structural factors that determine developments in housing affordability.

2 LITERATURE REVIEW

Housing affordability has become an increasingly analysed phenomenon in the social sciences and urban research over the last decade; it has assumed the status of a key determinant of social justice and territorial cohesion. The literature in the European context shows that this is a dynamic and multidimensional issue that cannot be reduced to a price/income ratio. Instead, it is necessary to take into account the structural, demographic, market and institutional variables that affect the ability of households to secure adequate housing without a disproportionate financial burden. The following review aims to identify the main trends and findings from current empirical research on housing affordability in Europe, with a particular focus on intergenerational, spatial and socio-economic inequalities, as well as identified cost stressors and policy responses.

Empirical research has highlighted significant differences in housing affordability between social groups, regions, and generations. Young households, tenants on the open market, and residents of large cities face a higher risk of stresses from costs, and housing insecurity. Case studies from Amsterdam (Barranco et al., 2021), Greece (Radin, 2024), and Portugal (Fragosi et al., 2023) document how a combination of structural factors and market pressures exacerbate inequalities in access to housing, not only between countries, but also within them.

In addition to housing prices, broader economic and demographic variables such as energy prices, tax policy, interest rates, population density or household size also affect cost stress (Cesarski, 2020; Bahramov, 2024; Biljanovska et al., 2024). Several studies link housing affordability to other indicators of social inclusion and stability, such as migrant unemployment in Italy (Simionescu, 2021). Housing affordability is becoming a social and generational issue in Germany (Espinosa et. al., 2022). Empirical findings (Arneric et al., 2024) confirm the importance of several structural and demographic factors in shaping housing affordability. Some authors highlight the growing importance of the housing-energy cost nexus (Egner and Krappr, 2025; Costarelli, 2025; Murarasu and Bujor, 2022). This complexity highlights the need for linked and cross-sectoral analyses.

Comparisons of housing affordability reveal important differences not only between countries but also within them, e.g., between municipalities in Portugal (Fragosi et al., 2023) or regions in Poland and the Czech Republic (Slavata et al., 2024) and Belgium (Winters, 2025). The dynamics of accessibility are shown to be sensitive to local market and political conditions.

Sabatini (2024) maps policy instruments such as increasing social housing stock, rent regulation, cooperation with NGOs, targeted programmes for vulnerable groups and housing subsidies. The final section highlights the importance of cities as actors in housing policy, draws attention to the need for vertical coordination between levels of government, and proposes a research and policy agenda that responds to the complexity of housing affordability in the urban context.

A synthesis of the available literature confirms that housing affordability is a complex phenomenon affected by the interplay of price, regulatory, demographic and energy factors. Empirical studies clearly show that none of the dimensions of affordability are isolated; in other words, economic parameters are intertwined with social and environmental contexts. For future research and policy practice, this implies a need for a multilevel, sectorally linked approach that responds to the complexity and variability of current housing (in)affordability in Europe.

The paper aims to identify and quantify the key factors influencing the Housing Cost Overburden Rate (HCOR) of households in the European Union countries, through the use of machine learning (Random Forest) methods and interpretive tools (SHAP), and also to examine their interactions in the context of a multidimensional housing affordability framework. The analysis aims to provide an empirically grounded and analytically transparent basis for the formulation of systemic, spatially sensitive and cross-sectorally interconnected housing policies in Europe.

3 MATERIAL AND METHODS

The selection of analytical variables is based on a knowledge base of research on housing affordability in the European context and on a synthesis of recommendations in the relevant professional literature (e.g., Housing Europe, OECD, Eurostat). Indicators capturing key dimensions of supply, demand, prices, incomes, investment behavior, and institutional interventions were taken into account. Emphasis was placed on the availability of data on a national level, the standardization of measurement, and comparability between countries. The variables were processed into a single analytical framework and included in a machine learning model as input factors influencing the target indicator, i.e., the Housing Cost Overburden Rate (HCOR). It expresses the share of the population that spends more than 40% of their disposable income on housing costs. This indicator is an internationally recognized and standardized measure of a household's financial burden, whice is published regularly by Eurostat. Its main advantages are its high degree of comparability across countries, its coverage of a wide range of cost items (including rent, mortgage payments, energy, and ancillary costs), and its independence from the form of housing tenure, which increases its social inclusiveness and analytical relevance. Unlike indicators based solely on property prices or loan amounts, the HCOR captures the social dimension of housing affordability, in particular the impact of housing expenditure on living standards and the economic vulnerability of the population. Given the asymmetric distribution and occurrence of extreme values in some countries, the variable was transformed; and in the analytical part of the work, *HCOR* (household expenditure on housing) always refers to the natural logarithm of *HCOR*, i.e., ln(*HCOR*).

Tab. 1 *Overview of analytical variables, abbreviations and data sources used*

Indicator	Acronym	Data Source
Housing Cost Overburden Rate	*HCOR*	Eurostat
Price to Income Ratio	*PricetoIncome*	Eurostat
Number of building permits per 1,000 population	*Permits*	Eurostat
Annual change in House Price Index (HPI)	*HPIch*	Eurostat
Share of rental housing	*Tenure*	Eurostat
Price to Rent Ratio (capital cities)	*PriceToRent*	Numbeo
Share of short-term rentals per 1,000 population	*STRent*	Horwath HTL and AllTheRooms
Share of unoccupied dwellings	*UnOccupied*	Eurostat
Number of rooms per person	*RoomsPerPerson*	Eurostat
Number of household members	*PersPerHousehold*	Eurostat
Change in the share of the urban population over the last 10 years	*DUrb10*	Eurostat
Migration balance per 1,000 population	*MigSaldo*	Eurostat
Number of immigrants per 1,000 population	*Imig1000Inh*	Eurostat
Share of social housing	*SocHousing*	Housing Europe
Share of public expenditure on housing (% of GDP)	*GovExpHousing*	Eurostat
Gross Fixed Capital Formation in Dwellings (% of GDP)	*FixCapDwell*	Eurostat

In our modeling, we used the latest available values for the analyzed variables.

In the first step, we present descriptive statistics for all the variables analyzed. In the next step, given the relatively low number of observations (27 EU member states) and the relatively high number of explanatory variables, based on a review of the literature, we decided to use the VSURF (Variable Selection Using Random Forests) algorithm to reduce the dimensionality of the input space and select the most relevant predictors for modeling the target variable (Genuer et al., 2010)

To interpret the outputs of the model constructed we used the Random Forest method, which, unlike linear regression models, does not require the independence of variables. For this reason, we did not address the spatial autocorrelation of the variables. In addition, the Random Forest method is capable of modeling nonlinear dependencies. In the next step we applied the SHAP (SHapley Additive exPlanations) method, which allows us to quantify the contribution of the individual explanatory variables to the values predicted. We then computed the average absolute SHAP values, which provide an aggregate measure of the importance of each variable across the dataset. To robustly determine their order of importance, we constructed confidence intervals using bootstrapping, thereby accounting for the variability of the estimates, depending on the data selection. Subsequently, we also constructed a SHAP interaction matrix to identify interaction effects between the explanatory variables in the model (Genuer and Poggi, 2020; Kim and Kim, 2022).

All the calculations were performed in the R computing environment (R Core Team, 2025; Kuhn, 2008; Liaw and Wiener, 2002; Wickham et al., 2023; Wickham,2016; Wickham et al., 2024; Slowikowski, 2024; Molnar et al., 2018; Genuer et al., 2022).

4 RESULTS AND DISCUSSION

Housing affordability varies considerably across the EU. The following analysis aims to identify the factors that influence it and to assess the extent of their impact.

4.1 Spatial differentiation of the cost burden of housing expenditure

From the perspective of the analytical work, this development is important because it reflects the combined effects of structural, market and policy factors that affect housing affordability. Between 2013 and 2024, a slight decline in the housing cost overburden rate, i.e., from 11.6% (2013) to 8.2% (2024) could be observed in the EU. The decline until 2020 (minimum 7.8%) was linked to favourable economic conditions and available credit. After 2021, the trend stopped due to the pandemic, rising energy prices, and inflation, with a slight decrease again in 2024, which may signal stabilisation.

The median housing overburden rate in the EU fell from 8.9% (2013) to 5.6% (2020), which was followed by a slight increase and a decline again in 2024. Greece has long been showing outliers (up to 45.5% in 2015), despite a decline

after 2016. However, high burdens persist, even in strong economies such as Germany, Denmark and the Netherlands. These countries have a high cost of living and often a well-developed rental sector, but despite a strong social system, they can be vulnerable groups in the housing market.

The spatial differentiation highlights the profound differences in housing affordability across the EU (Fig. 2). The results confirm the need for a context-sensitive housing policy that takes into account local market conditions, household structure, and housing tenure. Greece stands out dramatically with by far the highest HCOR (28.9%). Central Europe shows intermediate values, which may reflect a combination of rising housing prices and insufficiently flexible support systems. The lowest HCOR values are found in Cyprus (2.4%), Croatia (3.7%) and Slovenia (3.8%), i.e., countries where, either owner-occupation, historically high levels of owner-occupation and/or culturally shared households are prevalent, which may reduce costs.

Slovakia has a relatively high level of household housing expenditure burden, with Austria narrowly outperforming Austria on this indicator. However, when assessing the standard of living, it is necessary to analyse the broader structure of consumption expenditure, as housing costs represent only one of its components.

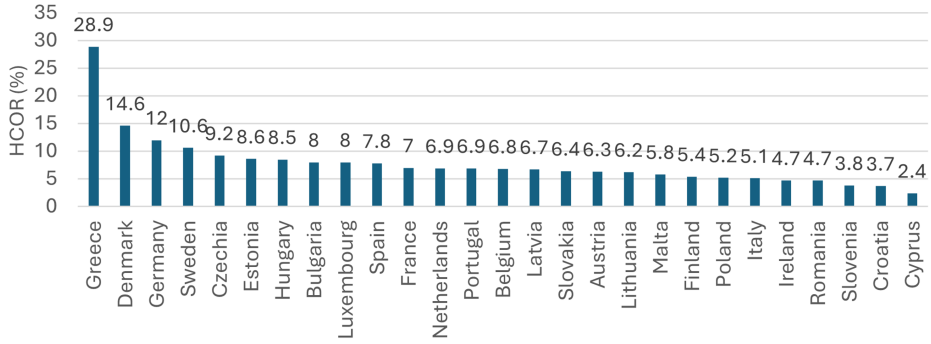

Fig. 1 *EU countries' housing cost overburden rates in 2024*

4.2 Structuring determinants of housing availability according to their function in an analytical model

When conducting analytical research on housing availability, the way variables are organized plays a key role in interpreting results and identifying mechanisms that influence market developments. Such a structure not only facilitates their theoretical anchoring in a broader context but also allows for a more accurate interpretation of their direct and indirect impact on predictive modelling.

Systemic factors are at the heart of the housing affordability issue; they are variables that describe the underlying balance between supply and demand, price and income levels, and determination of the structural risk of unaffordability. Their effects tend to be long-term, and relatively stable; they determine the baseline level of housing affordability.

Dynamic demand factors capture changes in housing demands; they can result from population movements, migration, demographic changes or specific social phenomena. They are variables that reflect shorter-term fluctuations in demand and the spatial distribution of the population, which can lead to regional pressures on the housing stock.

Moderating factors represent interventions by public administrations, institutions, and regulatory mechanisms that influence the extent to which systemic or demand factors translate into real impacts on households. These interventions can be either dampening or reinforcing.Financial and investment factors reflect market expectations, investor motivations, and speculative behaviour in the housing market. They can distort the perception of housing as a social need and contribute to its transformation into a financial asset, which can affect the price level and affordability of housing for the general population.

The multilevel breakdown of factors allows for more sophisticated modelling, in which not only the direct effect of individual variables can be captured, but also their interactions with each other in the context of housing affordability. The conceptual scheme thus developed promotes transparency in the analytical procedure and provides a framework for interpreting the outputs of the machine learning model.

Based on the VSURF algorithm, the following variables were identified as relevant: *PricetoIncome, PersPerHousehold, Tenure, PriceToRent, Permits* and *DUrb10*.

The standardised house price-to-income ratio (PricetoIncome) is a systemic indicator that captures how the current relationship between house prices and incomes differs from its long-run level. It is calculated as the ratio of the current house price-to-income ratio to its average value over the whole period from 2000 to the latest available data. An indicator value of 100 indicates that the current situation corresponds to the long-term average.

The number of household members (*PersPerHousehold*) is a basic demographic indicator with a direct impact on housing expenditures and their distribution among members. Larger households may spend more on housing in absolute terms, but the costs are shared among more members, which may reduce the level of the financial burden on the individual. Conversely, small or single-person households face a higher risk of becoming overburdened, as all costs are borne by the individual. This indicator thus mediates the differences in economic pressure on housing across household types and is an important socio-economic determinant affecting *HCOR*.

The systemic factor of the share of rental housing (*Tenure*) in the total amount of housing reflects the structure of the housing stock and the institutional conditions of access to

Tab. 2 *Descriptive statistics and the functional classification of variables used in the model*

Acronym	Minimum	Maximum	Average	Median	Standard Deviation	Factor
HCOR	0.88	3.36	1.93	1.92	0.46	target, modelled variable
PricetoIncome	53.19	127.61	96.77	98.56	16.07	Structural
Permits	0.93	15.95	4.52	3.68	3.09	Structural
HPIch	-8.20	11.10	2.19	1.90	4.88	Structural
Tenure	5.70	52.80	25.86	26.60	11.86	Structural
PriceToRent	13.70	37.00	23.67	25.40	6.07	Structural, Financial and Investment
UnOccupied	4.12	40.07	20.61	18.20	9.29	Structural, Financial and Investment
RoomsPerPerson	1.10	2.30	1.66	1.60	0.33	Structural
DUrb10	-0.31	5.55	1.92	2.19	1.39	Dynamic Demand
MigSaldo	-6.20	10.80	1.67	2.20	3.58	Dynamic Demand
Imig1000Inh	1.09	76.42	18.46	15.36	14.56	Dynamic Demand
PersPerHousehold	1.90	3.10	2.33	2.30	0.29	Dynamic Demand, Moderating
SocHousing	0.00	30.00	6.39	3.70	7.35	Moderating
GovExpHousing	0.00	3.94	0.37	0.22	0.71	Moderating
STRent	2.00	136.00	20.67	10.00	26.10	Moderating
FixCapDwell	2.00	8.60	4.30	3.90	1.58	Financial and Investment

housing. In countries or regions with a higher share of rental housing, the housing market tends to be more flexible and open to different social groups, including young, mobile or low-income households.

The ratio of the property price to the annual rent (*PriceToRent*) is an indicator of the structural balance between the ownership and rental segments of the housing market. It is one of the systemic factors. It reflects long-term interactions between supply and demand, as well as the market expectations and investment strategies of the market participants. Therefore, this indicator also acts secondarily, as an investment factor, as it indicates the attractiveness of the housing market in terms of the return on investment. High values of the Price-to-Rent Ratio may motivate investors to make speculative purchases, while low values may indicate a more stable or less profitable market.

The inclusion of the number of building permits per 1,000 inhabitants (Permits) as a systemic factor is methodologically justified because this indicator directly reflects the structural aspect of the housing market, namely, the supply capacity of the residential sector. It is a preliminary indicator of the development of construction, which signals how flexibly the market responds to demand, population growth or price signals. Building permits are therefore an input into the construction process, not just as a reflection of current demand, which makes them a fundamental determinant of housing supply. They also affect price pressure; a lack of new permissions increases tensions in the market, which can lead to rising house prices and in turn reduce housing affordability.

Using the change in the proportion of the urban population over the past decade (*DUrb10*) as an indicator of urbanization is an analytically robust approach to capturing long-term trends in the spatial distribution of a population. This indicator reflects systematic population shifts that fundamentally affect housing market demand; this increase pressure on housing stock in targeted urbanized areas, contribute to rising prices, reduce the affordability of rental housing, and increase the overall financial burden on households. They are among the dynamic demand factors.

The above determinants/factors were used to construct a prediction model using the Random Forest method. This method aims to explain the target variable *ln*(*HCOR*), which is a measure of the excess burden of housing expenditures on households.

The number of randomly selected predictors (*mtry = 1*) and the number of trees (*ntree = 1350*) were optimized using 5-fold cross-validation with the caret package. The resulting model achieved a coefficient of determination $R^2 \approx 0.7154$, which indicates the relatively high ability of the model to capture the excessive financial burden on households due to housing expanses.

Overall, despite the relatively small number of measurements (n = 27 countries) and the limited number of features (6), the model provides a robust and well-interpretable

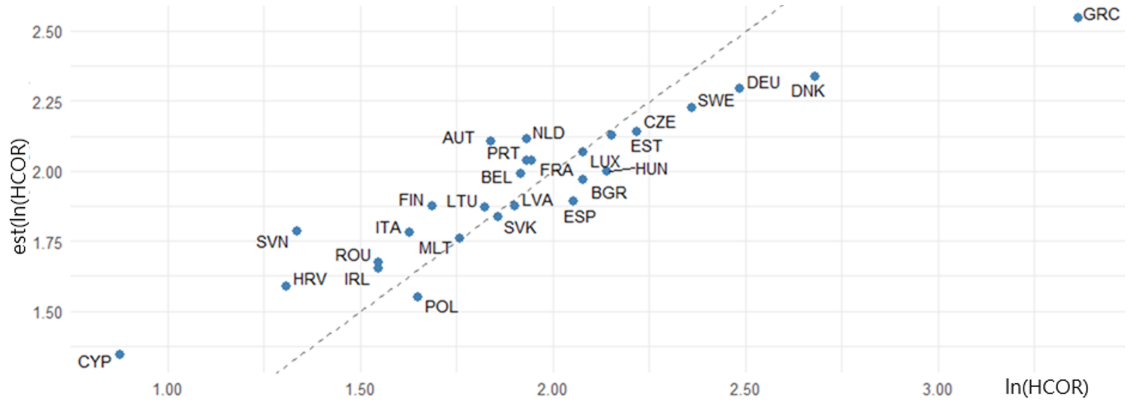

Fig. 2 *Predicted vs. actual values of ln(HCOR) for EU countries (2024)*

prediction of the level of the burden of housing costs in the European area.

To interpret the model, we used the SHAP (SHapley Additive exPlanations) method, which quantifies the contribution of individual variables to specific predictions while maintaining an additive consistency. SHAP values quantify the direction and magnitude of the effect of individual variables on the model prediction. Positive SHAP values contribute to a higher value of the modelled variable (ln(*HCOR*)), while negative SHAP values decrease it. This interpretation allows us to distinguish between factors that increase the risk of the stress from costs and those that mitigate it. Figures 3 to 8 display the SHAP dependence plots, which illustrate the nonlinear and individually specific relationships between the values of the input variables and their impact on the prediction of ln(*HCOR*). These visualisations allow us to identify critical thresholds at which the effect of a variable changes significantly, for example, from protective to risky. In this way, they provide valuable insights into which factors and what intervals contribute most to housing cost stress in individual EU Member States.

PricetoIncome has a nonlinear, strongly increasing impact on the share of the population experiencing stress from costs. The key point is the threshold around 96, above which the risk begins to increase significantly. This graph clearly illustrates the countries for which this indicator is one of the main factors contributing to the deterioration of housing affordability. Slovakia is approaching this threshold. It can be said that France is at the threshold beyond which stress from housing costs will begin to increase.

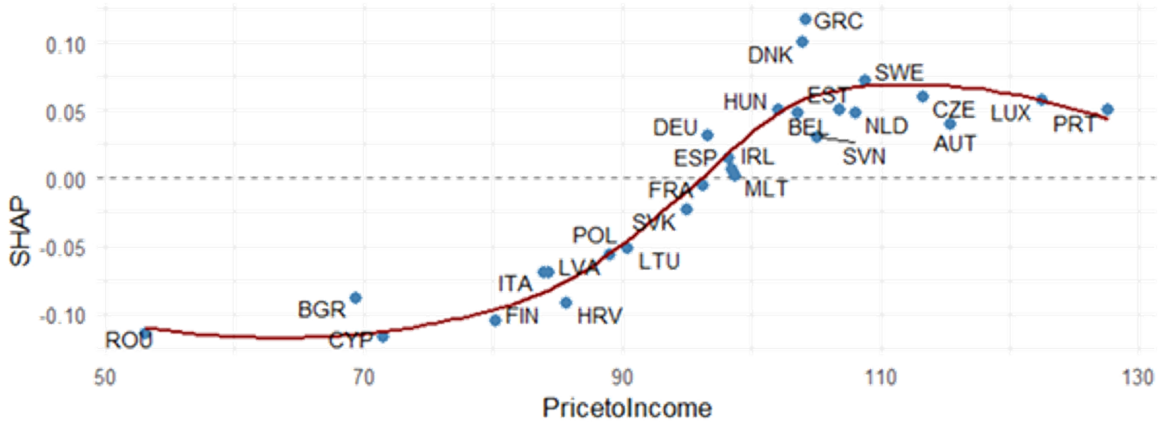

Fig. 3 *SHAP dependence plot for PricetoIncome*

The *PersPerHousehold* variable has a significant indirect effect; the more people living in one household, the lower the model is estimate of the stress from costs. This result can be interpreted to imply that shared housing lowers the unit housing costs of households, thereby increasing housing affordability. The graph also helps to identify the threshold (around 2.4) at which the effect breaks from negative to positive. The *PersPerHousehold* variable shows a positive effect for Portugal, Spain, Romania, Cyprus, Ireland, Croatia, Poland and Slovakia, where multigenerational or multi-person households are historically and culturally typical. This social context may contribute to lower rates of being overburdened by housing costs (HCOR), as the sharing of multiple household members allows for the spreading of housing costs while reducing individual burdens. This type of housing can also buffer the impacts of housing unaffordability, particularly for younger or economically weaker groups.

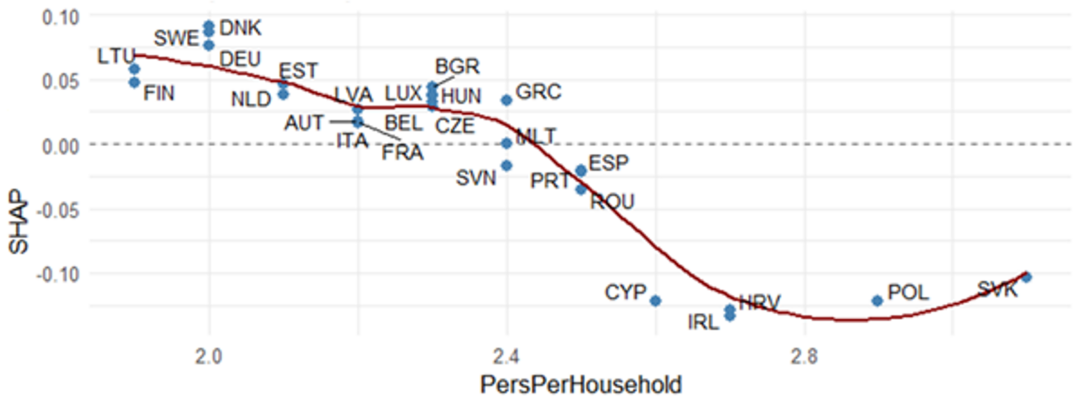

Fig. 4 *SHAP dependence plot for PersPerHousehold*

The proportion of people paying in market rent (*Tenure*) has emerged as a significant determinant of the degree of the stress costs. SHAP analysis confirmed the non-linear nature of this relationship: at lower values (up to around 30%), the effect is negligible, but increases significantly beyond this threshold. This break suggests that a high share of market rental housing, which is typical of, for example, Germany or Denmark, increases the risk of excessive financial stress, probably due to price stress and the limited availability of regulated rents. The results point to the need for intervention measures in countries with a large rental sector. Slovakia, as well as all post-socialist countries plus the Southern European countries, are at the lower end of the range (up to around 30 %), which indicates that market rental housing does not prevail here.

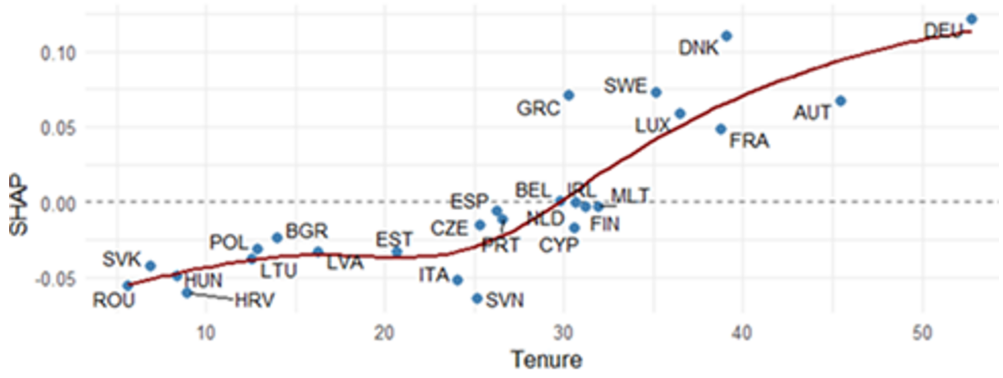

Fig. 5 *SHAP dependence plot for Tenure*

The *PriceToRent* variable has a significant impact on the estimated share of the population under stress from costs; the higher this ratio, the greater the pressure on households. The graph shows that a change from low to high *PriceToRent* values leads to a clear break in the impact around the value of 25. Once this threshold is exceeded, housing affordability deteriorates rapidly, which is clearly reflected in the model through the rising SHAP values. This raises an important policy warning: markets with high *PriceToRent* values can generate unsustainable housing costs for a large part of the population.

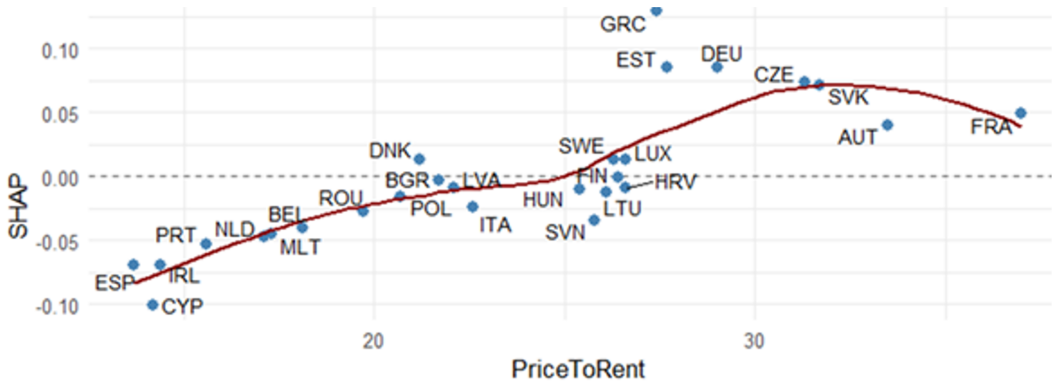

Fig. 6 *SHAP dependence plot for PriceToRent*

The *Permits* variable has a significant indirect non-linear effect, i.e., the more building permits per 1,000 inhabitants are issued, the lower the model is estimate of stress from costs from the key point which is approximately 4.5 permits per 1,000 inhabitants. Above this value the *Permits* impact on SHAP values are negative, which indicates that higher building intensity begins to have a dampening effect on the predicted cost stress. This effect is most pronounced in countries such as Cyprus and Malta, where high permits are likely to alleviate market stress and contribute to greater housing affordability. The finding suggests that an effective housing policy needs to exceed a certain threshold of construction to be able to buffer stress from costs. Low and uncoordinated construction may even temporarily increase pressure on the market, while sufficiently intensive construction acts as a stabilising factor.

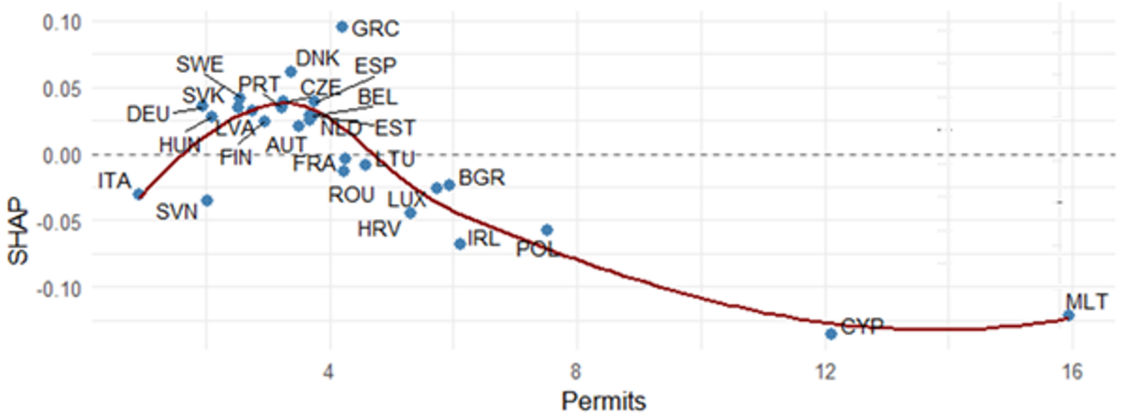

Fig. 7 *SHAP dependence plot for Permits*

If the increase in the urban population share (*DUrb10*) exceeds 1.7 percentage points over the most recent decade, this signals increased demand in urbanised areas. This demographic pressure often translates into rising housing prices, reduced rental availability and increased financial burdens on households. The urbanisation trend thus emerges as a significant dynamic demand factor that influences housing affordability. Slovakia has not reached the 1.7 percentage point threshold, where indicates that the *DUrb10* variable has a positive effect on the *HCOR*. The Czech Republic is approximately at the 1.7 threshold. Countries that have seen their urban population share grow by more than 1.7 percentage points over the last ten years also have a higher level of household expenditure on housing.

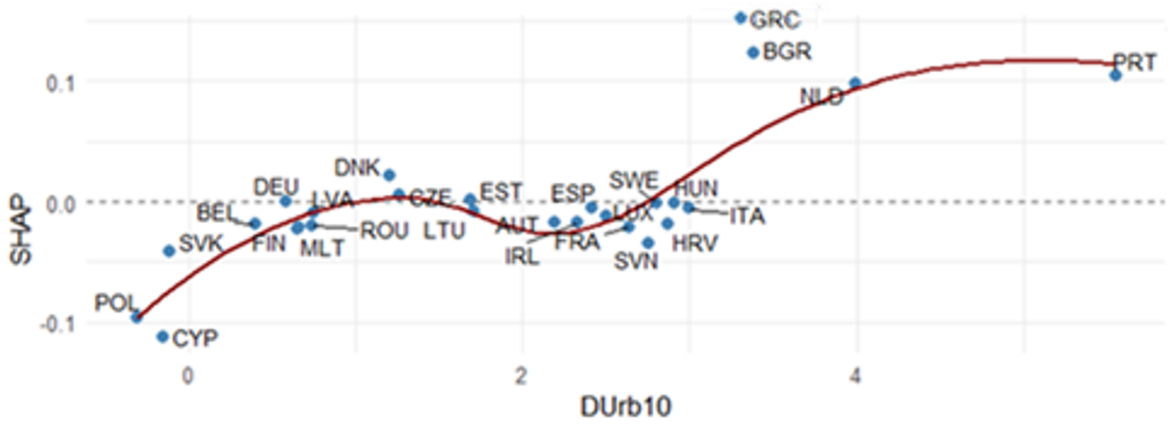

Fig. 8 *SHAP dependence plot for Permits*

The thresholds specified for the variables analyzed are only indicative, and their integration into policies depends on the type of variable.

Figure 8 shows the mean absolute SHAP values for the six explanatory variables, which are supplemented by the 95% confidence intervals obtained by bootstrapping. These values quantify the contribution of each variable to the prediction of the logarithm of the HCOR indicator under the Random Forest model. Importantly, the confidence intervals of all the variables overlap significantly with each other, i.e., they have a non-empty intercept. This means that the differences between the mean absolute SHAP values of the individual variables are not statistically significant. In other words, it cannot be said with sufficient confidence that any one variable has a statistically significantly higher or lower impact on the phenomenon modelled than the others. Analytically, this implies that all the selected variables (*PricetoIncome*, *PersPerHousehold*, *Tenure, PriceToRent*, Permits and *DUrb10*) should be considered comparatively equivalent in terms of their contribution to explaining the variability of the log-transformed *HCOR* (Housing Cost Overburden Rate) indicator when interpreting the results. This equivalence strengthens the robustness of the model and supports a multidimensional view of the determination of the stress of housing costs.

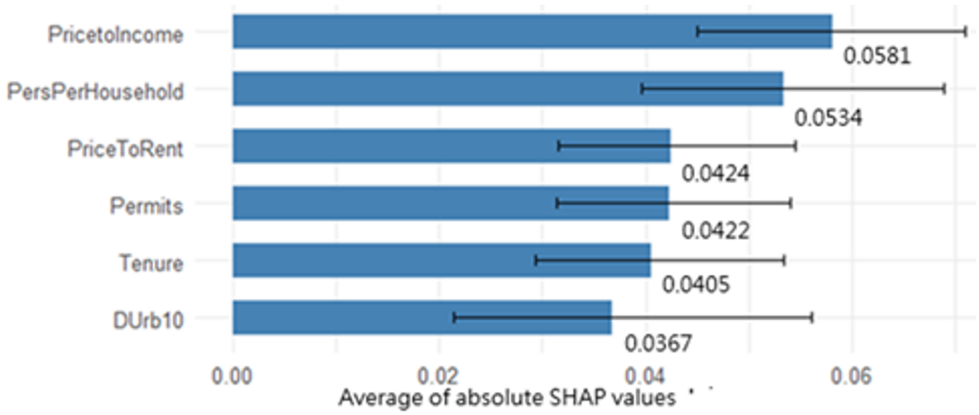

Fig. 9 *Average absolute SHAP values with 95% confidence intervals for ln(HCOR) prediction*

Factors do not affect the value of ln(*HCOR*) in isolation, but often act in correlation with each other. Therefore, in the next step, we were also interested in the strength of their interactions, i.e., the extent to which the effect of one variable varies with the value of another. This analysis provides a deeper insight into the complex structure of the relationships between the determinants of the stress of housing costs.

An analysis of SHAP interactions revealed that the effect of several factors on ln(*HCOR*) is not isolated but depends strongly on the values of the other variables. The strongest interaction was identified between the proportion of people living in market rented accommodation (Tenure) and the PriceToRent indicator. This juxtaposition of a high proportion of rented housing and relatively high rents points to market tensions in an environment where there is an insufficient supply of affordable or social housing.

The interaction between *Tenure* and the number of Building Permits (*Permits*) suggests that market rents are extremely sensitive to the level of supply, i.e., without sufficient construction, the risk of the unaffordability of housing increases more rapidly.

The interaction observed between *PricetoIncome* and *PersPerHousehold* also suggests that the negative impact of low affordability of housing is amplified in households with more members, where there is a higher need for sharing expenditures.

From a policy perspective, these results support the need for an integrated approach to housing regulation. Effective interventions should not be unilateral (e.g., construction only or rental support), but must reflect the interconnectedness of the market, and the demographic and structural aspects of housing. Such a policy design should be multi-level, spatially differentiated and able to respond in a targeted way to the combined sources of the stress of household costs.

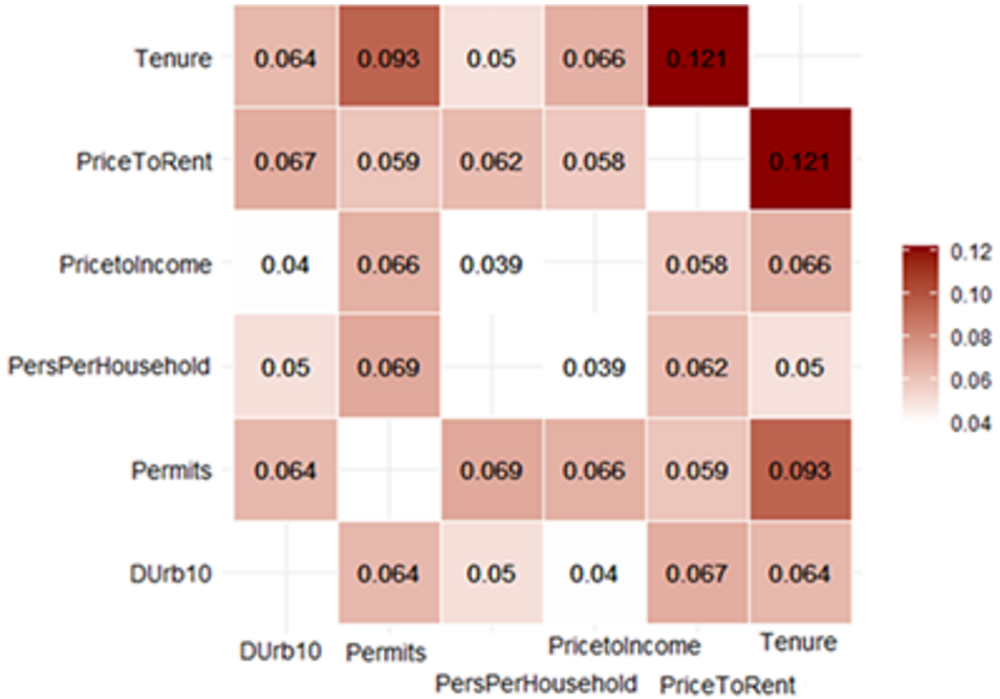

Fig. 10 *SHAP Interaction Matrix of Predictor Variables*

The observed interactions observed between the variables confirm that the effectiveness of individual housing policy instruments, e.g., construction, rent regulation or support for social housing, is not universal, but depends on the context of the broader market and demographic conditions. Effective housing policy therefore requires an integrated approach that links the instruments of construction, affordable rent, regulation and demographic balance as an interlinked system. At the same time, the effects of interaction, which can vary significantly between countries and regions, highlight the need for a spatially sensitive and differentiated policy design. Analytical tools using SHAP interactions offer a robust framework for assessing these interconnected phenomena and can serve as a basis for evidence-based, contextually- tailored housing policy design.

5 CONCLUSIONS

The problem of the stress of costs cannot be grasped only through traditional economic categories, but requires a multidimensional approach that combines market structures, demographic realities, and public policy instruments into a single interpretive framework. The relationships between variables and cost stresses are not linear, which confirms the need to use models capable of capturing non-linear effects such as Random Forest in combination with SHAP analysis.

An important methodological contribution of this analysis is the systematic breakdown of the explanatory variables into analytical categories, namely, systemic, dynamic demand, moderating and investment/financial factors. This framing allows for a better understanding of the mechanisms through which each variable influences housing affordability, while providing an analytical structure suitable for the formulation of comprehensive policies.

The development of housing is closely linked to the construction industry, but is also linked to a number of other disciplines and sectors, from planning to the use of housing. From the perspective of the real estate market, housing construction functions as a primary supply mechanism that directly influences the balance between supply and demand. Its insufficient volumes, which is often the result of regulatory, financial or spatial constraints, leads to rising house prices and rents, thereby increasing the cost burden on households and worsening housing affordability. This effect is particularly pronounced in metropolitan areas and in countries with a high share of market rents.

Another important methodological contribution of the analysis is the identification of thresholds at which the direction or intensity of the affect of individual variables on the log-transformed HCOR changes. These points represent analytically based thresholds that can serve as reference points for targeting public interventions. They can serve as an important analytical basis for the design of public policies aimed at reducing cost stress, in particular when defining target groups, defining threshold warnings and setting regulatory or support mechanisms.

The findings highlight the need for a coherent, multi-level and cross-sectoral approach to housing policy that reflects not only market mechanisms (e.g., price-to-income ratios) but also structural and socio-demographic contexts (e.g., household size and composition, housing tenure, or the share of the rental sector). As none of the variables showed a statistically dominant effect, it is clear that unilateral measures such as subsidies, tax breaks or deregulation cannot by themselves solve the problem of housing affordability.

The new Energy and Housing portfolio in the European Commission presents an opportunity to link the social dimension of housing more systematically with the objectives of climate transformation and energy security. At the same time, it can contribute to deepen support of the Member States in formulating inclusive, integrated, and data-driven housing strategies

REFERENCES

ARNERIĆ, Josip; KIKEREC, Matej; SKOKO, Branimir. *Trends and drivers of housing affordability in the EU: Insights from panel data analysis.* Online. Croatian Review of Economic, Business and Social Statistics. vol. 10 (2024), No. 2, pp. 49-62. Available at: https://doi.org/10.62366/crebss.2024.2.004. [viewed 2025-03-10].

BAHRAMOV, Fuad. *Assessing housing affordability: Analysis of property tax systems in European countries.* Online. The Annals of the University of Oradea. Economic Scientes. vol. 33 (2024), No.1, pp. 237-245. Available at: https://anale.steconomiceuoradea.ro/en/wp-content/uploads/2024/11/AUOES.July_.2024.27.pdf. [viewed 2025-03-10].

BARRANCO, Ricardo; Chris JACOBS-CRISIONI and Sjoerdje VAN HEERDEN (2021). *House price dynamics and affordability in the city of Amsterdam.* Publications Office of the European Union, Luxembourg. ISBN 978-92-76-43360-6.

BILJANOVSKA, Nina; Chenxu FU and Deniz IGAN. *A new dataset on housing affordability.* Online. SUERF Policy Brief, no. 768 (2024). Available at: https://www.suerf.org/wp-content/uploads/2024/01/f_74208afa67a707e43206c70ed62f5642_82135_suerf.pdf. [viewed 2025-03-10].

CESARSKI, Maciej. *Housing cost overburden rate of household and their types in Poland: 2007–2018.* Online. Polityka Społeczna, vol. 558 (2020), No. 9, pp. 25-30. Available at: https://doi.org/10.5604/01.3001.0014.4280. [viewed 2025-02-15].

COSTARELLI Igor (2025). *Housing Costs Development in Italy: Implications for Tenure Inequalities.* In: EGNER, Björn and Max-Christopher KRAPP (eds.).. Housing in Crisis. Palgrave Macmillan, Cham, p. 1–12. ISBN 978-3-031-87266-2.

EGNER, Björn and Max-Christopher KRAPP (2025). *Introduction: Housing in Crisis Across Europe.* In: EGNER, Björn and Max-Christopher KRAPP (eds.). Housing in Crisis. Palgrave Macmillan, Cham, pp. 1–12. ISBN 978-3-031-87266-2.

ESPINOSA URIEL, Pablo; Patricia PELAYO-ROMERO, Maddalena MARTINI, Andreas A. JOBST, Eric BARTHALON, Nihal TEMUEGE and Christian JANSSEN. *House of Cards? Perspectives on European Housing.* Online. Social Science Research Network Working Paper, 1 Dec. 2022. Available at: http://dx.doi.org/10.2139/ssrn.4686629. [viewed 2025-04-12].

FRAGOSO JANUÁRIO João; Carlos OLIVEIRA CRUZ; Humberto VARUM and VÍTOR FARIA E SOUSA. *Is housing becoming less affordable? A study of affordability in the Portuguese housing market.* Property Management, vol. 41(2023), No. 5, pp. 698-728. Available at: https://doi.org/10.1108/PM-08-2022-0059. [viewed 2025-02-15].

GENUER, Robin; Jean-Michel POGGI and Christine TULEAU-MALOT (2010). *Variable selection using random forests.* Online. Pattern Recognition Letters, vol. 31, no. 14, pp. 2225–2236. Available at: https://doi.org/10.1016/j.patrec.2010.03.014. [viewed 2025-01-14].

GENUER, Robin and Jean-Michel POGGI. *Random Forests with R.* Online. Cham: Springer, 2020. Available at: https://link.springer.com/content/pdf/10.1007/978-3-030-56485-8.pdf. [viewed 2025-01-10].

GENUER, Robin; Jean-Michel POGGI and Christine TULEAU-MALOT (2022). *VSURF: Variable Selection Using Random Forests.* Online. R package version 1.2.0. Available at: https://doi.org/10.32614/CRAN.package.VSURF [viewed 2025-01-10].

KIM, Yesuel and Youngchul KIM (2022). *Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models.* Online. Sustainable Cities and Society, vol. 79, article 103677. Available at: https://doi.org/10.1016/j.scs.2022.103677 [viewed 2025-01-10].

KUHN, Max (2008). *Building Predictive Models in R Using the caret Package.* Online. Journal of Statistical Software, vol. 28, no. 5, pp. 1–26. Available at: https://doi.org/10.18637/jss.v028.i05 [viewed 2025-01-10].

LIAW, Andy and Matthew WIENER (2002). *Classification and Regression by randomForest.* Online. R News, vol. 2, no. 3, pp. 18–22. Available at: https://CRAN.R-project.org/doc/Rnews/ [viewed 2025-01-10].

MOLNAR, Christoph; Bernd BISCHL and Giuseppe CASALICCHIO (2018). *iml: An R package for Interpretable Machine Learning.* Online. Journal of Open Source Software (JOSS), vol. 3, no. 26, article 786. Available at: https://doi.org/10.21105/joss.00786 [viewed 2025-01-10].

MURĂRAŞU, Ioan-Cătălin and Răzvan BUJOR. *The Energy Crisis as a Factor of Aggravating Poverty in Bulgaria and Romania.* State's Social Expenditures Evolution. Online. Romanian Economic Journal, vol. 25 (2022), No. 84, pp. 95-103. Available at: https://doi.org/10.24818/REJ/2022/84/08. [viewed 2025-03-10].

RADIN, Ioannis. *Assessing the problem of affordable housing in Greece.* Online. Discover Cities, vol. 1 (2024), article No. 33. Available at: https://doi.org/10.1007/s44327-024-00037-z. [viewed 2025-02-15].

R CORE TEAM (2025). R: *A Language and Environment for Statistical Computing.* Online. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/ [viewed 2025-01-10].

SABATINI, Francesca (2024). *Urban Governance and the Housing Crisis: A Review and Research Agenda.* Online. In: CALABRÒ, Francesco; Livia MADUREIRA, Francesco Carlo MORABITO and María José PIÑEIRA MANTIÑÁN (eds.). Networks, Markets & People 2024. Lecture Notes in Networks and Systems, vol. 1188. Springer, Cham, pp. 121–130. Available at: https://doi.org/10.1007/978-3-031-74716-8_12 [viewed 2025-04-12].

SIMIONESCU, Mihaela. *The status of immigrants on Italian labour market in the context of economic decline: Evidence from survey, macroeconomic and big data.* Online. Economics, Management and Sustainability, 2021, vol. 6 (2021), No. 1, pp. 34-48. Available at: https://doi.org/10.14254/jems.2021.6-1.3. [viewed 2025-03-12].

SLAVATA, David; Eva ARDIELLI and Marta MACIEJASZ. *Comparison of Housing Affordability in Czech and Polish Regions.* Online. Review of Economic Perspectives, vol. 24 (2024), No. 1, pp. 17-36. Available at: https://doi.org/10.2478/revecp-2024-0002. [viewed 2025-03-21].

REFERENCES

SLOWIKOWSKI, Kamil (2024). ggrepel: *Automatically Position Non-Overlapping Text Labels with 'ggplot2'.* Online. R package version 0.9.6. Available at: https://doi.org/10.32614/CRAN.package.ggrepel [viewed 202-01-10].

WICKHAM, Hadley (2016). ggplot2: *Elegant Graphics for Data Analysis.* Online. New York: Springer-Verlag. Available at: https://ggplot2.tidyverse.org [viewed 2025-01-10].

WICKHAM, Hadley; Romain FRANÇOIS, Lionel HENRY, Kirill MÜLLER and Davis VAUGHAN (2023). dplyr: *A Grammar of Data Manipulation.* Online. R package version 1.1.4. Available at: https://doi.org/10.32614/CRAN.package.dplyr [viewed 2025-01-10].

WICKHAM, Hadley; Davis VAUGHAN and Matthias GIRLICH (2024). tidyr: *Tidy Messy Data.* Online. R package version 1.3.1. Available at: https://doi.org/10.32614/CRAN.package.tidyr [viewed 2025-01-10].

WINTERS, Sien (2025). *A Housing Crisis in Belgium: Perception or Reality?* In: EGNER, Björn and Max-Christopher KRAPP (eds.). Housing in Crisis. Palgrave Macmillan, Cham, pp. 13–30. ISBN 978-3-031-87266-2.