

Louvain School of Management

Residential Attractiveness and Housing Prices in Belgian Municipalities: A Composite Indicator Approach

Author: Lisa Dounia
Supervisor: Corentin Vande Kerckhove
Co-supervisor: Braima Mané
Academic Year 2025-2026
Master [120] in Business Engineering, Specialized Finality in Business
Analytics

Declaration Regarding AI Tool Usage in Master's Thesis

During the preparation of this master's thesis, I, Lisa Dounia, utilized the following AI tools for the purposes described below:

1. Academic writing and paraphrasing assistance

ChatGPT (OpenAI) was used to reformulate and paraphrase written passages in order to improve clarity and academic register.

2. LaTeX typesetting assistance

ChatGPT (OpenAI) was used to assist with the typesetting and formatting of the thesis document in LaTeX, including syntax guidance for mathematical expressions, tables, figures, and bibliographic integration.

3. Code assistance and debugging

Claude Code (Anthropic) was used to assist with the structuring, optimisation, and debugging of the Python codebase developed for the data pipeline and empirical analyses.

4. Translation

DeepL Translator was used to assist with the translation between French and English throughout the literature review and writing process.

5. Bibliographic reference formatting

Scribbr and ChatGPT (OpenAI) were used to assist with the formatting of bibliographic references in APA style.

6. Literature review support

NotebookLM (Google) was used to support the consultation and synthesis of academic sources during the literature review phase.

After using these tools, I diligently reviewed and edited all content produced. I take full responsibility for the final content presented in this thesis.

By signing this declaration, I affirm that the content of this master's thesis reflects my original work, augmented by the responsible use of AI.

Date: 28/05/2026

Signature:



Abstract

This thesis investigates the relationship between residential attractiveness and housing prices across Belgian municipalities from both explanatory and predictive perspectives. Residential attractiveness is defined as the capacity of a municipality to attract and retain residents through territorial characteristics likely to influence residential choices.

To operationalise this concept, a Residential Attractiveness Index (RAI) is constructed for 565 Belgian municipalities over the 2015–2024 period. The index combines 21 indicators grouped into eight dimensions: Accessibility, Demographics, Economy, Governance, Urbanism, Housing, Infrastructure, and Social Conditions. The empirical analysis combines econometric and machine learning methods. Cross-sectional relationships are estimated using between estimators with spatial standard errors, while temporal dynamics are analysed through two-way fixed effects models. Predictive performance is assessed using expanding-window cross-validation and hold-out forecasting at horizons of 0, 1, and 3 years with Ridge regression, Random Forest, and XGBoost.

The results highlight a strong explanatory but limited predictive role for residential attractiveness. Cross-sectionally, municipalities with higher RAI scores consistently exhibit higher housing prices, even after controlling for regional and provincial fixed effects. However, within municipalities, year-on-year changes in the RAI are not significantly associated with housing price adjustments. From a predictive perspective, Ridge regression consistently outperforms the non-linear models, while the composite RAI provides only limited forecasting gains above price persistence alone. Models based on the 21 individual indicators systematically outperform the composite-based specifications, suggesting that aggregation into a single score removes part of the predictive information contained in the underlying variables. Predictive performance also deteriorates substantially as the forecasting horizon increases.

Overall, the findings suggest that residential attractiveness is more strongly associated with persistent spatial differences in Belgian housing prices rather than with explaining or predicting their short-term evolution. This thesis contributes to the literature by constructing a municipal-level composite index of residential attractiveness in Belgium, by systematically comparing its explanatory and predictive roles, and by showing that disaggregated indicators outperform the composite in forecasting tasks.

Acknowledgements

I would first like to express my sincere gratitude to my supervisors, Professor Corentin Vande Kerckhove and Mr. Braima Mané, for their guidance and support throughout the production of this thesis.

I would also like to express my gratitude to all the teachers and professors who, from my earliest school years to this very day, encouraged me, believed in my potential, and helped me grow.

I am grateful to all the people who supported me and made these past five years so meaningful. A special thought goes to Nicolas Liva, my study buddy.

I would like to thank my mother, Caroline Guiliang-Toto, who has always seen in me the woman I am becoming and who passed on to me her resilience and perseverance.

Finally, I dedicate this thesis to Lisa Dounia, the one who never stopped dreaming, who always dared, and who always believed she was capable of it, convinced that the frontiers of the Universe were her only limits, because without you, none of this would have been possible.

Contents

1	Introduction	1
2	Literature Review	3
2.1	Concept and Determinants of Residential Attractiveness	3
2.2	Measuring Residential Attractiveness: Composite Indicators	5
2.3	Residential Attractiveness and Housing Prices	7
2.4	Research Gap	8
3	Methodology	10
3.1	Construction of the Residential Attractiveness Index	10
3.1.1	Conceptual Framework and Indicator Selection	10
3.1.2	Data Preprocessing and Distributional Diagnostics	11
3.1.3	Multivariate Structure and Variable Selection	13
3.1.4	Index Construction	13
3.1.5	Robustness check	14
3.2	Econometric Analysis	15
3.2.1	Descriptive diagnostics	15
3.2.2	Econometric Models	16
3.2.3	Robustness Checks	18
3.3	Predictive Modeling	19
3.3.1	Data Splitting Strategy	19
3.3.2	Prediction Tasks	20
3.3.3	Model Specifications	20
3.3.4	Algorithms	21
4	Results	23
4.1	Residential Attractiveness Index	23
4.1.1	Multivariate Analysis	23
4.1.2	Temporal Stability of the RAI	25
4.1.3	Robustness of the RAI	26
4.2	Econometric Results	28
4.2.1	Main Results	28
4.2.2	Robustness Checks	29
4.3	Predictive Results	30
4.3.1	Backtesting	30
4.3.2	Selected models and out-of-sample validation	32
4.3.3	Predictions	32
5	Discussion	34

5.1	Key Insights	34
5.2	Limitations and Directions for Future Research	35
5.3	Managerial and Societal Implications	38
6	Conclusion	40
	References	41
	Appendices	48
	Appendix A: Data Construction and Indicator Selection	48
	A.1 Initial Data Sources	48
	A.2 Definitions of Dimensions and Indicators	49
	A.3 Final Indicator Set	51
	Appendix B: Data Preprocessing	51
	B.1 Correspondence table for municipal mergers	51
	B.2 Missingness patterns and imputation methods	52
	B.3 Distributional diagnostics and transformation examples	53
	Appendix C: Correlation Matrix	58
	Appendix D: Principal Component Analysis	58
	D.1 Eigenvalues and Explained Variance	58
	D.2 Scree Plot	59
	D.3 Cumulative Explained Variance	60
	Appendix E: Missing data treatment for housing prices	60
	Appendix F: Descriptive Diagnostics	60
	F.1 Box–Cox Transformation	60
	F.2 Variance Decomposition	61
	F.3 Cross-Sectional and Within Associations	61
	Appendix G: Econometric Robustness	62
	G.1 Econometric Robustness Checks	62
	G.2 Spatial Dependence	62
	G.3 Alternative Conley Distance Cutoffs	62
	G.4 Rank-Based Within Specifications	62
	Appendix H: Hyperparameter search space	63
	Appendix I: Best Performing Ridge Specifications	63
	I.1 Estimated coefficients	63
	I.2 Model fit statistics	64
	Appendix J: GitHub Repository	64

List of Tables

1	Selected indicators included in the Residential Attractiveness Index	11
2	Expanding-window validation structure (development sample)	20
3	Model specifications	21
4	PCA component matrix	24
5	Rank correlations between the baseline RAI and alternative specifications . . .	27
6	Top 10 municipalities under alternative specifications	27
7	Leave-one-out sensitivity analysis of the RAI	27
8	Econometric model results	29
9	Backtest performance across horizons (development sample, target years 2017– 2023)	31
10	Selected α values for Ridge across backtest folds (development sample, target years 2017–2023)	31
11	Ridge versus OLS across backtest folds (development sample, target years 2017– 2023)	32
12	Hold-out evaluation of selected models (hold-out sample, target year 2024) . . .	32
13	Detailed raw data sources	48
14	Conceptual definition of dimensions	49
15	Definition of indicators	50
16	Final indicator set	51
17	Correspondence table between pre-merger and 2025 NSI codes	52
18	Missingness patterns and imputation methods by indicator	52
19	Eigenvalues and Explained Variance	58
20	Missingness patterns and imputation methods for median housing prices	60
21	Box–Cox estimation results	60
22	Between–within variance decomposition	61
23	Correlation between RAI and $\ln(\text{Price})$	61
24	Summary of robustness checks	62
25	Moran’s I test for spatial autocorrelation	62
26	Conley spatial standard errors under alternative distance cutoffs	62
27	Rank-based robustness checks for within estimates	63
28	Estimated Ridge coefficients by forecasting horizon	64
29	regularization parameters of the selected Ridge models	64

List of Figures

1	Distribution of the Residential Attractiveness Index	25
2	Spatial distribution of the RAI across Belgian municipalities (2024)	25
3	Consecutive year-on-year Spearman rank correlation	26
4	Rank trajectories of five selected municipalities	26
5	Historical and predicted median housing price for selected municipalities.	33
6	Distributional diagnostics and candidate transformations for indicator H1	53
7	Distributional diagnostics and candidate transformations for indicator U2	54
8	Distributional diagnostics and winsorisation for indicator G3	55
9	Distributional diagnostics and candidate transformations for indicator G1	56
10	Distributional diagnostics and candidate transformations for indicator A2	57
11	Correlation matrix of the candidate indicators.	58
12	Scree plot of PCA eigenvalues.	59
13	Cumulative explained variance across principal components.	60
14	Box–Cox profile likelihood and distribution of housing prices	61
15	Associations between RAI and $\ln(\text{Price})$	61

1 Introduction

In the first quarter of 2024, the median price of an attached or semi-detached house in Belgium was 259,500 euros. However, this national figure conceals substantial territorial disparities. The median price reached 499,000 euros in the Brussels-Capital Region, compared with 178,000 euros in Wallonia. At the municipal level, the differences are even more pronounced, ranging from Knokke-Heist to Colfontaine (Statbel, 2024). These figures highlight the existence of strong spatial disparities in housing prices across Belgian municipalities.

One possible explanation for these differences lies in residential attractiveness. In line with the literature, residential attractiveness is defined as the capacity of a place to attract and retain residents (Alexandre et al., 2010; Chaze, 2017; Poirot & Gérardin, 2010). Although housing prices are influenced by broader macroeconomic conditions, they also reflect how households value different locations (Guo & Bhat, 2001; Rosen, 1974). In this context, territorial characteristics such as accessibility, infrastructure, housing conditions or local amenities may shape a municipality's ability to attract and retain residents. Municipalities offering more attractive residential environments may therefore experience stronger housing demand, which can in turn be reflected in higher housing prices. Differences in residential attractiveness may thus contribute to explaining spatial price differentials. Nevertheless, the extent to which this relationship holds empirically, and whether it also contributes to predicting future housing price dynamics, remains uncertain.

Studying residential attractiveness nevertheless raises an important methodological challenge due to its multidimensional nature. Because it encompasses a broad range of heterogeneous territorial characteristics, it cannot be adequately captured by a single indicator. Composite indicators therefore provide a useful framework for synthesising multidimensional information into a single measure.

Against this background, this thesis is guided by the following main research question:

To what extent does residential attractiveness contribute to explaining and predicting housing prices across Belgian municipalities?

A complementary sub-question addresses the measurement dimension:

How can residential attractiveness be operationalised and measured at the municipal level in Belgium?

To address these questions, this thesis constructs a Residential Attractiveness Index (RAI) combining 21 indicators across eight dimensions of local living conditions. The indicators are normalized and aggregated through equal weighting and linear aggregation. The empirical analysis relies on a panel dataset covering 565 Belgian municipalities over the period 2015 to 2024 and combines econometric modelling with predictive analysis across different forecasting horizons.

The remainder of this thesis is structured as follows. The next chapter reviews the literature on residential attractiveness, composite indicators, and their relationship with housing prices, while highlighting the main research gap addressed in this study. Chapter 3 presents the methodology used to construct the Residential Attractiveness Index, together with the empirical framework. The empirical results are then reported in Chapter 4 and discussed in Chapter 5, alongside the main limitations, directions for future research and managerial and societal implications. Finally, Chapter 6 concludes by summarising the main findings and outlining the contributions of the study.

2 Literature Review

This section reviews the existing scientific literature on residential attractiveness and its relationship with housing prices. It is organised around four main objectives. First, it clarifies the concept of residential attractiveness and identifies its main determinants. Second, it discusses existing approaches used to measure multidimensional concepts. Third, it examines how residential attractiveness can be related to housing prices through existing theoretical and empirical contributions. Finally, it identifies the main gaps in the literature, with a particular focus on the Belgian context.

2.1 Concept and Determinants of Residential Attractiveness

Place attractiveness was not initially treated as a central concept in geographical research. For a long time, it was regarded as a secondary notion and was mainly approached through observed territorial flows (Chaze, 2017). However, as processes such as globalisation and decentralisation intensified, territories at multiple spatial scales increasingly came to be seen as being in competition with one another. This gradually shifted attention beyond competitiveness alone towards a broader perspective, in which territories are assessed according to their capacity to attract and retain different kinds of resources (Alexandre et al., 2010). In this context, attractiveness progressively became an important point of reference in both academic research and policy-oriented discussions on territorial development (Alexandre et al., 2010; Chaze, 2017).

Building on this evolution, the literature has sought to clarify the concept of attractiveness. Place attractiveness is commonly defined as the ability of a territory to attract and retain mobile populations and activities (Poirot & Gérardin, 2010). Alexandre et al. (2010) further structure this notion around two complementary distinctions. First, attractiveness combines an objective dimension, related to a place's capacity to attract and retain flows of people, capital and activities, and a subjective dimension, which refers to how desirable the place is perceived to be. Second, attractiveness may be understood either as effective attractiveness, inferred from observed flows, or as potential attractiveness, which refers to latent or anticipatory dynamics revealed through intentions, declared preferences, or signals of possible relocation, even when these have not yet resulted in actual movements. In line with this view, Niedomysl (2006) cautions against inferring attractiveness solely from observed migration outcomes, since demographic changes may reflect structural constraints rather than underlying residential preferences.

Within this broader conceptual framework, the literature also distinguishes different forms of attractiveness depending on the agents and decisions considered. Drawing on several contributions, Chaze (2017) identifies two broad categories: capital attractiveness, which refers to the capacity of a territory to attract firms and investments, and people attractiveness, which refers to the capacity of a territory to attract and retain permanent and temporary residents. Residential attractiveness is therefore positioned as a specific component of people attractiveness, and is defined in this study as the capacity of a place to attract and retain individuals as residents

(Alexandre et al., 2010; Chaze, 2017).

Having clarified the concept, it is important to distinguish residential attractiveness from closely related notions. Residential attractiveness is closely related to, but conceptually distinct from, quality of life. Quality of life refers to individuals' subjective assessment of their life situation as experienced in the place where they live, shaped by personal values and aspirations as well as by the social, cultural and environmental context of that territory (World Health Organization, 2012). Residential attractiveness, by contrast, refers to a territory's ability to attract and retain inhabitants and reflects how places are evaluated in residential choice processes (Alexandre et al., 2010). Unlike quality of life, which focuses on the conditions of life and their perception by residents within a territory, residential attractiveness concerns the characteristics of territories as they are assessed in residential choice processes and in comparison with other places, rather than the well-being of the current population (Chaze, 2017).

Regarding its determinants, no universal consensus exists on a fixed set of factors shaping residential attractiveness. However, empirical studies show convergence at the level of broad domains. Housing and dwelling conditions repeatedly appear as central, encompassing housing supply, structural characteristics and affordability, which together shape access to residential opportunities (Alexandre et al., 2010; Chhetri et al., 2006; Gude et al., 2025; Hofer et al., 2024; Szaja, 2021). The neighbourhood environment forms a second key domain, integrating both physical and social dimensions, including environmental amenities, land use patterns and safety conditions (Alexandre et al., 2010; Guo & Bhat, 2001; Hofer et al., 2024; Szaja, 2021). Accessibility and connectivity also constitute a recurrent category, referring to spatial integration and proximity to everyday activities (Alexandre et al., 2010; Chhetri et al., 2006; Hofer et al., 2024). Local public goods and services, notably education and healthcare, complete this set of commonly identified domains (Alexandre et al., 2010; Guo & Bhat, 2001; Hofer et al., 2024). There is therefore a degree of convergence at the level of analytical categories, but not at the level of exact determinants or universal weights. In the context of the Paris-Dauphine study conducted under the Chaire Ville & Immobilier, Alexandre et al. (2010) propose a framework composed of eleven components: geography and climate, accessibility, demography, political organisation, economy, urbanism, housing, social conditions, infrastructures, culture and heritage, and environment and sustainable development policy. The authors also note that each component may include both objective and symbolic dimensions. Objective dimensions refer to measurable characteristics of a territory, whereas symbolic dimensions relate to perceptions, representations and territorial image. This distinction reflects the idea that residential attractiveness may derive both from material conditions and from more subjective evaluations of place.

Finally, beyond its structural multidimensionality, residential attractiveness is also recognised as preference-dependent. Determinants do not affect all individuals in the same way, and their relative importance varies according to socio-demographic characteristics and household profiles. Hooijen (2021) shows that perceptions of attractiveness depend on individual attributes

and subjective evaluations of place. Empirical evidence also confirms this heterogeneity in Belgium, where residential location preferences vary significantly across household segments defined by tenure status, education level, nationality and household composition (Cockx & Canters, 2020).

2.2 Measuring Residential Attractiveness: Composite Indicators

Because residential attractiveness is multidimensional, it cannot be adequately captured by a single variable. Composite indicators therefore provide an appropriate methodological response, as they make it possible to aggregate multiple individual indicators into a synthetic measure supported by an explicit conceptual framework (OECD et al., 2008). Their main advantage is that they summarise heterogeneous information into a single interpretable measure, thereby allowing relative comparison across territories and over time while retaining the underlying information base (OECD et al., 2008). By reducing a large number of variables into a more manageable form, they also improve readability and facilitate overall assessment.

At the same time, the construction of composite indicators raises important methodological challenges. The final result depends on a series of key choices, including indicator selection, normalisation, weighting and aggregation, all of which involve explicit assumptions (OECD et al., 2008). If these steps are not conceptually justified and transparently reported, the resulting index may produce biased interpretations, conceal weaknesses in certain dimensions through compensatory effects, or oversimplify complex realities. As emphasised in the OECD Handbook, a composite indicator should therefore be understood as a modelling tool whose validity depends on its coherence, robustness and fitness for its intended purpose (OECD et al., 2008).

The literature identifies several alternative methods for constructing a multidimensional score. Principal Component Analysis derives weights from the statistical structure of the data, which can strengthen internal coherence and reduce dimensionality, but it assigns importance according to variance rather than conceptual relevance and assumes full compensability between dimensions (Gómez-Limón & Riesgo, 2009). The Analytic Hierarchy Process explicitly incorporates expert judgement through hierarchical comparisons, which improves normative transparency, but it also introduces subjectivity into the weighting process and generally relies on additive aggregation (Gómez-Limón & Riesgo, 2009). Distance-based multi-criteria approaches evaluate each unit relative to an ideal benchmark and allow partial control over compensation, although the calibration of this compensation parameter remains methodologically open (Gómez-Limón & Riesgo, 2009). Optimisation-based models, including DEA-type frameworks, determine weights endogenously through mathematical programming, thereby reducing arbitrary assignment, but may suffer from limited discrimination power (Peng et al., 2017). Taken together, these different approaches reveal a trade-off between statistical objectivity, normative judgement and interpretability, which makes the choice of construction method a substantive methodological decision rather than a neutral technical step.

Notwithstanding these methodological challenges, several institutional and academic initiatives have operationalised residential attractiveness or closely related concepts through multidimensional frameworks combining thematic dimensions and indicators.

At the international level, the OECD Regional Attractiveness Compass provides a comprehensive framework for profiling territorial attractiveness across multiple domains reflecting a region's ability to attract investors, talent and visitors (Ryu et al., 2024). Rather than producing a single synthetic index, the Compass adopts a multidimensional approach, offering a set of composite indicators that each capture a specific dimension of attractiveness. This structure highlights the complexity of the concept and avoids reducing it to a single aggregate measure. At the European level, the EU Regional Competitiveness Index 2.0 follows a more aggregative approach, combining multiple indicators into a single composite index designed to assess how attractive regions are for firms and residents (Dijkstra et al., 2023). However, its primary focus remains on competitiveness, meaning that residential attractiveness is only captured indirectly and not as a central analytical construct.

In the Belgian context, several composite frameworks capture dimensions that are closely related to residential attractiveness, although they are generally applied at broader spatial scales or developed for different analytical purposes. The OECD Regional Attractiveness Compass, for instance, is implemented at the level of 44 territorial units corresponding to OECD TL3 regions (Fadic et al., 2019; Ryu et al., 2024), while the EU Regional Competitiveness Index is available at the provincial level. At a more disaggregated scale, the Indice des Conditions de Bien-Être (ICBE), developed by the Walloon Institute for Evaluation, Foresight and Statistics (IWEPS), constitutes a rare example of a composite index applied at the municipal level in Belgium (IWEPS, 2015). Although explicitly framed as a measure of quality of life, it incorporates several dimensions that are directly relevant to residential choice, such as material conditions, access to services and environmental quality, and therefore illustrates the feasibility of composite measurement at this scale.

Beyond Belgium, several international and academic initiatives show that residential attractiveness can be explicitly measured at finer spatial scales. The UBS Residential Attractiveness Indicator ranks Swiss municipalities according to their attractiveness as places to live (Hofer et al., 2024), capturing residential appeal at the municipal level while presenting results within larger regional groupings. Academic contributions provide further evidence in this direction. For example, Szaja (2021) applies the Vector Measure Construction Method to construct a synthetic indicator ranking municipalities in the West Pomeranian Voivodeship in Poland according to residential attractiveness, enabling both hierarchical comparison and territorial classification. Similarly, the Dutch Leefbaarometer, developed by the Dutch Ministry of the Interior, provides a detailed and fine-grained measure of residential liveability at a very local scale (Mandemakers et al., 2021). However, despite this growing body of work, no measure of residential attractiveness at the municipal level has been identified in the Belgian context.

2.3 Residential Attractiveness and Housing Prices

While residential attractiveness concerns how individuals evaluate places in the context of residential choices, housing prices capture how territorial characteristics are reflected in the market. Empirical evidence points to substantial spatial heterogeneity in housing prices. In Belgium, housing prices vary not only across the Brussels-Capital, Flemish and Walloon Regions, but also across municipalities within each region (De Bruyne & Van Hove, 2006; Statbel, 2024). These variations are associated with local factors such as income levels, demographic dynamics, housing stock characteristics and geographical position, especially proximity to major economic centres, as well as differences in public goods, fiscal contexts and natural amenities (Hilber, 2011).

Such spatial differentiation is commonly interpreted through the mechanism of capitalisation, whereby the value households assign to local attributes becomes embedded in property prices (Hilber, 2011). This reasoning is rooted in the hedonic pricing framework developed by Rosen (1974), which views good as a bundle of characteristics whose implicit prices are revealed through market equilibrium. In this context, households do not only pay for the dwelling itself, but also for the set of local conditions associated with its location, such as accessibility, infrastructure, amenities or environmental quality. If municipalities differ systematically in these territorial characteristics, and if households value these differences, housing prices may adjust accordingly.

At the same time, interpreting housing prices as a direct measure of territorial characteristics requires caution. Housing markets operate as spatial equilibria in which households sort across locations according to their income, preferences and constraints (Guo & Bhat, 2001; Rosen, 1974). As a result, observed prices reflect not only the value of local attributes, but also the willingness and ability of different population groups to pay for them. In addition, some territorial characteristics may themselves evolve in response to housing market dynamics, implying that local attributes are not always fully exogenous (Finlay & Williams, 2024; Hilber, 2011).

Housing prices are also shaped by broader spatial and structural mechanisms. Geographical and regulatory constraints may limit housing supply and amplify price responses to demand (Hilber, 2011). Moreover, municipalities are not isolated markets, as property values are influenced by proximity to employment centres and neighbouring territories (De Bruyne & Van Hove, 2006). Spatial disparities in housing prices should therefore be understood as the outcome of interacting economic, territorial and spatial processes rather than as the mechanical capitalisation of isolated local attributes.

Against this background, the relationship between residential attractiveness and housing prices remains difficult to characterise in the literature. While both concepts are central to the analysis of spatial disparities, their interaction is rarely examined in a direct and integrated way. In particular, an important difficulty lies in the fact that housing prices are closely related to residential attractiveness while also evolving as a function of it. This creates ambiguity in the

interpretation of their relationship and helps explain why the two concepts are often not explicitly linked within a unified empirical framework. The most explicit discussion of this issue is provided by Alexandre et al. (2010), who analyse residential attractiveness across French urban agglomerations. In their study, attractiveness is measured through migration dynamics and explained using a large set of territorial characteristics. However, housing prices are not included in the econometric model. The authors explicitly justify this choice by highlighting a fundamental analytical difficulty. While housing prices can influence residential attractiveness by attracting or repelling households depending on their level, they also evolve as a function of the very attractiveness they are supposed to reflect. Housing prices are therefore potentially both a cause and a consequence of residential attractiveness, which makes their integration difficult from an econometric perspective. As a result, their inclusion is left as a direction for future research rather than being directly addressed in the model. Although the relationship between residential attractiveness and housing prices is rarely examined directly within a unified framework, several studies nevertheless analyse how specific territorial characteristics are associated with local housing prices. For example, Öner (2017) examine Swedish municipalities and estimate the effect of retail accessibility on housing prices using panel data and fixed effects, while controlling for socio-economic and fiscal characteristics. Similarly, De Bruyne and Van Hove (2006) show, in the case of Belgian municipalities, that housing prices are influenced by distance and travel time to major economic centres, beyond income and real estate characteristics. While these studies do not explicitly mobilise the concept of residential attractiveness, the territorial dimensions they analyse, such as accessibility or proximity to employment centres, correspond to components commonly associated with it. The literature therefore provides evidence that territorial dimensions commonly associated with residential attractiveness contribute to explaining spatial housing price disparities, but examines these dimensions separately rather than within a unified multidimensional framework.

2.4 Research Gap

The literature on territorial attractiveness provides a solid conceptual and empirical foundation for understanding how places attract and retain populations. It highlights the multidimensional nature of residential attractiveness, and composite indicators have emerged as a relevant methodological approach to synthesise these different dimensions into a single measure.

However, residential attractiveness remains only partially developed as a distinct empirical concept. In many studies, it is not directly measured, but rather embedded within broader notions such as quality of life or territorial competitiveness. As a result, it is often approached indirectly, without being operationalised as a dedicated measure.

This limitation is particularly evident at the empirical level. While several composite indices include dimensions related to residential attractiveness, they are generally constructed for different purposes or at more aggregated spatial scales. In the Belgian context, existing indicators

mainly operate at the regional or provincial level, and no composite index specifically designed to measure residential attractiveness at the municipal level has been identified. This constitutes an important gap, as municipalities represent a key level at which residential decisions are made and housing market differences emerge.

In addition, the relationship between residential attractiveness and housing prices remains insufficiently explored. While individual territorial characteristics have been linked to housing prices, residential attractiveness has not been directly examined in relation to housing price differences across Belgian municipalities.

This thesis aims to address these gaps by constructing a composite Residential Attractiveness Index specifically designed for Belgian municipalities. The objective is to operationalise residential attractiveness as the capacity of a municipality to attract and retain residents through a multidimensional set of local characteristics. The analysis then examines the extent to which this index contributes both to explaining spatial differences in housing prices and to predicting future price dynamics across municipalities. By doing so, this thesis contributes to the literature in two main ways. First, it develops a composite index specifically designed to measure residential attractiveness at the municipal level in Belgium, whereas no such index has been identified in the existing literature to the best of our knowledge. Second, it extends the analysis of territorial attractiveness beyond explanatory relationships by also assessing its predictive contribution to housing prices. The next chapter presents the methodological framework used to construct the index and evaluate its empirical relevance.

3 Methodology

This section presents the methodology used to construct the Residential Attractiveness Index (RAI) and evaluate its relationship with housing prices. It proceeds in three steps, the construction of the RAI as a composite indicator, the analysis of its explanatory power using cross-sectional and panel models, and a dedicated predictive framework to assess its forecasting performance.

3.1 Construction of the Residential Attractiveness Index

The construction of the Residential Attractiveness Index follows five main steps, conceptual framework and indicator selection, data preprocessing, multivariate diagnostics, index construction, and robustness analysis.

3.1.1 Conceptual Framework and Indicator Selection

The Residential Attractiveness Index (RAI) conceptualises residential attractiveness as the capacity of a municipality to attract and retain residents through the local conditions it offers. The resulting index provides a general representation of residential attractiveness rather than a preference-specific measure targeting particular household groups. The aim is to construct a relative synthetic measure for 565 Belgian municipalities over the period 2015–2024. Residential attractiveness is modelled as a multidimensional phenomenon structured around distinct but interrelated dimensions reflecting the main local conditions that shape residential location choices. Building on the framework proposed by Alexandre et al. (2010), the structure is adapted to the Belgian municipal context by excluding dimensions considered less relevant at this scale and introducing a governance dimension to better capture institutional and policy-related aspects. This leads to a specification based on eight dimensions: accessibility, demography, economy, urban form, housing, social conditions, infrastructure, and governance.

Indicators are selected based on explicit criteria, including theoretical relevance, consistency with the dimensional structure, municipal-level interpretability, non-redundancy, and clear directionality. Practical constraints related to data availability, reliability, and cross-municipal comparability are also taken into account. The selection of indicators is guided by the objective of measuring realised territorial characteristics. In line with the composite indicator literature, which emphasises the need to align the type of indicators with the concept being measured (OECD et al., 2008), the dataset relies primarily on output indicators, reflecting observable outcomes. Input indicators are excluded, as they capture resources rather than realised characteristics. Some process indicators are retained where they directly contribute to shaping these observable characteristics. The dataset includes 22 indicators, whose main characteristics are reported in Table 1. All indicators are drawn from official open-access Belgian administrative and statistical sources. Additional details on data construction and selection are provided in Appendix A.

Table 1: Selected indicators included in the Residential Attractiveness Index

Dimension	ID	Indicator (unit)	Type	Time coverage
Accessibility	A1	Density of public transport stops (stops/km ²)	Output	2026
	A2	Road network density (km/km ²)	Output	2026
	A3	Distance to nearest general hospital site (km)	Output	2026
Demography	D1	Population growth (YoY, %)	Output	2015–2024
	D2	Share of young population (0–18, %)	Output	2015–2024
	D3	Share of elderly population (65+, %)	Output	2015–2024
Economy	E1	Median household income (€)	Output	2015–2023
	E2	Unemployment rate (%)	Output	FL: 2017–2023, WAL, BXL : 2015–2024
Governance	G1	Additional municipal income tax rate, IPP (%)	Process	2015–2023
	G2	Municipal debt (€/capita)	Output	2015–2023
	G3	Ordinary budget balance (€/capita)	Output	WAL, BXL: 2015–2023, FL: 2020–2024
Housing	H1	Newly authorised dwellings (per 1,000 inhabitants)	Process	2015–2024
	H2	Share of new buildings (built after 2011, %)	Output	2015–2024
	H3	Share of old buildings (built before 1900, %)	Output	2015–2024
Infrastructure	I1	Educational establishments (per 1,000 children aged 3–17)	Process	2026
	I2	Childcare facilities (per 1,000 children aged 0–2)	Process	2026
	I3	Presence of railway station (0/1)	Process	2026
Social conditions	SC1	Crime rate (offences per 1,000 inhabitants)	Output	2015–2024
	SC2	Risk of poverty rate (%)	Output	2015–2023
Urban form	U1	Share of green and natural cadastral land (%)	Output	2015–2024
	U2	Population density (inh/km ²)	Output	2015–2024
	U3	Land-use diversity, Shannon entropy (index)	Output	2015–2024

Note: FL = Flanders, WAL = Wallonia, and BXL = Brussels-Capital Region. D3 is excluded from the final dataset due to high collinearity with D2 ($r = -0.847$).

3.1.2 Data Preprocessing and Distributional Diagnostics

In 2025, several Belgian municipalities were merged, leading to changes in their identification codes (NSI codes) provided by Statbel (n.d.). Although the analysis covers the period before these mergers, some datasets are already reported using the updated NSI codes, while others still rely on the previous ones. To ensure consistency across data sources, all observations are harmonised to the 2025 NSI classification. When pre-merger NSI codes are used, values are aggregated to match the new municipalities and reassigned to the corresponding updated NSI codes.

A subset of indicators is only available for a single reference year (2026), as they derive from administrative or geospatial inventories without historical time series. These include, in

particular, infrastructure and accessibility variables. To integrate them into the panel, these indicators are treated as time-invariant over the study period and their 2026 values are replicated across all years, under the assumption that these structural characteristics evolve slowly over time and remain approximately stable over the period considered. This nevertheless introduces a limited look-ahead bias, since 2026 information is used to characterise earlier years in the panel (Blanchet et al., 2023).

Missing values are addressed using a structured imputation framework reflecting both the temporal and spatial structure of the data. Observations are classified into four categories: municipality–year gaps (isolated missing values within otherwise complete time series), left truncation (missing values at the beginning of the series), right truncation (missing values at the end), and persistent municipality-level missingness (repeated absence for specific municipalities). While part of this missingness can be attributed to administrative publication rules, the remaining cases are assumed to arise from data production and reporting processes and are therefore treated as Missing at Random (OECD et al., 2008). Imputation follows a sequential pipeline. Missing values are first filled using last observation carried forward (LOCF) for right-truncated series and backward observation carried forward (BOCF) for left-truncated series, thereby exploiting each municipality’s own observed trajectory whenever possible. Remaining gaps are then imputed using the average of geographically adjacent municipalities at time t , under the assumption that nearby municipalities provide locally comparable reference values. While the procedure is designed to rely primarily on information available at time t , some look-ahead bias may remain due to the use of BOCF, which relies on future observations for early-period gaps. This reflects a trade-off between bias and variance in missing data treatment, where allowing limited use of future information can improve the stability of the imputed series at the cost of a small bias (Blanchet et al., 2023).

The distribution of each indicator is examined to detect extreme values and asymmetry that may influence the composite index (OECD et al., 2008). Outliers are identified using the interquartile range (IQR) criterion. These observations are retained, as they reflect structural heterogeneity across municipalities rather than measurement error. To address asymmetry, skewness is computed for each indicator. Variables with absolute skewness greater than 1 are transformed using a set of candidate transformations adapted to their support, including square-root, logarithmic ($\log(x)$ or $\log(1 + x)$ when zeros are present), inverse ($1/x$), polynomial (x^2 , x^3), and arcsine square-root transformations for proportions (Hair et al., 2019). After each transformation, skewness is recomputed and transformations that do not reduce skewness to the interval $[-1, 1]$ are discarded. When multiple candidates satisfy this criterion, the least aggressive transformation is retained (Hair et al., 2019). Final choices are validated using histogram and Q–Q plot inspection. When no transformation achieves acceptable symmetry, winsorisation at the 1st and 99th percentiles has been considered (Boudt et al., 2020). Additional details regarding the preprocessing procedures are provided in Appendix B.

3.1.3 Multivariate Structure and Variable Selection

Multivariate analysis is used as a preliminary diagnostic step to examine the internal structure of the dataset before normalisation. It aims to assess the coherence of the selected indicators, detect potential redundancy, and verify whether the empirical structure of the data is consistent with the multidimensional theoretical framework underlying the index (OECD et al., 2008). The panel is first aggregated at the municipal level by averaging each indicator over 2015–2024, in order to capture persistent cross-sectional differences across municipalities.

Pearson correlations are then analysed to identify strong linear relationships and potential overlap between indicators, while recognising that correlation alone does not necessarily imply conceptual redundancy (OECD et al., 2008). The corresponding correlation matrix is reported in Appendix C.

Principal Component Analysis is used as a diagnostic tool to examine the internal structure of the dataset (OECD et al., 2008). The variables are standardised prior to the decomposition to ensure comparability. Then, the relevance of each component is assessed using its eigenvalue and its contribution to total explained variance. The number of meaningful components is primarily determined using the Kaiser criterion (eigenvalue > 1), complemented by inspection of the scree plot. Finally, component loadings are used to identify the variables contributing most strongly to each component and to assess potential redundancy among indicators (see Appendix D).

3.1.4 Index Construction

Because the indicators entering the RAI are expressed in heterogeneous units, a **normalization** step is required to transform them into comparable quantities before aggregation. Following standard practice (Mazziotta & Pareto, 2021; OECD et al., 2008), z-score standardisation is used as the baseline method, while alternative normalisations are considered as robustness checks. Indicator polarity is incorporated directly into the transformation so that higher values always reflect higher residential attractiveness. The baseline normalisation is defined as:

$$(1) y_{cj} = \frac{x_{cj} - \mu_j}{\sigma_j} \quad (2) y_{cj} = -\frac{x_{cj} - \mu_j}{\sigma_j}$$

where (1) applies to positively oriented indicators and (2) to negatively oriented indicators, x_{cj} denotes the value of indicator j for municipality c , and μ_j , σ_j are the mean and standard deviation of indicator j (Mazziotta & Pareto, 2021). To avoid look-ahead bias, normalisation parameters (μ_j , σ_j) are estimated exclusively on observations available up to and including year t , ensuring that no future information contaminates the standardisation of year t values (Sasse et al., 2023).

Once normalised, indicators are grouped into eight dimensions, and a **weighting** scheme is applied. Within each dimension, equal weighting is adopted. If dimension d contains k_d indicators, each indicator receives a weight $w_j = 1/k_d$, so that the dimension score is computed

as:

$$D_{c,d} = \frac{1}{k_d} \sum_{j \in d} y_{cj}$$

Finally, dimension scores are combined through an **aggregation** step using linear aggregation. Each dimension receives equal weight, so that the composite index for municipality c is:

$$RAI_c = \frac{1}{D} \sum_{d=1}^D D_{c,d}, \quad D = 8$$

While additive aggregation is easy to interpret, it implies full compensability across dimensions (OECD et al., 2008).

3.1.5 Robustness check

Robustness check is conducted to assess the sensitivity of the index to key methodological choices, namely normalisation, weighting, and aggregation (OECD et al., 2008). Four alternative specifications are considered, each modifying one component of the baseline.

First, normalisation robustness is assessed by replacing z-score standardisation with min–max rescaling, applied at each year t using only information available at t to preserve temporal consistency. The transformation is defined as:

$$(1) y_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (2) y_{ij} = 1 - \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

for positive (1) and negative (2) polarity, respectively (Mazziotta & Pareto, 2021; OECD et al., 2008).

Second, sensitivity to the within-dimension weighting structure is evaluated by replacing equal indicator weights with PCA-based weights (OECD et al., 2008). For each dimension, weights are derived from the absolute loadings of the first principal component and final RAI aggregates the eight dimensions using equal weights :

$$w_j^{PCA} = \frac{|a_{1j}|}{\sum_{j'} |a_{1j'}|}$$

Third, aggregation robustness is examined by replacing additive aggregation with geometric aggregation across dimensions:

$$RAI_c^G = \left(\prod_{d=1}^D D_{c,d} \right)^{1/D}$$

Since geometric aggregation requires strictly positive values, dimension scores are shifted by adding a constant whenever the minimum value is non-positive. The geometric mean reduces compensability relative to the arithmetic mean (Mazziotta & Pareto, 2013; OECD et al., 2008).

Fourth, an alternative specification based on the Mazziotta–Pareto Index (MPI) is implemented. Unlike the baseline index, which relies on a hierarchical aggregation structure through dimension scores, the MPI directly aggregates all normalised indicators into a single composite measure:

$$RAI_c^{MPI} = \bar{x}_c - \frac{s_c^2}{\bar{x}_c}$$

where $\bar{x}_c$ and s_c denote the mean and standard deviation of the normalised indicators for municipality c . By penalising dispersion across indicators, the MPI reduces the degree of compensability and favours municipalities displaying more balanced performances across dimensions (Mazziotta & Pareto, 2013). The rank stability of the four specifications is then assessed using Spearman’s rank correlation coefficient.

In addition, a leave-one-out sensitivity analysis is conducted to assess the contribution of each dimension to the index. The RAI is recomputed by excluding one dimension at a time, and the resulting municipal rankings are compared to the baseline. Following Saisana et al. (2005), the average shift in ranks is used to quantify the average displacement in municipal positions when a given dimension is excluded. Spearman’s rank correlation provides a complementary measure of overall ranking similarity. Both metrics are averaged across years to obtain dimension-specific indicators of structural influence.

3.2 Econometric Analysis

The econometric analysis examines the relationship between residential attractiveness and housing prices across Belgian municipalities over the period 2015–2024. It first documents the main patterns in the data, then estimates the relationship using econometric models, and finally assesses the robustness of the results.

3.2.1 Descriptive diagnostics

The dependent variable is the median transaction price of houses (2, 3, or 4+ facades) at the municipal level, excluding apartments, denoted $Price_{it}$, where i indexes municipalities and t indexes years (Statbel, n.d.-f). Missing values in median prices are first imputed using the median value of geographically adjacent municipalities at the same time t , ensuring a complete panel before transformation (see Appendix E). To determine the appropriate functional form, a Box–Cox transformation is applied (Box & Cox, 1964). The transformation parameter λ is estimated together with its 95% confidence interval based on the profile log-likelihood. Since $\lambda = 0$ lies within this interval, the logarithmic specification for median housing price has been retained.

To assess the feasibility of fixed-effects estimation, a between–within variance decomposition is performed for both RAI_{it} and $\ln(Price_{it})$, writing each variable as $x_{it} = \bar{x}_i + (x_{it} - \bar{x}_i)$. Total, between, and within standard deviations, together with the within-to-total ratio, are reported to verify that neither variable is time-constant within municipalities, as required for fixed-effects

identification (Wooldridge, 2010, p. 269).

Finally, the association between RAI_{it} and $\ln(\text{Price}_{it})$ is examined along both dimensions of the panel (Hsiao, 2003, pp. 4–5). A between-municipality Pearson correlation $r(\overline{RAI}, \overline{\ln(\text{Price})})$ is computed on municipality time-means, capturing whether more attractive municipalities persistently command higher prices. A within-municipality correlation $r(RAI_{it} - \overline{RAI}, \ln(\text{Price}_{it}) - \overline{\ln(\text{Price})})$ is computed on demeaned variables, capturing whether year-to-year changes in attractiveness within a municipality are associated with changes in prices. All detailed results for the descriptive diagnostics are provided in Appendix F.

3.2.2 Econometric Models

The empirical strategy relies on two complementary sources of variation. The between estimator captures cross-sectional differences across municipalities, while the within estimator exploits variation within municipalities over time. These two approaches rely on different identification assumptions and are therefore not directly comparable (Wooldridge, 2010).

The **between estimator** is obtained by applying OLS where each observation corresponds to a municipality averaged over the period 2015–2024. By construction, this approach removes all time variation and relies only on differences across municipalities. The baseline specification is:

$$\overline{\ln(\text{Price})}_i = \alpha + \beta \overline{RAI}_i + u_i$$

Consistency requires that $\overline{RAI}_i$ is uncorrelated with the error term u_i , implying the following exogeneity condition::

$$\mathbb{E}(u_i \mid \overline{RAI}_i) = 0$$

In practice, this means that no unobserved municipal characteristic should simultaneously affect residential attractiveness and housing prices (Wooldridge, 2010). This is a strong assumption and is unlikely to hold perfectly in practice. For example, municipalities with a long-standing history of economic prosperity may exhibit both higher RAI values and higher housing prices due to underlying factors such as wealth or urban structure that are not fully captured by the index. This may lead to omitted variable bias in β (Wooldridge, 2013).

Despite this limitation, the between estimator is retained because the research question focuses on differences in housing prices across municipalities, making cross-sectional variation the relevant source of identification. The estimated coefficient β should therefore be interpreted with caution, as it may partly reflect unobserved structural differences rather than the effect of attractiveness alone.

Municipalities that are geographically close are likely to share unobserved local characteristics, such as labour market conditions or land use regulations, which can induce spatial correlation in the residuals. Standard heteroskedasticity-robust errors do not account for this type of dependence and therefore tend to underestimate standard errors (Conley, 1999). To address

this issue, standard errors are corrected using the approach of Conley (1999), which allows for residual correlation across municipalities within a given distance threshold while assuming that correlation becomes negligible beyond this cutoff. The choice of this threshold is inherently somewhat arbitrary and a baseline cutoff of 25 km is adopted throughout the analysis.

All subsequent between-estimator specifications rely on the same identification assumption. The exogeneity condition therefore remains unchanged, and standard errors are treated using the same spatial correction.

To further control for geographic heterogeneity, specifications including province and region fixed effects are also estimated:

$$\overline{\ln(\text{Price})}_i = \alpha + \beta \overline{\text{RAI}}_i + \delta_g + u_i$$

The second empirical approach relies on the **within estimator** using a Two-Way Fixed Effects (TWFE) specification:

$$\ln(\text{Price}_{it}) = \alpha_i + \lambda_t + \beta \text{RAI}_{it} + \varepsilon_{it}$$

This model exploits within-municipality variation over time. Municipality fixed effects α_i control for all time-invariant characteristics, while time fixed effects λ_t capture common shocks affecting all municipalities. Consequently, the coefficient β is identified exclusively from changes occurring within the same municipality over time (Wooldridge, 2010).

Consistency of the TWFE estimator requires the following strict exogeneity condition:

$$\mathbb{E}(\varepsilon_{it} \mid \text{RAI}_{i1}, \dots, \text{RAI}_{iT}, \alpha_i, \lambda_t) = 0$$

In practice, this implies that changes in residential attractiveness are not themselves influenced by factors that simultaneously affect housing prices over time. This assumption is likely to be restrictive in the present context, as reverse causality constitutes a major concern. As noted by Alexandre et al. (2010), prices can simultaneously be both a determinant and an outcome of attractiveness. For example, rising housing prices may attract higher-income households or stimulate investment in local amenities and infrastructure, thereby improving the observable indicators entering the RAI. As a result, the RAI may become correlated with unobserved factors affecting housing prices, violating the strict exogeneity assumption required for consistent TWFE estimation (Wooldridge, 2010). In the absence of a valid instrumental variable strategy, the estimated coefficient should therefore be interpreted as a conditional association rather than as a causal effect.

An additional consideration arises from the construction of the RAI itself. The index is standardised using an expanding reference distribution over time. Consequently, its variation reflects not only changes in a municipality's own characteristics, but also mechanical variation induced by changes in the underlying standardisation distribution. Within-municipality variation there-

fore combines genuine changes in attractiveness with this mechanical component, which may introduce measurement noise and attenuate the estimated coefficient.

Given the panel structure of the data, standard errors are clustered at the municipality level to account for serial correlation within municipalities over time. Ignoring this dependence would lead to underestimated standard errors (Cameron & Miller, 2015). All subsequent within-estimator specifications rely on the same identification assumption and use the same treatment of standard errors.

To account for delayed effects, additional lagged specifications are estimated for $\ell = 1, 2, 3$:

$$\ln(\text{Price}_{it}) = \alpha_i + \lambda_t + \beta \text{RAI}_{i,t-\ell} + \varepsilon_{it}$$

3.2.3 Robustness Checks

Six robustness checks are conducted to assess the stability of the main results.

First, the lagged within-estimator specifications are complemented by a panel Granger causality test (Granger, 1969) to examine whether past values of the RAI contain predictive information for future housing prices beyond the persistence already captured by lagged prices. For each lag length $k \in \{1, \dots, 3\}$, a restricted model regressing $\ln(\text{Price}_{it})$ on its own k lags and two-way fixed effects is compared with an unrestricted model that additionally includes k lags of the RAI. The null hypothesis that all lagged RAI coefficients are jointly equal to zero is tested using a cluster-robust Wald F -statistic evaluated against an $F(k, G - 1)$ distribution, where G denotes the number of municipalities. The test is repeated for different lag lengths to ensure that the conclusions are not driven by a particular specification choice.

Second, the sensitivity of the results to the construction of the RAI is assessed by re-estimating the baseline between-estimator and within-estimator models using the alternative index specifications presented in Section 3.2.5. This robustness check evaluates whether the main findings remain stable across different normalisation, weighting, and aggregation procedures.

Third, residual spatial dependence is examined using Moran's I (Moran, 1950) computed on the residuals of the between-estimator and within-estimator models. The neighbourhood structure is defined using a Queen contiguity matrix with row-standardised weights, and statistical significance is assessed through 999 permutation tests (Hope, 1968). This robustness check serves both to evaluate whether the models leave substantial spatial autocorrelation unexplained and to assess whether the use of Conley standard errors is justified in the between-estimator specifications.

Fourth, the robustness of the results to spatial dependence is examined using Conley standard errors (Conley, 1999) in the between-estimator specifications. A baseline distance threshold of 25 km is adopted, while additional specifications using alternative thresholds of 10 km and 50 km are also estimated to assess the sensitivity of the results to the choice of cutoff distance.

Fifth, all within-estimator specifications are replicated using rank-based measures of the RAI. Since the baseline index is standardised using an expanding reference distribution, this transformation removes the mechanical variation induced by the normalisation procedure. The objective is therefore to assess whether the within-estimator results are driven by the construction of the index rather than by underlying economic relationships.

Finally, the analysis is replicated in levels by replacing $\ln(\text{Price}_{it})$ with housing prices expressed in euros. Both the between-estimator and within-estimator baseline specifications are re-estimated in order to verify that the main conclusions are not driven by the logarithmic transformation of the dependent variable.

3.3 Predictive Modeling

The application of machine learning to housing price prediction has expanded substantially over the past two decades. Housing prices are influenced by interactions between dwelling characteristics and territorial attributes (Hilber, 2011; Rosen, 1974), which may involve non-linear relationships. In this context, machine learning methods provide a flexible framework capable of capturing such patterns without imposing strong parametric assumptions.

Early studies showed that flexible machine learning models could outperform traditional linear regression in out-of-sample prediction, as they better capture non-linear relationships between housing characteristics and transaction prices (Limsombunchai et al., 2004). Later research confirmed these results across different housing markets and machine learning algorithms (Ho et al., 2021; Park & Bae, 2015). In particular, ensemble methods such as Random Forest and gradient boosting have often been identified among the best-performing approaches (Ho et al., 2021). Similarly, Pai and Wang (2020) show that machine learning models trained on transaction data can achieve higher predictive accuracy than traditional valuation methods.

Despite these advantages, machine learning approaches also present important limitations. Compared with traditional regression models, they generally offer lower interpretability, making it more difficult to identify the individual contribution of predictors to housing price variation (Ho et al., 2021; Limsombunchai et al., 2004). Their predictive performance also depends strongly on the quality, representativeness, and size of the training data (Ho et al., 2021).

The present analysis extends this literature by framing housing price prediction as a temporal forecasting problem at the municipal scale and by evaluating whether residential attractiveness, measured through the RAI, provides additional predictive information for future housing price dynamics across Belgian municipalities. The following subsections present the data splitting strategy, prediction tasks, model specifications, algorithms, and evaluation metrics.

3.3.1 Data Splitting Strategy

The full sample (2015–2024) is divided into two non-overlapping subsets. The development sample covers the years used for model training and validation; the hold-out sample corresponds to the final unseen evaluation period (target year 2024) and is used only once, after all modelling

decisions have been finalised.

The analysis considers three forecasting horizons h , where the objective is to predict median housing prices at time $t + h$ using information available at time t . For each horizon, features are drawn from year t and the target corresponds to $\ln(\text{Price}_{i,t+h})$. For $h = 0$, the development sample covers feature years 2015–2023 and the hold-out uses feature year 2024 targeting 2024 prices. For $h = 1$, the development sample covers feature years 2015–2022 and the hold-out uses feature year 2023 targeting 2024 prices. For $h = 3$, the development sample covers feature years 2015–2020 and the hold-out uses feature year 2021 targeting 2024 prices. The complete fold structure is reported in Table 2.

Table 2: Expanding-window validation structure (development sample)

Horizon	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7
$h = 0$	2015–2016 → 2017	2015–2017 → 2018	2015–2018 → 2019	2015–2019 → 2020	2015–2020 → 2021	2015–2021 → 2022	2015–2022 → 2023
$h = 1$	2015–2016 → 2018	2015–2017 → 2019	2015–2018 → 2020	2015–2019 → 2021	2015–2020 → 2022	2015–2021 → 2023	–
$h = 3$	2015–2016 → 2020	2015–2017 → 2021	2015–2018 → 2022	2015–2019 → 2023	–	–	–

3.3.2 Prediction Tasks

The predictive analysis is formulated as a supervised regression problem to forecast median housing prices at the municipality level. The dependent variable is the natural logarithm of the median housing price, $\ln(\text{Price}_{i,t})$. This transformation ensures that prediction errors are evaluated on a scale where they approximate proportional deviations rather than absolute differences, which is appropriate given the substantial heterogeneity in price levels across Belgian municipalities.

Predictions are carried out for horizons $h \in \{0, 1, 3\}$. These horizons are chosen to evaluate contemporaneous, one-year-ahead, and longer-term predictions while remaining compatible with the limited temporal dimension of the panel. When $h = 0$, the model produces a contemporaneous nowcast, while $h \geq 1$ corresponds to genuine multi-step-ahead forecasts. For each horizon, explanatory variables are observed at time t , and the target variable is $\ln(\text{Price}_{i,t+h})$. Housing prices are known to display temporal persistence at the aggregate level (Case & Shiller, 1988), which motivates the inclusion of lagged prices as predictors in each model specification. For each horizon, a single pooled model is estimated across municipalities. The prediction problem is therefore written as

$$\hat{y}_{i,t+h} = f(X_{i,t}; \theta),$$

where $\hat{y}_{i,t+h}$ denotes the predicted value, $X_{i,t}$ the set of predictors observed at time t , and $f(\cdot)$ denotes a model-specific prediction function estimated separately for each horizon and specification.

3.3.3 Model Specifications

Four model specifications are estimated for each forecasting horizon. As reported in Table 3, the model M0 serves as a baseline and relies only on past housing prices. M1 adds the RAI, M2 further includes macroeconomic controls, mortgage rate (European Central Bank, 2024)

and economic sentiment indicator constructed from daily Belgian news articles through textual sentiment analysis (Sentometrics, 2024), while M3 replaces the RAI with its 21 underlying indicators in order to evaluate whether the disaggregated dimensions contain additional predictive information beyond the aggregated index. The specification of housing prices differs across forecasting horizons. For $h = 0$, the target variable is $\ln(\text{Price}_{i,t})$, and models rely on the lagged price $\ln(\text{Price}_{i,t-1})$ to prevent data leakage, since the contemporaneous price corresponds to the variable being predicted. This contemporaneous specification is not intended as a real-time forecasting exercise, but rather to assess whether the RAI and contemporaneous market conditions contain additional predictive information beyond lagged prices alone. By contrast, for $h \geq 1$, the target becomes $\ln(\text{Price}_{i,t+h})$, while the contemporaneous price $\ln(\text{Price}_{i,t})$ is included as a persistence control.

Table 3: Model specifications

Horizon	Model	Features
$h = 0$ (target: $\ln(\text{Price}_{i,t})$)		
	M0	$\ln(\text{Price}_{i,t-1})$
	M1	$\ln(\text{Price}_{i,t-1})$, $\text{RAI}_{i,t}$
	M2	$\ln(\text{Price}_{i,t-1})$, $\text{RAI}_{i,t}$, mortgage_rate_t , sentiment_t
	M3	$\ln(\text{Price}_{i,t-1})$, $X_{1,i,t}, \dots, X_{K,i,t}$
$h \in \{1, 3\}$ (target: $\ln(\text{Price}_{i,t+h})$)		
	M0	$\ln(\text{Price}_{i,t})$
	M1	$\ln(\text{Price}_{i,t})$, $\text{RAI}_{i,t}$
	M2	$\ln(\text{Price}_{i,t})$, $\text{RAI}_{i,t}$, mortgage_rate_t , sentiment_t
	M3	$\ln(\text{Price}_{i,t})$, $X_{1,i,t}, \dots, X_{K,i,t}$

Note: i indexes municipalities, t denotes time, and $k = 1, \dots, K = 21$ indexes the normalised indicators in the RAI.

Before estimation, the same imputed dataset as in the econometric analysis is used. For the linear model considered here, ridge regression, predictors are standardised prior to estimation, as ridge solutions are not equivariant under scaling of the inputs (Hastie et al., 2009). Tree-based models are not standardised, since their performance does not depend on predictor scales.

3.3.4 Algorithms

Three models are considered to compare predictive performance. Ridge regression is used as a regularised linear benchmark, as the ℓ_2 penalty shrinks coefficient estimates and improves numerical stability in predictive settings with multiple correlated predictors (Hastie et al., 2009; Hoerl & Kennard, 1970). Random Forest is implemented as a bagging-based tree ensemble that captures non-linearities and interaction effects while reducing predictive variance through the aggregation of decorrelated trees (Breiman, 2001). XGBoost is used as a boosting-based approach in which trees are sequentially fitted to remaining prediction errors (Chen & Guestrin, 2016). As an additional robustness check, an unregularised OLS specification is also estimated to assess whether the ℓ_2 regularisation implemented in Ridge regression provides measurable predictive gains.

Model performance is evaluated using a time-series backtesting procedure based on a rolling forecasting origin with an expanding window (Hyndman & Athanasopoulos, 2021). For each

validation year T , models are trained on all observations from years strictly prior to T and evaluated on observations from year T . This preserves the temporal ordering of the data and ensures that only past information is used to predict future outcomes. The number of backtesting iterations depends on the prediction horizon, as the target year must remain within the development sample. As a result, seven validation years are available for $h = 0$, six for $h = 1$, and four for $h = 3$.

Within each backtesting iteration, hyperparameters are selected using only the years included in the corresponding training sample. For example, if a model is evaluated on year $T = 2019$, hyperparameter tuning is carried out exclusively using observations from years prior to 2019. The backtesting validation year is therefore never used during hyperparameter selection. The tuning step follows an expanding-window structure within the training sample. Models are repeatedly estimated on earlier years and validated on subsequent years, with up to three validation splits depending on the number of years available in the training sample. For each candidate configuration, performance is evaluated using negative mean squared error.

Hyperparameter configurations are sampled using `RandomizedSearchCV` with 60 random draws. This choice balances search coverage and computational feasibility, consistent with evidence that randomized search can efficiently explore hyperparameter spaces with a limited number of evaluations (Bergstra & Bengio, 2012). For Ridge regression, feature scaling is re-estimated separately within each training split, ensuring that scaling parameters are computed using training data only. Random seeds are fixed at the horizon, specification, and algorithm level to ensure reproducibility. The search spaces are reported in Appendix H.

Model selection is based on average backtesting performance across validation years. Predictive performance is assessed using RMSE, MAE, and R^2 , computed on the logarithmic scale. These metrics are first used within the inner cross-validation folds to select hyperparameters, then averaged across backtesting folds to identify the best-performing specification. Once the best specification and hyperparameters are identified, the corresponding model is retrained on the full development sample using the hyperparameters selected during the last backtest fold and evaluated on the hold-out sample.

This section presented the empirical framework underlying the analysis. It introduced the indicators used to capture residential attractiveness, detailed the data processing and the construction of the composite index. It also outlined the econometric framework used to examine the relationship between residential attractiveness and housing prices, together with the robustness checks implemented. Finally, the predictive framework was presented, including the forecasting tasks, model specifications, and evaluation procedures. The next section reports the empirical findings.

4 Results

This chapter presents the empirical results of the analysis. It first examines the construction and behaviour of the Residential Attractiveness Index (RAI). It then evaluates the relationship between the RAI and housing prices, distinguishing between cross-sectional differences across municipalities and within-municipality dynamics over time. Finally, the chapter assesses the predictive relevance of residential attractiveness by comparing forecasting models across different horizons and examining the resulting municipal price predictions.

4.1 Residential Attractiveness Index

The following results present the main properties of the Residential Attractiveness Index, including its internal structure, distribution, temporal stability, and robustness to alternative construction methods and dimension exclusions.

4.1.1 Multivariate Analysis

Prior to constructing the RAI, a **Pearson correlation analysis** was conducted across the 22 candidate indicators to identify potential redundancies (see Appendix C). Several pairs of indicators exceeded the threshold of $|r| > 0.80$. However, only one pair was considered conceptually redundant, namely D2 (share of young population) and D3 (share of elderly population), with $r = -0.847$. As both indicators describe the same demographic balance from opposite ends of the age distribution, retaining both would have disproportionately increased the weight of the demographic dimension without providing substantial additional information. D3 was therefore excluded. Other strongly correlated indicators were nevertheless retained when they captured different aspects of residential attractiveness. For instance, A1 (public transport stops per km²) and U2 (population density), despite a correlation of $r = 0.853$, reflect different underlying dimensions of municipalities. Their strong correlation is therefore interpreted as a structural feature of urban environments rather than as evidence of redundancy. The final dataset includes 21 indicators grouped into eight dimensions.

A **Principal Component Analysis** is conducted on the 21 standardised indicators to examine the internal structure of the dataset and assess the extent to which indicators capture distinct or overlapping information (OECD et al., 2008). The objective is not to derive weights, as the index relies on an equal-weighting scheme, but rather to provide diagnostic insight into the correlation structure of the data.

Based on the Kaiser criterion, five components are retained, explaining 69.3% of the total variance. A large share of this variation is concentrated in the first two components, which together account for 47.6%, indicating that a limited number of underlying patterns structure the data (see Appendix D).

As reported in Table 4, the first component represents the main source of variation across municipalities. It opposes municipalities characterised by high unemployment, elevated poverty,

high crime rates, and dense transport infrastructure to those with higher incomes, a more recently built housing stock, greater land-use diversity, and better childcare provision. This pattern reflects a composite axis along which urbanisation, accessibility, and socio-economic conditions co-vary. The second component reflects a different territorial configuration. Municipalities scoring high on this component tend to combine a larger share of old buildings, more green and natural land, greater land-use diversity, stronger educational infrastructure, and greater distance to the nearest hospital. This component therefore appears to capture differences in land-use organisation. The remaining components explain a smaller share of the variance and are less clearly structured. Demographic and governance indicators are spread across these dimensions, suggesting that they do not align with a single dominant pattern in the data.

These results have implications for the interpretation of the composite index. Although the index uses an equal-weighting scheme, the actual contribution of each indicator to differences in RAI scores depends on the underlying correlation structure. Since the first component explains a large portion of the overall variance, indicators related to socio-economic conditions, accessibility, and urban density contribute more to the dispersion of the index than indicators associated with governance or demographic characteristics, which are less aligned with the dominant structure of the data.

Table 4: PCA component matrix

Code	PC1	PC2	PC3	PC4	PC5
SC2	0.896	0.273	-0.045	-0.104	0.036
A1	0.807	-0.341	0.260	-0.196	0.096
E2	0.789	0.515	-0.006	-0.002	-0.072
SC1	0.776	-0.018	-0.422	0.096	0.104
U2	0.754	-0.342	0.261	-0.247	0.153
A2	0.752	-0.404	0.213	-0.044	0.086
H2	-0.746	-0.109	0.065	0.113	0.255
E1	-0.721	-0.329	0.209	0.115	0.065
U3	-0.580	0.616	-0.002	0.134	-0.158
I2	-0.428	0.163	-0.184	-0.216	-0.076
A3	-0.320	0.575	0.053	-0.201	0.490
H3	0.281	0.733	0.162	-0.012	-0.086
I1	0.271	0.715	-0.217	-0.052	0.150
U1	-0.132	0.679	-0.006	0.013	0.483
G1	0.025	0.504	0.221	0.348	-0.474
D2	0.288	0.244	0.687	0.334	0.169
H1	0.157	-0.444	-0.535	0.370	0.128
G2	0.208	0.144	-0.493	0.296	0.374
I3	0.353	-0.049	-0.279	0.499	-0.260
D1	-0.092	-0.423	0.314	0.584	0.360
G3	-0.346	-0.338	-0.186	-0.456	0.018

Note: Loadings $\geq |0.40|$ in bold. Indicators sorted by dominant component.

Concerning the **distribution**, the RAI displays substantial variation across municipalities, with a standard deviation of 0.291 and values ranging from -1.086 to 1.079 over the 2015–2024 period. As shown in Figure 1, the distribution is slightly asymmetric, with a wider dispersion among lower values, indicating that municipalities with lower attractiveness tend to be more heterogeneous.

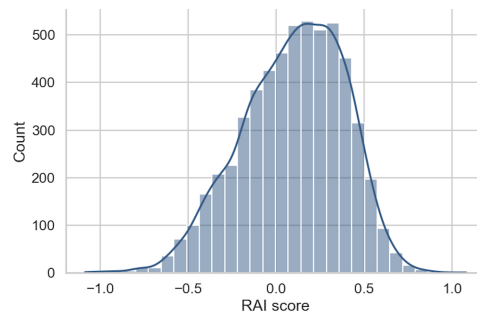


Figure 1: Distribution of the Residential Attractiveness Index

The spatial distribution (Figure 2) reveals a clear but non-uniform regional structure. Higher RAI values are concentrated in a broad northern and central belt, particularly across much of Flanders and the municipalities surrounding the Brussels area. Lower values are more frequent in the south-central part of the country, especially in several municipalities of Hainaut, Namur, and parts of Liège. However, the map also shows pockets of relatively high RAI values in the south-east, in Luxembourg, indicating substantial within-region heterogeneity. More generally, neighbouring municipalities often display similar scores, while local deviations remain visible, suggesting that residential attractiveness is spatially structured but not spatially uniform.

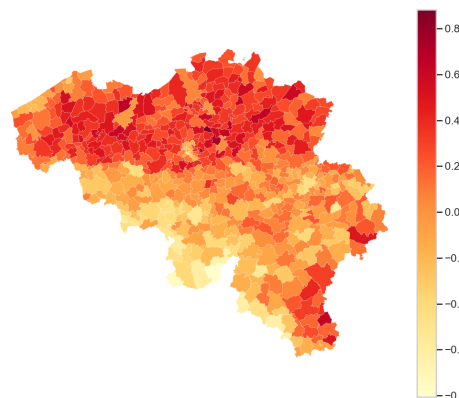


Figure 2: Spatial distribution of the RAI across Belgian municipalities (2024)

4.1.2 Temporal Stability of the RAI

Before analysing temporal stability, it is important to clarify the interpretation of RAI scores over time. Since the index is standardised separately each year, scores measure a municipality's relative position rather than an absolute level of residential attractiveness. As a result, score levels are not directly comparable across years, whereas rankings remain informative for assessing whether municipalities maintain their relative position over time.

Against this background, the results point to a high degree of stability in the ranking structure. Consecutive year-on-year Spearman correlations range between 0.933 and 0.960 (Figure 3), which indicates that the ordering of municipalities changes only marginally from one year to the next. In other words, most municipalities tend to remain close to their previous position in the ranking. Over a longer horizon, the correlation between 2015 and 2024 decreases to $\rho = 0.87$, suggesting that small year-to-year adjustments accumulate over time and lead to more noticeable

reordering.

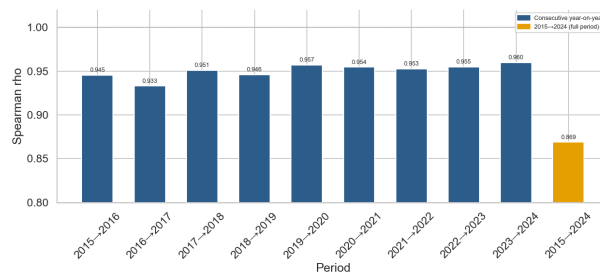


Figure 3: Consecutive year-on-year Spearman rank correlation

To illustrate the diversity of individual dynamics underlying this aggregate stability, Figure 4 presents the rank trajectories of five randomly selected municipalities over the period. Tielt-Winge (red), Mouscron (yellow), and Amblève (dark blue) follow relatively gradual trends over time, with the first two respectively improving and declining in the national ranking. By contrast, Avelgem (light blue) and Daverdisse (green) display more pronounced fluctuations, highlighting that a stable overall ranking structure can still coexist with meaningful year-to-year changes for individual municipalities.

The case of Daverdisse illustrates how the index can react to isolated shocks. The sharp improvement in 2021 is mainly driven by population growth (D1), which increases from -1.5% in 2020 to $+2.7\%$ in 2021, a rise of about 4.2 percentage points. While this change is moderate in raw terms, it is amplified after standardisation and disproportionately affects the index. As population growth returns to more typical levels in 2022 (-0.8%) and other indicators remain broadly stable, the 2021 peak reflects a temporary fluctuation rather than a structural change in residential attractiveness. The implications of this sensitivity are discussed in Section 5.2.

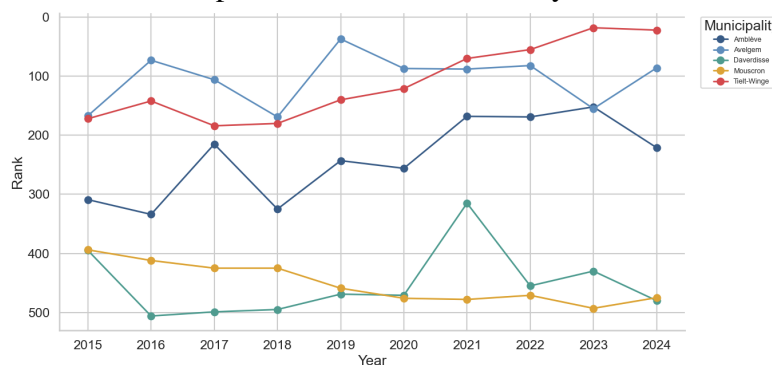


Figure 4: Rank trajectories of five selected municipalities

4.1.3 Robustness of the RAI

The baseline RAI is compared with several alternative specifications that vary the normalisation, weighting, and aggregation choices. Rank correlations are computed using period-average scores across the 565 municipalities. Three main results emerge as reported in Table 5. First, using PCA-based weights or replacing the arithmetic mean with geometric aggregation leads to almost identical rankings. This indicates that the baseline assumptions of equal weighting and additive aggregation do not drive the results. Second, replacing z-score standardisation

with min–max normalisation introduces slightly more variation, consistent with the greater sensitivity of min–max normalisation to extreme values. However, the overall ranking remains strongly preserved, suggesting that the results are robust to the choice of normalisation. Finally, the specification based on the Mazziotta–Pareto Index produces a markedly different ranking, highlighting the sensitivity of the results to the aggregation structure. While the baseline index relies on a fully compensatory approach, the Mazziotta–Pareto Index allows only limited compensation and penalises unbalanced indicator profiles. The ranking therefore remains relatively stable across compensatory methods, but changes substantially when indicator imbalance is explicitly penalised.

Table 5: Rank correlations between the baseline RAI and alternative specifications

Comparison	ρ
Baseline vs. Min-max	0.929
Baseline vs. PCA-weighted	0.986
Baseline vs. Geometric mean	0.988
Baseline vs. MPI	−0.382

The comparison of top-ranked municipalities confirms this pattern. Rankings remain largely stable across compensatory approaches, with only a limited number of municipalities showing sensitivity. In contrast, the MPI leads to a systematic reordering of municipalities, reflecting its fundamentally different aggregation logic.

Table 6: Top 10 municipalities under alternative specifications

Municipality	Base	Min-max	PCA	Geom.	MPI
Boortmeerbeek	1	1	1	1	195
Herent	2	2	2	2	243
Attert	3	64	5	4	421
Oud-Heverlee	4	4	3	5	228
Pelt	5	3	12	3	197
Wezembeek-Oppeem	6	94	4	16	322
Zwalm	7	10	11	15	234
Kortenbergh	8	15	10	7	224
Olen	9	11	14	6	200
Rotselaar	10	7	16	10	252

Note: Rankings are based on period-average scores over 2015–2024. Lower rank indicates higher residential attractiveness.

To assess the structural robustness of the index, a leave-one-out sensitivity analysis is conducted. Each dimension is sequentially excluded and the RAI is recomputed on the remaining seven dimensions. This allows the influence of each dimension on the overall municipal ranking to be evaluated. Two complementary metrics are reported: the mean Spearman rank correlation (ρ) with the baseline index and the average absolute rank shift, which measures the average displacement in municipal rankings.

Table 7: Leave-one-out sensitivity analysis of the RAI

Dimension	ρ	average absolute rank shift
Social Conditions	0.934	46
Accessibility	0.948	40
Governance	0.956	35
Economy	0.959	34
Housing	0.967	32

Dimension	ρ	average absolute rank shift
Infrastructure	0.975	28
Urbanism	0.981	24
Demographics	0.985	21

As shown in Table 7, the exclusion of each dimension preserves a high correlation with the baseline ranking, indicating that the index is not driven by a single dimension. However, the magnitude of the rank changes differs across dimensions. Social Conditions and Accessibility generate the largest perturbations, while Demographics and Urbanism have a more limited effect. These differences are consistent with the PCA results (Section 4.1.1), where accessibility, urban density, and socio-economic indicators contribute strongly to the dominant component, whereas other dimensions are more weakly structured or dispersed across components. This suggests that the observed sensitivity reflects the underlying correlation structure of the data rather than an explicit weighting mechanism. The results indicate that the RAI is structurally robust, but not dimension-neutral. While some dimensions exert a stronger influence on the ranking, the absence of a clear discontinuity in the sensitivity pattern does not support the exclusion of any dimension on empirical grounds alone.

4.2 Econometric Results

The following results examine the relationship between residential attractiveness and housing prices, before assessing the robustness of the findings.

4.2.1 Main Results

Table 8 highlights a marked contrast between the between and within estimates.

The **between estimator** produces large and highly statistically significant coefficients across all specifications. In the baseline model, a one-unit increase in the RAI is associated with a 0.733 log-point difference in median housing prices, implying that two municipalities separated by one standard deviation in residential attractiveness are associated with a price difference of approximately 24% ($e^{0.733 \times 0.291} - 1$). This coefficient remains almost unchanged when region fixed effects are introduced (0.739), suggesting that the relationship is not solely driven by broad regional differences. When province fixed effects are included, the coefficient decreases to 0.597, indicating that part of the association reflects spatially correlated factors at a more aggregated geographic level. A large and highly significant coefficient nonetheless persists within provinces, indicating that municipalities with higher residential attractiveness are associated with higher housing price levels even when comparing geographically similar areas.

The **within-estimator coefficients** remain small and statistically non-significant at the 10% level across both contemporaneous and lagged specifications. Although the coefficient at lag 3 reaches weak significance ($p = 0.093$), this effect is not supported by neighbouring lags and therefore does not provide evidence of a robust delayed relationship. Granger causality tests reported in Section 4.2.2 further confirm that the RAI does not contain predictive information for housing

price changes at any horizon. The Pearson correlation analysis supports this interpretation (see Appendix F). Although the within-demeaned correlation between the RAI and log housing prices is relatively high ($r = 0.58$), this association disappears once time fixed effects are introduced in the TWFE specification. This suggests that the observed co-movement primarily reflects common temporal trends rather than municipality-specific dynamics.

Table 8: Econometric model results

Specification	$\hat{\beta}$	SE	p-val.	Sig.
<i>Panel A, Between</i>				
RAI only	0.733	0.095	<0.001	***
RAI + region FE	0.739	0.107	<0.001	***
RAI + province FE	0.597	0.069	<0.001	***
<i>Panel B, Within (TWFE)</i>				
Baseline (lag = 0)	0.016	0.023	0.494	
Lag = 1	0.010	0.022	0.638	
Lag = 2	0.013	0.023	0.560	
Lag = 3	0.047	0.028	0.093	.

Note: Dependent variable: $\ln(\text{median housing price})$. *** $p < 0.001$, . $p < 0.10$.

4.2.2 Robustness Checks

A series of additional robustness checks is conducted to assess the sensitivity of the results to alternative specifications, spatial dependence corrections, index constructions, and transformations of the dependent variable. Detailed results are reported in Appendix G.

First, the Granger causality tests show that past values of the RAI do not provide additional predictive information for housing prices once past housing prices are taken into account. The null hypothesis of no Granger causality cannot be rejected for any of the lag lengths considered. This also clarifies the marginally significant coefficient observed at $\ell = 3$ in Table 8. Since the effect is neither supported by neighbouring lags nor jointly significant in the Granger causality test, it does not indicate a systematic dynamic relationship between the RAI and housing prices over time.

Second, the econometric results remain broadly stable across alternative RAI constructions. In the between estimator specifications, all compensatory aggregation methods continue to produce large and statistically significant coefficients. By contrast, the specification based on the Mazziotta–Pareto Index does not produce a significant coefficient, indicating that the strength of the cross-sectional relationship is sensitive to the degree of compensation allowed between indicators in the aggregation procedure. In the within estimator specifications, the estimated coefficients remain statistically insignificant across all alternative index constructions. Although the Mazziotta–Pareto specification produces a statistically significant coefficient, its magnitude remains effectively zero and therefore does not indicate a meaningful economic relationship.

Third, Moran's I tests reveal substantial positive spatial autocorrelation in the residuals of the between specifications, indicating that neighbouring municipalities tend to exhibit correlated residuals. This result supports the use of Conley spatial standard errors in the cross-sectional

analysis. By contrast, no significant spatial autocorrelation is detected in the residuals of the within specifications once municipality and time fixed effects are included, suggesting that these effects absorb most of the persistent spatial structure in the data.

Fourth, the between-estimator results remain qualitatively unchanged when using alternative Conley distance cutoffs of 10 km and 50 km, indicating that the main conclusions are not sensitive to the choice of spatial correction bandwidth.

Additionally, the results remain unchanged when the continuous RAI is replaced by within-year rankings, indicating that the absence of a within effect is not driven by the method used for standardization of the RAI.

Finally, the analysis is replicated using housing prices expressed in levels rather than logarithms. The between-municipality relationship remains large and statistically significant, while the within-municipality estimates remain close to zero and not statistically significant. This confirms that the findings are not driven by the logarithmic transformation of the dependent variable.

4.3 Predictive Results

The following results examines predictive performance across forecasting horizons through backtesting and hold-out evaluation, before presenting the municipal price forecasts generated by the selected models.

4.3.1 Backtesting

The backtesting results reported in Table 9 highlight four main patterns. The first two patterns focus on the predictive performance of RAI-based specifications (M1 and M2), while the last two isolate the specific predictive contribution of the RAI composite itself relative to alternative feature sets and price persistence.

The blue-highlighted entries, which identify the best-performing RAI-based specification (M1 or M2) at each forecasting horizon, reveal two consistent findings. First, Ridge consistently achieves the lowest prediction errors across all horizons and specifications. At $h = 0$, the RMSE reaches 0.0984 under M1, compared to 0.1068 for RF and 0.1092 for XGB. The same ordering is observed at $h = 1$ and $h = 3$, suggesting that the predictive relationship is captured sufficiently well by a regularised linear specification and that more flexible models do not generate additional gains.

Second, the preferred RAI-based specification varies across forecasting horizons. At $h = 0$, M1 Ridge slightly outperforms M2 Ridge, suggesting that mortgage rates and the economic sentiment indicator do not improve contemporaneous predictions and slightly increase forecasting error. From $h = 1$ onward, however, M2 Ridge becomes the best-performing RAI-based specification. This pattern suggests that macroeconomic variables become progressively more informative at longer forecasting horizons.

The best overall specification at each forecasting horizon, indicated in green in Table 9, is

M3 Ridge across all three horizons. Since M3 replaces the aggregated RAI composite with its 21 underlying indicators, this systematic outperformance suggests that the aggregation into the RAI leads to a loss of predictive information. By compressing multiple indicators into a single measure, the index reduces the model's ability to exploit variation across dimensions. When indicators are used separately, this additional information improves predictive performance. The performance gap between M1 Ridge and M3 Ridge widens with the forecasting horizon, from approximately 1.7% at $h = 0$ to 2.1% at $h = 1$ and 7.0% at $h = 3$, suggesting that richer feature sets become increasingly relevant at longer forecasting horizons.

Finally, comparing M0 and M1 Ridge directly isolates the incremental contribution of the RAI composite beyond price persistence alone. RMSE decreases from 0.0990 to 0.0984 at $h = 0$, from 0.0982 to 0.0977 at $h = 1$, and from 0.1131 to 0.1125 at $h = 3$. Although the reduction in RMSE is observed at all horizons, it remains below 0.6% in every case, suggesting that past prices already capture most of the predictive signal, while the additional predictive contribution of the aggregated RAI composite remains limited.

Table 9: Backtest performance across horizons (development sample, target years 2017–2023)

Spec	Algorithm	$h = 0$		$h = 1$		$h = 3$	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
M0	RF	0.1032	0.9023	0.1007	0.9082	0.1149	0.8812
M0	Ridge	0.0990	0.9106	0.0982	0.9126	0.1131	0.8847
M0	XGB	0.1058	0.8971	0.1028	0.9044	0.1172	0.8762
M1	RF	0.1068	0.8958	0.1046	0.9009	0.1179	0.8747
M1	Ridge	0.0984	0.9117	0.0977	0.9136	0.1125	0.8860
M1	XGB	0.1092	0.8909	0.1063	0.8976	0.1171	0.8765
M2	RF	0.1036	0.9019	0.1008	0.9082	0.1105	0.8901
M2	Ridge	0.0990	0.9103	0.0961	0.9164	0.1066	0.8979
M2	XGB	0.1060	0.8967	0.1027	0.9044	0.1107	0.8898
M3	RF	0.0980	0.9123	0.0986	0.9120	0.1082	0.8948
M3	Ridge	0.0967	0.9146	0.0956	0.9172	0.1048	0.9013
M3	XGB	0.1032	0.9026	0.1034	0.9027	0.1093	0.8923

Note: **Green** denotes the best overall result within each forecasting horizon, while **Blue** denotes the best result among specifications including the RAI composite score (M1–M2).

The behaviour of the regularisation parameter α provides additional insight into the estimation problem. Although α is searched over a wide range (10^{-5} to 10^3), the selected values rapidly converge toward the lower bound from the third fold onward (Table 10). This pattern suggests that Ridge increasingly behaves like an unpenalised linear model as the training sample expands. Larger α values are selected only in the earliest folds, where the limited number of observations likely increases estimation instability.

Table 10: Selected α values for Ridge across backtest folds (development sample, target years 2017–2023)

Horizon	Model	$T = 2017$	$T = 2018$	$T = 2019$	$T = 2020$	$T = 2021$	$T = 2022$	$T = 2023$
$h = 0$	M1 Ridge	1.00	3.45	1.03×10^{-5}	2.12×10^{-5}	1.11×10^{-5}	1.13×10^{-5}	1.03×10^{-5}
$h = 1$	M2 Ridge	4.20	1.51×10^{-5}	1.02×10^{-5}	1.45×10^{-5}	1.02×10^{-5}	1.00×10^{-5}	—
$h = 3$	M2 Ridge	1.66	1.06×10^{-5}	1.00×10^{-5}	3.19×10^{-5}	—	—	—

To assess whether the ℓ_2 penalty remains useful once the training sample becomes larger, OLS is estimated as a robustness check on the same folds and specifications. Table 11 shows that Ridge and OLS produce virtually identical RMSE and R^2 values across all forecasting horizons. The results therefore suggest that the predictive gains from regularisation become negligible once sufficient training data are available.

Table 11: Ridge versus OLS across backtest folds (development sample, target years 2017–2023)

Horizon	Spec	Ridge RMSE	OLS RMSE	Ridge R^2	OLS R^2
$h = 0$	M1	0.0984	0.0984	0.9117	0.9117
$h = 1$	M2	0.0961	0.0961	0.9164	0.9164
$h = 3$	M2	0.1066	0.1066	0.8979	0.8980

4.3.2 Selected models and out-of-sample validation

The hold-out evaluation reported in Table 12 provides an out-of-sample assessment of predictive performance across forecasting horizons.

At $h = 0$, validation and hold-out RMSE remain very close, suggesting that the model generalises relatively well at the contemporaneous horizon.

At $h = 1$, the gap between validation and hold-out performance becomes more pronounced, with hold-out RMSE exceeding validation RMSE by 16.5%. This corresponds approximately to prediction errors of 11.9% in housing prices ($e^{0.1120} - 1 \approx 11.9\%$). The distribution of prediction errors also shifts, as the model underestimates prices in a large majority of municipalities (441 out of 565). This suggests that the model struggles to capture the housing price level observed in 2024 when forecasting one year ahead using 2023 characteristics.

At $h = 3$, predictive performance deteriorates further, with hold-out RMSE exceeding validation RMSE by 33.5%. This corresponds approximately to prediction errors of 15.3% in housing prices ($e^{0.1423} - 1 \approx 15.3\%$). In contrast to the $h = 1$ results, the model predominantly overestimates prices, with 506 municipalities overestimated against only 59 underestimated. This pattern likely reflects the longer forecasting gap at $h = 3$, where 2021 characteristics are used to predict 2024 prices, limiting the model's ability to account for subsequent changes in housing market conditions.

Overall, the widening gap between validation and hold-out RMSE across horizons indicates that predictive performance deteriorates as the forecasting horizon increases. These results should nevertheless be interpreted with caution, since the hold-out evaluation relies on a single target year (2024) and may therefore partly reflect conditions specific to that period rather than a broader structural pattern.

Table 12: Hold-out evaluation of selected models (hold-out sample, target year 2024)

Horizon	Spec	Model	RMSE validation	RMSE Hold-out	$\Delta\%$
0	M1	Ridge	0.0984	0.0975	-0.9
1	M2	Ridge	0.0961	0.1120	+16.5
3	M2	Ridge	0.1066	0.1423	+33.5

4.3.3 Predictions

Three main patterns emerge from the hold-out evaluation. First, the $h = 0$ specification produces the most accurate predictions in all three municipalities, with errors of +10.0%, -3.6%, and +0.1% for Dinant, Lochristi, and Molenbeek-Saint-Jean respectively. This result is consistent with the RMSE ranking reported in Table 12. Second, the $h = 3$ specification overestimates prices in all three municipalities (-15.4%, -21.3%, and -5.7%). Third, the $h = 1$ specification systematically underestimates prices, with errors of +17.2%, +5.6%, and +8.4%. In Lochristi, this represents a clear deterioration compared with the near-zero errors observed during the

backtesting period.

The ex-post evaluation for 2025 confirms the tendency of the $h = 1$ specification to underestimate prices across the three municipalities, with errors reaching +21.6%, +13.5%, and +12.6% respectively. Results for $h = 3$ are more heterogeneous (+12.4%, +1.0%, and -4.7%). Although the prediction error for Lochristi is close to zero, this result should be interpreted cautiously given the strong overestimation observed during the 2024 hold-out evaluation.

Beyond 2025, the $h = 3$ specification projects substantial price declines by 2026 relative to 2024 observed prices, amounting to -20.5%, -8.7%, and -11.6% for Dinant, Lochristi, and Molenbeek-Saint-Jean respectively. Since no comparable declines are observed in the historical sample, these projections are unlikely to reflect realistic market dynamics.

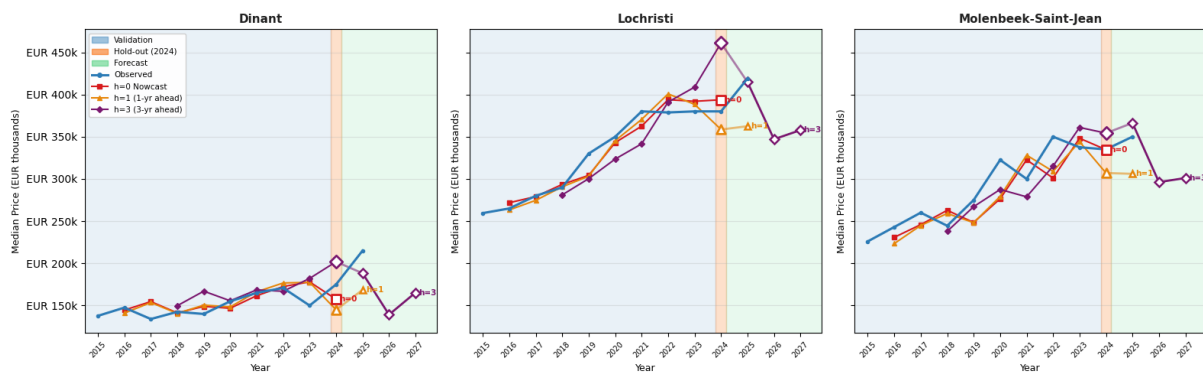


Figure 5: Historical and predicted median housing price for selected municipalities.

In conclusion, this chapter showed that the association between the RAI and housing prices depends on whether the analysis focuses on cross-sectional differences or within-municipality dynamics over time. The predictive analysis further indicated that forecasting performance differs across horizons and model specifications. The next chapter discusses these findings, together with the limitations and managerial implications of the study.

5 Discussion

This section discusses the main findings on the relationship between residential attractiveness and housing prices across Belgian municipalities, the limitations of the analysis and possible directions for future research, as well as the managerial and societal implications of the research.

5.1 Key Insights

The main research question examined the extent to which residential attractiveness explains housing prices and contributes to their prediction across Belgian municipalities. The study provides several key insights regarding these two dimensions.

From an explanatory perspective, the RAI exhibits a strong cross-sectional association with housing prices. Municipalities with higher RAI scores consistently display higher housing price levels, and this relationship remains robust across specifications, including when controlling for regional and provincial fixed effects. By contrast, year-on-year changes in municipal RAI scores are not associated with detectable price adjustments within the same municipality. The RAI therefore appears more informative for explaining the spatial distribution of housing prices across municipalities than their short-term evolution over time.

From a predictive perspective, the contribution of residential attractiveness beyond price persistence remains relatively limited. Past housing prices already capture a large share of the information reflected in the RAI composite, leaving only limited additional predictive value for the index across forecasting horizons. The specification based on the 21 individual indicators systematically outperforms the composite-based specifications, however, suggesting that aggregating the indicators into a single RAI score removes part of the predictive information contained in the underlying variables.

Across forecasting horizons, Ridge regression consistently outperforms the non-linear alternatives, suggesting that the relationship between municipal characteristics and housing prices is better approximated by a linear specification. In addition, macroeconomic variables improve predictive performance for forward-looking horizons, while contributing little to the contemporaneous specification, indicating that their informational value is mainly prospective. Although predictive performance remains relatively satisfactory for the contemporaneous horizon, it deteriorates substantially as the forecasting horizon increases. In particular, hold-out errors at the one- and three-year horizons correspond approximately to prediction errors of 11.9% and 15.3% in housing prices, respectively, which remain economically large. Taken together, these findings suggest that residential attractiveness provides only limited predictive accuracy for forecasting future housing prices across Belgian municipalities.

The hold-out results also reveal different patterns of forecast bias across horizons, although the following interpretations should be understood as plausible explanations rather than formally established mechanisms. At the one-year horizon, the model systematically underestimates realised prices. This may reflect the unusual post-2022 market environment, in which housing

prices continued to increase despite a sharp rise in mortgage rates, from 1.46% in 2021 to 3.59% in 2023. Because the training sample predominantly associates higher interest rates with weaker housing market conditions, the model appears to interpret elevated rates as a signal of slower price growth, leading to an underestimation of realised prices. At the three-year horizon, the opposite pattern emerges. Predictions rely on 2021 characteristics and macroeconomic conditions, a period still marked by historically low interest rates and strong housing market momentum. Consequently, the model tends to extrapolate more optimistic price dynamics, resulting in price projections that exceed the more moderate levels ultimately observed in 2024 as market conditions tightened.

Regarding the measurement sub-question, the results suggest that residential attractiveness can be meaningfully operationalised at the municipal level in Belgium through a multidimensional composite indicator. The RAI captures stable and interpretable territorial differences across municipalities while remaining sensitive to local variations over time. At the same time, the robustness analyses show that the measurement of residential attractiveness is not entirely neutral to methodological choices. While municipal rankings remain highly stable across compensatory specifications, they change substantially under the Mazziotta–Pareto approach, indicating that the treatment of compensation between indicators has important implications for how residential attractiveness is quantified. The results also suggest that some dimensions, particularly socio-economic conditions and accessibility, contribute more strongly to the observed spatial hierarchy than others. This likely reflects the underlying correlation structure of the data rather than explicit weighting choices.

Taken together, these findings point to two complementary interpretations, although both should be understood as hypotheses rather than empirically established causal mechanisms. First, the weak within-municipality association may reflect the structural inertia of residential attractiveness, as many dimensions captured by the RAI evolve slowly over time. Second, the limited predictive contribution of the RAI could suggest that territorial characteristics associated with residential attractiveness are already partly reflected in existing housing prices through capitalisation mechanisms. Overall, the RAI therefore appears more informative for understanding long-run spatial differences in housing prices than for explaining or predicting short-term housing price dynamics.

5.2 Limitations and Directions for Future Research

This study presents several limitations that help clarify the scope and interpretation of the results. First, the structure of the RAI itself may limit its ability to capture short-term housing market dynamics. Many indicators included in the index evolve only gradually over time, which reduces the amount of meaningful within-municipality variation available for estimation. As a result, changes in the RAI may be too slow or too diffuse to be associated with short-term housing price movements. Future research could explore alternative specifications relying on more reactive

indicators in order to better capture short-run local dynamics that may attract and retain residents. Moreover, the index remains sensitive to fluctuations in individual indicators. As illustrated in Section 4.1.2, relatively small absolute changes in indicators with low dispersion may generate large variations after standardization. Since the RAI is constructed through the aggregation of standardised indicators, these amplified values can exert a disproportionate influence on the final score and temporarily affect municipal rankings. Consequently, some short-term movements in the index may reflect statistical amplification effects rather than genuine structural changes in residential attractiveness. While the overall ranking structure remains highly stable, the interpretation of local year-to-year trajectories should therefore be approached with caution. Future research could further assess this source of uncertainty by evaluating the sensitivity of scores and rankings to small perturbations in the data or by constructing confidence intervals around index values through resampling techniques.

Furthermore, the RAI relies on an implicit assumption of homogeneous residential preferences across the population. As a population-averaged measure of a municipality's capacity to attract and retain residents, the index assumes that local characteristics contribute similarly to residential attractiveness across individuals. In practice, however, residential preferences are likely to differ across demographic groups. Families with children, retired households, and young professionals may value different aspects of a municipality depending on their needs and lifestyles. Consequently, a municipality that scores highly on the aggregate index may be particularly attractive to some groups while being less attractive to others. Future research could explore this heterogeneity by developing targeted sub-indices designed for specific population profiles, allowing for a more granular representation of residential attractiveness.

The econometric analysis is also subject to several limitations. First, the relationship between residential attractiveness and housing prices should be interpreted with caution due to potential endogeneity issues. As discussed by Alexandre et al. (2010), housing prices may simultaneously be both a determinant and an outcome of residential attractiveness. Consequently, the RAI cannot necessarily be considered exogenous to the housing market. Unobserved municipal characteristics may jointly influence housing prices and residential attractiveness in the between specifications, while housing market dynamics may also affect future values of the RAI in the within framework. These mechanisms may induce correlation between the explanatory variable and the error term, thereby violating the exogeneity assumptions underlying the estimators. The estimated coefficients should therefore be interpreted as conditional associations rather than causal effects.

In addition, a related limitation concerns the spatial dimension of the analysis. The presence of significant spatial autocorrelation in the residuals of the between model, as indicated by Moran's I , suggests that municipality-specific characteristics alone may not fully capture the spatial processes underlying housing prices. In practice, residential attractiveness may also depend on the characteristics of neighbouring municipalities, such as proximity to employment cen-

tres, infrastructure, amenities, or attractive residential environments located beyond municipal boundaries. Since the current RAI is constructed exclusively from municipality-level indicators, these potential spatial spillovers are not explicitly incorporated into the index. Future research could therefore extend the framework by integrating indicators capturing the characteristics of neighbouring municipalities.

The predictive analysis presents several additional limitations. Data availability represents an important constraint, as the analysis relies on an annual panel covering a relatively short time period. This limits the amount of temporal variation available for estimation and becomes particularly restrictive for the predictive analysis, where the number of backtest folds and hold-out years decreases substantially at longer forecasting horizons. As a result, predictive performance may partly reflect period-specific market conditions rather than stable forecasting ability. Future research could therefore benefit from longer time series, higher-frequency data, or access to private datasets allowing the analysis to cover a larger number of market cycles and economic regimes.

The predictive framework itself also remains limited in several respects. The analysis was restricted to Ridge, Random Forest, and XGBoost, while alternative implementations such as LightGBM (Ke et al., 2017) and CatBoost (Prokhorenkova et al., 2018) rely on different algorithmic refinements and have demonstrated competitive or superior performance relative to XGBoost on tabular datasets (Bentéjac et al., 2021). Whether these implementations could outperform the Ridge baseline in the present context remains an open empirical question.

Furthermore, although the M3 specification systematically outperforms the other models, the contribution of individual indicators to predictive accuracy remains difficult to isolate because correlated variables tend to share the estimated signal. Methods such as permutation importance (Breiman, 2001) could help identify which indicators carry the predictive signal beyond housing prices, and thereby inform future index design.

The predictive framework also does not explicitly account for the hierarchical structure of the data. Municipalities are nested within provinces and regions, and units sharing the same administrative level likely exhibit correlated unobserved characteristics that the current models do not account for. Hierarchical models incorporating random effects at the province or region level could better capture this multilevel structure and potentially improve out-of-sample generalisation (Gelman & Hill, 2007).

Additionally, the models only generate point forecasts and do not quantify predictive uncertainty. Conformal prediction methods could address this limitation by constructing prediction intervals around each forecast, thereby providing a direct measure of predictive reliability (Vovk et al., 2005).

Finally, a limitation arises from the measurement of housing prices. The dependent variable is defined as the municipal median price of all types of houses, which provides a representative level but does not capture the full distribution of prices within each municipality. As a result,

two municipalities with similar medians may exhibit very different internal market structures, for instance in terms of dispersion or segmentation. This aggregation smooths out variation within municipalities and may conceal important internal dynamics, such as rising prices in specific neighbourhoods or increasing disparities within the local housing market. Future research could address this limitation by refining the dependent variable, for instance by considering distributional measures such as price dispersion or quantiles within each municipality.

5.3 Managerial and Societal Implications

Beyond its academic contribution, the RAI provides practical insights for stakeholders involved in territorial planning and housing market analysis, while offering a framework to evaluate the broader societal externalities associated with residential attractiveness.

For **public authorities**, including municipalities and regional institutions, the leave-one-out sensitivity analysis (Table 7) and the PCA analysis (Section 4.1.1) highlight the importance of social conditions and accessibility in shaping residential attractiveness. This suggests that policies and investments targeting these dimensions may offer the greatest potential for improving a municipality's relative position. The RAI may therefore support territorial planning and resource allocation decisions by helping public authorities identify weaknesses, benchmark their performance against comparable municipalities, and monitor the evolution of residential attractiveness over time.

For **valuation professionals, investors, and market analysts**, the relationship between the RAI and housing prices suggests that residential attractiveness is largely reflected in long-run price levels. The between-estimator can therefore be used to derive an expected price level for each municipality based on its residential attractiveness profile. Comparing this expected value with observed market prices generates a residual that may provide useful market signals. Positive residuals may reflect speculative pressure, strong investor demand, or unobserved local amenities, whereas negative residuals may indicate relative undervaluation or local deterioration not yet fully incorporated into transaction prices. The RAI may therefore serve as a structural benchmark for identifying municipalities that appear relatively over- or undervalued given their underlying characteristics.

For **households**, the RAI provides an additional tool for comparing municipalities beyond housing prices alone. Municipalities within the same price range may differ substantially in terms of accessibility, infrastructure, social conditions, or governance quality. Consulting the RAI alongside housing prices may therefore help households identify municipalities offering a comparatively more attractive living environment within a similar budget range.

From a broader societal perspective, higher residential attractiveness may also generate unintended consequences for housing affordability. As municipalities improve their attractiveness through better infrastructure, public services, or social conditions, these improvements may become capitalised into housing prices. Consequently, the municipalities offering the most

favourable living environments may also become increasingly financially inaccessible. This pattern already appears visible in Belgium. In 2024, the median housing price reached 351,750 EUR in the top RAI decile. By comparison, the average new mortgage granted in Belgium amounted to only 187,632 EUR (Banque Nationale de Belgique, 2025b), meaning that the typical loan value remained substantially below the purchase price observed in the most attractive municipalities. In addition, Belgian prudential regulations require first-time buyers to provide a minimum down payment of 10% (Banque Nationale de Belgique, 2025a). For a property priced at 351,750 EUR, this represents an upfront payment of approximately 35,000 EUR, excluding notary fees and transaction costs. For many lower- and middle-income households, such financial requirements may constitute a substantial barrier to accessing these municipalities. Over time, these dynamics may reinforce socio-spatial inequalities, as municipalities with the greatest capacity to attract and retain residents become increasingly accessible primarily to higher-income households.

Taken together, the discussion highlighted that residential attractiveness was strongly associated with spatial differences in housing prices across Belgian municipalities, while providing limited information regarding their short-term evolution and prediction. It also underlined several methodological limitations, future research directions, and potential managerial and societal applications of the RAI. The next chapter concludes the thesis and summarises its main contributions.

6 Conclusion

This thesis examined the extent to which residential attractiveness contributes to explaining and predicting housing prices across Belgian municipalities, as well as how it can be operationalised and measured at the municipal level.

This thesis showed that residential attractiveness can be meaningfully represented through a multidimensional composite index combining 21 indicators across eight dimensions. Residential attractiveness appeared to reflect persistent territorial characteristics embedded within Belgian municipalities, and municipalities characterised by higher levels of residential attractiveness generally also exhibited higher housing price levels. Short-term changes within a municipality, however, did not appear to immediately translate into corresponding price changes, and the predictive contribution of the composite index remained limited. Predictive performance deteriorated substantially as the forecasting horizon increased, and aggregating indicators into a composite score removed part of the information relevant for forecasting. Overall, these findings suggest that residential attractiveness, when measured through a composite index, is more informative for understanding persistent differences in housing prices across municipalities than for explaining or predicting their short-term evolution.

From an academic perspective, this thesis provides one of the first municipal-level operationalisation of residential attractiveness in Belgium through a multidimensional composite index based on publicly available data. Beyond the construction of the index itself, the analysis highlights an important asymmetry between explanation and prediction. The thesis therefore contributes both methodologically, by proposing a transparent and replicable framework, and empirically, by showing that residential attractiveness appears more closely related to long-term spatial price structures than to short-term housing price dynamics.

Finally, while this thesis documents persistent differences in residential attractiveness across municipalities, it also raises an important policy question: to what extent can local authorities actually influence their capacity to attract and retain residents? If some dimensions of residential attractiveness can be improved through targeted local policies and investments, such as infrastructure, accessibility or public services, then municipalities may have real levers to strengthen their residential attractiveness. Better understanding which aspects of residential attractiveness can realistically be influenced by public action, and which are largely determined by factors beyond local control, therefore represents an avenue for future research.

References

- Actiris. (n.d.). Chiffres [dataset]. <https://www.actiris.brussels/fr/citoyens/chiffres/>
- Agentschap Digitaal Vlaanderen. (2026). Ogc api features haltes de lijn [dataset]. <https://www.vlaanderen.be/datavindplaats/catalogus/ogc-api-features-haltes-de-lijn>
- Agentschap Onderwijs Vlaanderen. (n.d.). Onderwijsaanbod lijst [dataset]. <https://data-onderwijs.vlaanderen.be/onderwijsaanbod/lijst?n=1+2&hz=true&hs=111+312>
- Alexandre, H., Cusin, F., & Juillard, C. (2010). L'attractivité résidentielle des agglomérations françaises. *Observateur de l'immobilier*, 76, 3–66. <https://hal.science/hal-00672601>
- Banque Nationale de Belgique. (2025a). Instruments macroprudentiels – immobilier : FAQ. <https://www.nbb.be/fr/supervision-financiere/politique-macroprudentielle/instruments-macroprudentiels/immobilier/faq>
- Banque Nationale de Belgique. (2025b). Quelles évolutions pour le marché des crédits hypothécaires ? <https://www.nbb.be/fr/actualites-et-evenements/actualites/quelles-evolutions-pour-le-marche-des-credits-hypothecaires>
- BBCDR. (n.d.). Finpro [dataset]. <https://finpro.bbcd.be>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305. <https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf>
- Blanchet, J., Hernandez, F., Nguyen, V. A., Pelger, M., & Zhang, X. (2023). Bayesian imputation with optimal look-ahead-bias and variance tradeoff. *arXiv preprint arXiv:2202.00871*. <https://doi.org/10.48550/arXiv.2202.00871>
- Boudt, K., Todorov, V., & Wang, W. (2020). Robust distribution-based winsorization in composite indicators construction. *Social Indicators Research*, 149, 375–397. <https://doi.org/10.1007/s11205-019-02259-w>
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–252. <http://www.jstor.org/stable/2984418>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brussels Open Data. (n.d.). Codes ins-nis et codes postaux - belgique [dataset]. <https://opendata.brussels.be/explore/dataset/codes-ins-nis-postaux-belgique/>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>

- Case, K. E., & Shiller, R. J. (1988, February). *The efficiency of the market for single-family homes* (NBER Working Paper No. 2506) (Available at SSRN: <https://ssrn.com/abstract=421766>). National Bureau of Economic Research.
- Chaze, M. (2017). *L'attractivité territoriale : Éléments de définition* (Research Report). UMR Territoires. <https://halshs.archives-ouvertes.fr/halshs-02520675>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chhetri, P., Stimson, R. J., & Western, J. (2006). Modelling the factors of neighbourhood attractiveness reflected in residential location decision choices. *Studies in Regional Science*, 36(2), 393–417. <https://doi.org/10.2457/srs.36.393>
- Cockx, L., & Canters, F. (2020). Determining heterogeneity of residential location preferences of households in Belgium. *Applied Geography*, 124, 102271. <https://doi.org/10.1016/j.apgeog.2020.102271>
- Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics*, 92(1), 1–45. [https://doi.org/10.1016/S0304-4076\(98\)00084-0](https://doi.org/10.1016/S0304-4076(98)00084-0)
- De Bruyne, K., & Van Hove, J. (2006). Explaining the spatial variation in housing prices: An economic geography approach. <https://doi.org/10.2139/ssrn.947671>
- Dijkstra, L., Papadimitriou, E., Cabeza Martinez, B., De Dominicis, L., & Kovacic, M. (2023). *EU regional competitiveness index 2.0: 2022 edition (revised May 2023)* (tech. rep.). Publications Office of the European Union. <https://doi.org/10.2776/46106>
- European Central Bank. (2024). Monetary financial institutions interest rate statistics [Accessed: 2026]. <https://data.ecb.europa.eu/data/datasets/MIR/MIR.M.U2.B.A2C.A.R.A.2250.EUR.N>
- Fadic, M., Garcilazo, E., Moreno Monroy, A. I., & Veneri, P. (2019). *Classifying small (TL3) regions for cross-country comparisons* (OECD Regional Development Working Papers No. 2019/10). OECD Publishing. <https://doi.org/10.1787/b902cc00-en>
- Finlay, J., & Williams, T. C. (2024). *Housing demand, inequality, and spatial sorting*. https://trevorewilliams.github.io/files/FinlayWilliams_Housing.pdf
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/Hierarchical models*. Cambridge University Press.
- Gómez-Limón, J. A., & Riesgo, L. (2009). Alternative approaches to the construction of a composite indicator of agricultural sustainability: An application to irrigated agriculture in the Duero basin in Spain. *Journal of Environmental Management*, 90(11), 3345–3362. <https://doi.org/10.1016/j.jenvman.2009.05.023>
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438. <https://doi.org/10.2307/1912791>

- Gude, V., Melendez, R., & Islam, M. M. (2025). Factors influencing anticipated migration patterns across metropolitan cities in the United States between 2018 and 2022. *Frontiers of Urban and Rural Planning*, 3(1). <https://doi.org/10.1007/s44243-025-00053-6>
- Guo, J. Y., & Bhat, C. R. (2001). *Residential location choice modeling: Accommodating sociodemographic, school quality and accessibility effects* (tech. rep.). The University of Texas at Austin. <http://caee.webhost.utexas.edu/prof/bhat/ABSTRACTS/residential.pdf>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hilber, C. A. L. (2011). The economic implications of house price capitalization: A survey of an emerging literature. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2403043>
- Ho, W. K., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, 38(1), 48–70. <https://doi.org/10.1080/09599916.2020.1832558>
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Hofer, K., Rieder, T., & Walert, F. (2024). *Residential attractiveness indicator* (tech. rep.). UBS Switzerland AG, Chief Investment Office GWM. <https://www.ubs.com/global/en/media/display-page-ndp/en-20241022-ubs-residential-attractiveness.html>
- Hooijen, I. (2021). *Place attractiveness: A study of the determinants playing a role in residential settlement behaviour* [Doctoral dissertation, Maastricht University]. <https://doi.org/10.26481/dis.20210205ih>
- Hope, A. C. A. (1968). A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(3), 582–598. <http://www.jstor.org/stable/2984263>
- Hsiao, C. (2003). *Analysis of panel data* (2nd ed.). Cambridge University Press.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
- IBSA. (n.d.). Finances des communes bruxelloises [dataset]. <https://ibsa.brussels/themes/finances-publiques/finances-des-communes-bruxelloises>
- Infrabel. (n.d.). Operational points of the railway network [dataset]. <https://infrabel.opendatasoft.com/explore/dataset/operationele-punten-van-het-netwerk/table/>
- IWEPS. (2015). *Indice des conditions de bien-être en Wallonie (ICBE) : Rapport de recherche (2e exercice, juin 2015)* (tech. rep.). IWEPS. https://www.iweps.be/wp-content/uploads/2017/01/ICBE_2015_rapport_de_recherche_final-19aout.pdf

- IWEPS. (n.d.-a). Catalogue des indicateurs statistiques - walstat (indicateur 235300) [dataset]. https://walstat.iweeps.be/walstat-catalogue.php?indicateur_id=235300
- IWEPS. (n.d.-b). Catalogue des indicateurs statistiques - walstat (indicateur 811500) [dataset]. https://walstat.iweeps.be/walstat-catalogue.php?indicateur_id=811500
- IWEPS. (n.d.-c). Catalogue des indicateurs statistiques - walstat (indicateur 811501) [dataset]. https://walstat.iweeps.be/walstat-catalogue.php?indicateur_id=811501
- IWEPS. (n.d.-d). Catalogue des indicateurs statistiques - walstat (indicateur 811506) [dataset]. https://walstat.iweeps.be/walstat-catalogue.php?indicateur_id=811506
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Light-GBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
- Limsombunchai, V., Gan, C., & Lee, M. (2004). House price prediction: Hedonic price model vs. artificial neural network. *American Journal of Applied Sciences*, 1(3), 193–201. <https://doi.org/10.3844/ajassp.2004.193.201>
- Mandemakers, J., Leidelmeijer, K., Burema, F., Halbersma, R., Middeldorp, M., & Veldkamp, J. (2021). *Instrumentontwikkeling Leefbaarometer 3.0* (tech. rep.). Atlas Research. <https://www.leefbaarometer.nl/resources/LBM3Instrumentontwikkeling.pdf>
- Mazziotta, M., & Pareto, A. (2021). Everything you always wanted to know about normalization (but were afraid to ask). *Rivista Italiana di Economia Demografia e Statistica*, 75(1), 41–52. <https://www.rieds-journal.org/rieds/article/view/29/27>
- Mazziotta, M., & Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one. *Social Indicators Research*, 116(2), 591–606. <https://doi.org/10.1007/s11205-013-0379-3>
- Meine Kinderbetreuung. (n.d.). Betreuungsangebot [dataset]. <https://www.meinekinderbetreuung.be/de/betreuungsangebot>
- Ministère de la Fédération Wallonie-Bruxelles. (2025). Fwb age fichier signalétique des établissements d'enseignement de la fédération wallonie-bruxelles (fase) 2025 [dataset]. <https://www.odwb.be/explore/dataset/fwb-age-fichier-signaletique-des-etablissements-d-enseignement-de-la-federation-/information/>
- Ministerium der Deutschsprachigen Gemeinschaft. (n.d.). Ostbelgien bildung [dataset]. <https://ostbelgienbildung.be/desktopdefault.aspx/tabid-2185/>
- Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17–23. <https://doi.org/10.2307/2332142>
- Niedomysl, T. (2006). *Migration and place attractiveness* [Doctoral dissertation, Uppsala University]. <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-6873>
- OECD, European Union, & European Commission Joint Research Centre. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. <https://doi.org/10.1787/9789264043466-en>

- ONE. (n.d.). Recherche de milieux d'accueil [dataset]. <https://my.one.be/search?theme=ACCUEIL>
- Öner, Ö. (2017). Retail city: The relationship between place attractiveness and accessibility to shops. *Spatial Economic Analysis*, 12(1), 72–91. <https://doi.org/10.1080/17421772.2017.1265663>
- Open Data Wallonie-Bruxelles. (2025). Communes gemeente belgium [dataset]. <https://data.europa.eu/data/datasets/https-www-odwb-be-explore-dataset-communesgemeente-belgium-?locale=undefined>
- OpenStreetMap contributors. (n.d.). Openstreetmap [dataset]. <https://www.openstreetmap.org>
- Opgroeien. (n.d.). Kinderopvangzoekers [dataset]. <https://www.opgroeien.be/kinderopvangzoekers>
- OTW. (2020). Tec gtfs [dataset]. <https://transportdata.be/dataset/tec-gtfs>
- Pai, P.-F., & Wang, W.-C. (2020). Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences*, 10(17), 5832. <https://doi.org/10.3390/app10175832>
- Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications*, 42(6), 2928–2934. <https://doi.org/10.1016/j.eswa.2014.11.040>
- Peng, C., Wu, X., Fu, Y., & Lai, K. K. (2017). Alternative approaches to constructing composite indicators: An application to construct a sustainable energy index for APEC economies. *Operational Research*, 17(3), 747–759. <https://doi.org/10.1007/s12351-016-0235-z>
- Poirot, J., & Gérardin, H. (2010). L'attractivité des territoires : Un concept multidimensionnel. *Mondes en développement*, 149(1), 27–41. <https://doi.org/10.3917/med.149.0027>
- Police fédérale belge. (n.d.). Tableau des statistiques de faits criminels [dataset]. <https://www.police.be/statistiques/fr/criminalite/faits-criminels/tableau>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, 6639–6649. https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55. <https://www.jstor.org/stable/1830899>
- Ryu, I. S., Gorelova, A., Flood, M., & Donoso Salas, J. I. (2024). *A granular approach to measuring regional attractiveness* (OECD Regional Development Papers No. 100). OECD Publishing. <https://doi.org/10.1787/4f0866f1-en>
- Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2), 307–323. <https://doi.org/10.1111/j.1467-985X.2005.00350.x>
- Sasse, L., Nicolaisen-Sobesky, E., Dukart, J., Eickhoff, S. B., Götz, M., Hamdan, S., Komeyer, V., Kulkarni, A., Lahnakoski, J., Love, B. C., Raimondo, F., & Patil, K. R. (2023). On

- leakage in machine learning pipelines. *arXiv preprint*. <https://doi.org/10.48550/ARXIV.2311.04179>
- Sentometrics. (2024). Economic sentiment indicator [Accessed: 2026]. <https://sentometrics.com/fr/timo-browser?indicator=Economic+Sentiment+Indicator>
- SPF Finances. (n.d.). Taux de la taxe communale additionnelle à l'impôt des personnes physiques [dataset]. <https://data.gov.be/fr/datasets/tauxippcom>
- SPF Santé publique. (n.d.). Établissements de soins de santé [dataset]. <https://www.health.belgium.be/fr/professionnels/professionnels-sante/sante-humaine/organisation-soins-sante/etablissements-soins-sante>
- Statbel. (2024). House prices in belgium, 2024. <https://statbel.fgov.be/en/news/house-prices-2024>
- Statbel. (n.d.-a). Datalab - revenu disponible administratif par commune [dataset]. <https://statbel.fgov.be/fr/open-data/datalab-revenu-disponible-administratif-par-commune>
- Statbel. (n.d.). Géographie et classifications territoriales. <https://statbel.fgov.be/fr/propos-de-statbel/methodologie/classifications/geographie>
- Statbel. (n.d.-b). Land use according to the land register [dataset]. <https://statbel.fgov.be/en/themes/housing/land-use-according-land-register>
- Statbel. (n.d.-c). Parc des bâtiments [dataset]. <https://statbel.fgov.be/fr/themes/construction-logement/parc-des-batiments>
- Statbel. (n.d.-d). Permis de bâtir par mois, région, province, arrondissement et commune [dataset]. <https://statbel.fgov.be/fr/open-data/permis-de-batir-par-mois-region-province-arrondissement-et-commune>
- Statbel. (n.d.-e). Structure de la population [dataset]. <https://statbel.fgov.be/fr/themes/population/structure-de-la-population>
- Statbel. (n.d.-f). Ventes de biens immobiliers selon la nature du bien dans l'acte de vente - belgique [dataset]. <https://statbel.fgov.be/fr/open-data/ventes-de-biens-immobiliers-selon-la-nature-du-bien-dans-lacte-de-vente-belgique>
- Steunpunt Werk. (n.d.). Vlaamse arbeidsrekening: Pendel [dataset]. <https://www.steunpuntwerk.be/cijfers/vlaamse-arbeidsrekening>
- STIB-MIVB. (n.d.). Stops by line - production [dataset]. <https://stibmivb.opendatasoft.com/explore/dataset/stops-by-line-production/export/>
- Szaja, M. (2021). Applying the VMCM method to assess the level of residential attractiveness of municipalities in the West Pomeranian Voivodeship. *Procedia Computer Science*, 192, 4701–4710. <https://doi.org/10.1016/j.procs.2021.09.248>
- Vovk, V., Gammernan, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer. <https://doi.org/10.1007/b106715>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

- Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western Cengage Learning.
- World Health Organization. (2012). *WHOQOL: Measuring quality of life* (tech. rep. No. WHO/HIS/HSI Rev.2012.03). World Health Organization. <https://www.who.int/tools/whoqol>

Appendices

Appendix A Data Construction and Indicator Selection

This appendix documents the data underlying the Residential Attractiveness Index (RAI). It details the initial data sources, provides precise definitions of the dimensions and indicators, and presents the final set of variables retained for the analysis.

A.1 Initial Data Sources

The construction of the RAI relies on a wide range of raw datasets collected from national and regional institutions. Table 13 summarises the main data sources, their time coverage, and their role in the analysis.

Table 13: Detailed raw data sources

Area	Raw Data	Coverage	Source	Use
Brussels	Geolocated public transport stops (bus, tram, metro)	2026	(STIB-MIVB, n.d.)	A1
Flanders	Geolocated public transport stops (bus, tram)	2026	(Agentschap Digitaal Vlaanderen, 2026)	A1
Wallonia	Geolocated public transport stops (bus)	2026	(OTW, 2020)	A1
Belgium	Road network (motorway, trunk, primary, secondary, tertiary)	2026	(OpenStreetMap contributors, n.d.)	A2
Belgium	General hospital campus sites (geolocated)	2026	(SPF Santé publique, n.d.)	A3
Belgium	Total population and population by age	1992–2025	(Statbel, n.d.-e)	D1, D2, D3
Flanders	Unemployment rate	2017–2023	(Steunpunt Werk, n.d.)	E2
Wallonia	Unemployment rate	2008–2024	(IWEPS, n.d.-a)	E2
Brussels	Unemployment rate	1992–2024	(Actiris, n.d.)	E2
Belgium	Median disposable income (MS_MEDIAN)	2015–2023	(Statbel, n.d.-a)	E1
Belgium	Municipal additional personal income tax rate (IPP)	1990–2023	(SPF Finances, n.d.)	G1
Brussels	Municipal debt	2010–2023	(IBSA, n.d.)	G2
Wallonia	Municipal debt per capita	2013–2023	(IWEPS, n.d.-d)	G2
Flanders	Municipal debt per capita	2014–2024	(BBCDR, n.d.)	G2
Brussels	Ordinary budget balance	2010–2023	(IBSA, n.d.)	G3
Flanders	Ordinary budget per capita	2020–2025	(BBCDR, n.d.)	G3
Wallonia	Ordinary revenues and expenditures	2013–2023	(IWEPS, n.d.-b, n.d.-c)	G3
Belgium	Newly authorised residential dwellings	1996–2024	(Statbel, n.d.-d)	H1
Belgium	Total number of buildings	1995–2028	(Statbel, n.d.-c)	H2, H3
Belgium	Buildings constructed after 2011	1995–2025	(Statbel, n.d.-c)	H2

Area	Raw Data	Coverage	Source	Use
Belgium	Buildings constructed before 1900	1995–2025	(Statbel, n.d.-c)	H3
Flanders	Educational institutions (kindergarten to secondary)	2026	(Agentschap Onderwijs Vlaanderen, n.d.)	I1
German Community	Educational institutions (kindergarten to secondary)	2026	(Ministerium der Deutschsprachigen Gemeinschaft, n.d.)	I1
Wallonia-Brussels Federation	Educational institutions (kindergarten to secondary)	2026	(Ministère de la Fédération Wallonie-Bruxelles, 2025)	I1
Flanders	Childcare facilities	2026	(Opgroeien, n.d.)	I2
German Community	Childcare facilities	2026	(Meine Kinderbetreuung, n.d.)	I2
Wallonia-Brussels Federation	Childcare facilities	2026	(ONE, n.d.)	I2
Belgium	Railway stations and stops	2026	(Infrabel, n.d.)	I3
Belgium	Crime counts by municipality	2015–2024	(Police fédérale belge, n.d.)	SC1
Belgium	Risk of poverty rate	2015–2023	(Statbel, n.d.-a)	SC2
Belgium	Cadastral land-use categories (22 categories)	1983–2025	(Statbel, n.d.-b)	U1, U3
Belgium	Total population	1992–2025	(Statbel, n.d.-e)	U2
Belgium	Median sale price of individual houses	2010–2025	(Statbel, n.d.-f)	Target
Belgium	Municipality identifiers (NIS code, geometry) (2025)	–	(Open Data Wallonie-Bruxelles, 2025)	Matching
Belgium	Postal codes	–	(Brussels Open Data, n.d.)	Matching
Belgium	Surface area	2024	(Statbel, n.d.-b)	Scaling

A.2 Definitions of Dimensions and Indicators

Residential attractiveness is conceptualised as the capacity of a municipality to provide favourable living conditions through observable territorial characteristics. It is structured into eight complementary dimensions reflecting distinct mechanisms influencing residential choices.

Dimensions

Table 14: Conceptual definition of dimensions

Dimension	Definition
Accessibility	Dimension capturing the position of a municipality within transport and service networks.
Demography	Dimension capturing population structure and dynamics.
Economy	Dimension capturing local economic conditions.

Dimension	Definition
Governance	Dimension capturing the financial capacity and management of local authorities.
Housing	Dimension capturing the characteristics and dynamics of the housing stock.
Infrastructure	Dimension capturing the availability of key public services and facilities supporting daily life.
Social conditions	Dimension capturing safety, inequality and socio-economic vulnerability affecting living conditions.
Urban form	Dimension capturing land-use patterns and spatial organisation.

Indicators

Table 15: Definition of indicators

ID	Definition
A1	Number of public transport stops per square kilometre, capturing the spatial availability of public transport infrastructure within the municipality.
A2	Total length of structurally significant road segments relative to municipal surface area, capturing road accessibility and connectivity.
A3	Euclidean distance between the municipal centroid and the nearest general hospital site, capturing access to healthcare services.
D1	Year-on-year percentage change in total resident population, capturing demographic growth dynamics.
D2	Share of population aged 0–18 in total population, reflecting the presence of younger cohorts.
D3	Share of population aged 65 and over, capturing population ageing.
E1	Median disposable income of households, reflecting local purchasing power.
E2	Share of the labour force that is unemployed, capturing labour market conditions.
G1	Municipal surcharge rate applied to federal personal income tax, reflecting local fiscal pressure.
G2	Total municipal debt divided by population, capturing the level of indebtedness relative to municipality size.
G3	Difference between ordinary revenues and expenditures per capita, capturing the municipality's operating fiscal balance.
H1	Number of newly authorised residential dwellings per 1,000 inhabitants, capturing recent housing development intensity.
H2	Share of buildings constructed after 2011 in total building stock, reflecting recent residential development.
H3	Share of buildings constructed before 1900, capturing the historical structure of the housing stock.
I1	Number of educational establishments per 1,000 children aged 3–17, capturing availability of schooling infrastructure.
I2	Number of childcare facilities per 1,000 children aged 0–2, capturing availability of early childhood services.
I3	Binary indicator equal to 1 if at least one railway station or stop is present in the municipality, capturing access to rail transport.

ID	Definition
SC1	Number of recorded criminal offences per 1,000 inhabitants, capturing the level of criminal activity.
SC2	Share of population below the at-risk-of-poverty threshold, capturing socio-economic vulnerability.
U1	Share of cadastral surface corresponding to green and natural land uses, capturing environmental quality.
U2	Total resident population divided by municipal surface area, capturing settlement density.
U3	Shannon entropy index computed over cadastral land-use shares, measuring the degree of functional diversity in land use.

A.3 Final Indicator Set

Following the selection process, a set of indicators is retained to represent the different dimensions of residential attractiveness. Table 16 reports the final indicator set, along with their polarity and time coverage.

Table 16: Final indicator set

Dimension	ID	Polarity	Coverage
Accessibility	A1	+	2026
Accessibility	A2	+	2026
Accessibility	A3	-	2026
Demography	D1	+	1992–2025
Demography	D2	+	1992–2025
Economy	E1	+	2015–2023
Economy	E2	-	2008–2024
Governance	G1	-	1990–2023
Governance	G2	-	2013–2024
Governance	G3	+	2010–2025
Housing	H1	+	1996–2024
Housing	H2	+	1995–2025
Housing	H3	-	1995–2025
Infrastructure	I1	+	2026
Infrastructure	I2	+	2026
Infrastructure	I3	+	2026
Social Conditions	SC1	-	2015–2024
Social Conditions	SC2	-	2015–2023
Urban Form	U1	+	1983–2025
Urban Form	U2	+	1992–2025
Urban Form	U3	+	1983–2025

Appendix B Data Preprocessing

B.1 Correspondence table for municipal mergers

Table 17 presents the correspondence between pre-merger NSI codes and the updated municipalities resulting from the 2024 and 2025 municipal mergers (Statbel, n.d.).

Table 17: Correspondence table between pre-merger and 2025 NSI codes

Old NSI code	Municipality (before merger)	New NSI code	Municipality (2025 classification)
82003	Bastogne	82039	Bastogne
82005	Bertogne	82039	Bastogne
11002	Antwerp	11002	Antwerp
11007	Borsbeek	11002	Antwerp
23023	Galmaarden	23106	Pajottegem
23024	Gooik	23106	Pajottegem
23032	Herne	23106	Pajottegem
37012	Ruiselede	37021	Wingene
37018	Wingene	37021	Wingene
37007	Meulebeke	37022	Tielt
37015	Tielt	37022	Tielt
44012	De Pinte	44086	Nazareth-De Pinte
44048	Nazareth	44086	Nazareth-De Pinte
44034	Lochristi	44087	Lochristi
44073	Wachtebeke	44087	Lochristi
46014	Lokeren	46029	Lokeren
44045	Moerbeke	46029	Lokeren
44040	Melle	44088	Merelbeke-Melle
44043	Merelbeke	44088	Merelbeke-Melle
46003	Beveren	46030	Beveren-Kruibeke-Zwijndrecht
46013	Kruibeke	46030	Beveren-Kruibeke-Zwijndrecht
11056	Zwijndrecht	46030	Beveren-Kruibeke-Zwijndrecht
73006	Bilzen	73110	Bilzen-Hoeselt
73032	Hoeselt	73110	Bilzen-Hoeselt
73009	Looz	73111	Tongeren-Borgloon
73083	Tongeren	73111	Tongeren-Borgloon
71069	Ham	71071	Tessenderlo-Ham
71057	Tessenderlo	71071	Tessenderlo-Ham
71022	Hasselt	71072	Hasselt
73040	Kortesseem	71072	Hasselt

B.2 Missingness patterns and imputation methods

Table 18: Missingness patterns and imputation methods by indicator

Indicator	Pattern	Method	Count	% of observations
E1	Right truncation	LOCF	564	9.98
	Left truncation	BOCF	5	0.09
	Persistent	Spatial median	10	0.18
	Total	–	579	10.25
E2	Right truncation	LOCF	284	5.03
	Left truncation	BOCF	1136	20.11
	Persistent	Spatial median	10	0.18
	Total	–	1430	25.31
G1	Left truncation	BOCF	28	0.50
	Total	–	28	0.50
G2	Right truncation	LOCF	289	5.12
	Left truncation	BOCF	17	0.30
	Gap	Spatial median	5	0.09
	Total	–	311	5.50
G3	Right truncation	LOCF	288	5.10
	Left truncation	BOCF	1425	25.22
	Gap	Spatial median	1	0.02
	Persistent	Spatial median	10	0.18
	Total	–	1724	30.51
H1	Right truncation	LOCF	9	0.16
	Total	–	9	0.16
H2	Left truncation	BOCF	1128	19.96
	Persistent	Spatial median	10	0.18
	Total	–	1138	20.14

Continued on next page

Indicator	Pattern	Method	Count	% of observations
H3	Persistent	Spatial median	10	0.18
	Total	–	10	0.18
SC2	Right truncation	LOCF	564	9.98
	Left truncation	BOCF	5	0.09
	Persistent	Spatial median	10	0.18
	Total	–	579	10.25

B.3 Distributional diagnostics and transformation examples

To reduce extreme asymmetry and improve distributional balance, several candidate transformations were evaluated for skewed indicators. The figures below present representative examples of the preprocessing procedure using histograms and Q–Q plots.

Indicator H1 exhibited strong positive skewness together with zero values. The $\log(1+x)$ transformation substantially improved the distributional balance while preserving zero observations.

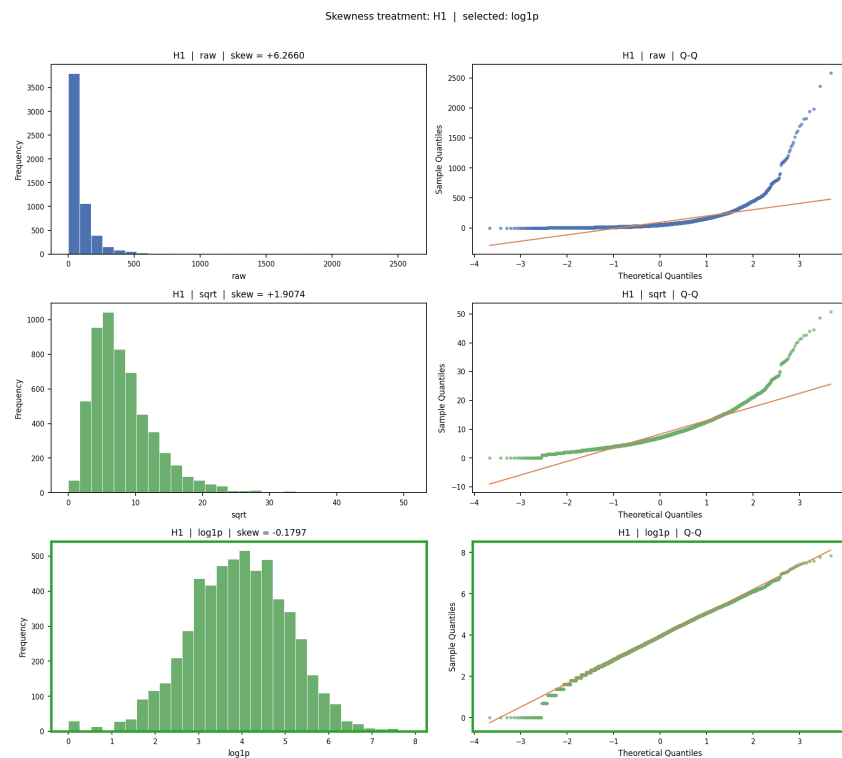


Figure 6: Distributional diagnostics and candidate transformations for indicator H1

Indicator U2 displayed substantial positive skewness. Among the candidate transformations evaluated, the logarithmic transformation provided the strongest reduction in asymmetry.

Skewness treatment: U2 | selected: log

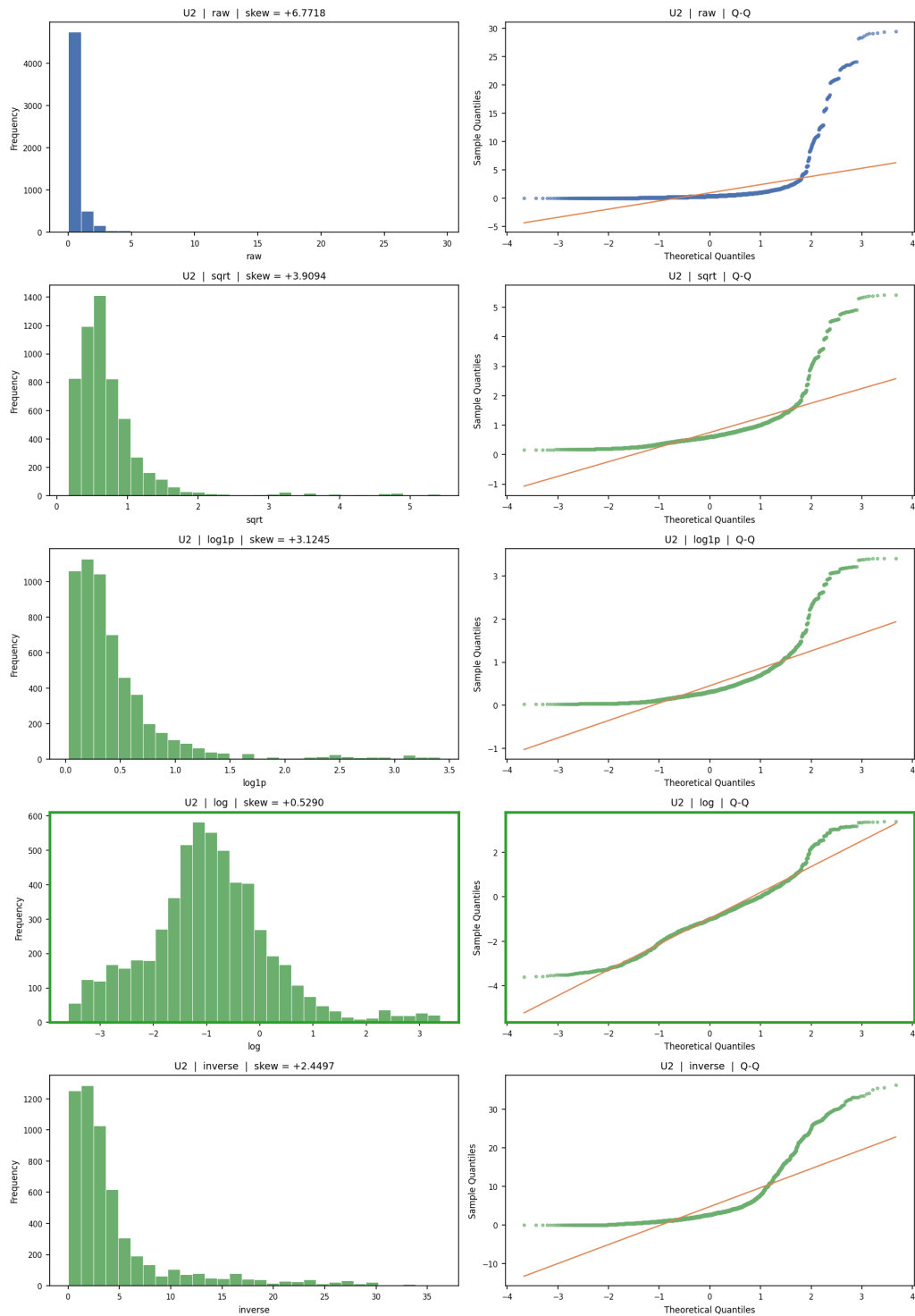


Figure 7: Distributional diagnostics and candidate transformations for indicator U2

For indicator G3, standard transformations remained insufficient to adequately reduce skewness. Winsorisation at the 1st and 99th percentiles was therefore retained to limit the influence of extreme values.

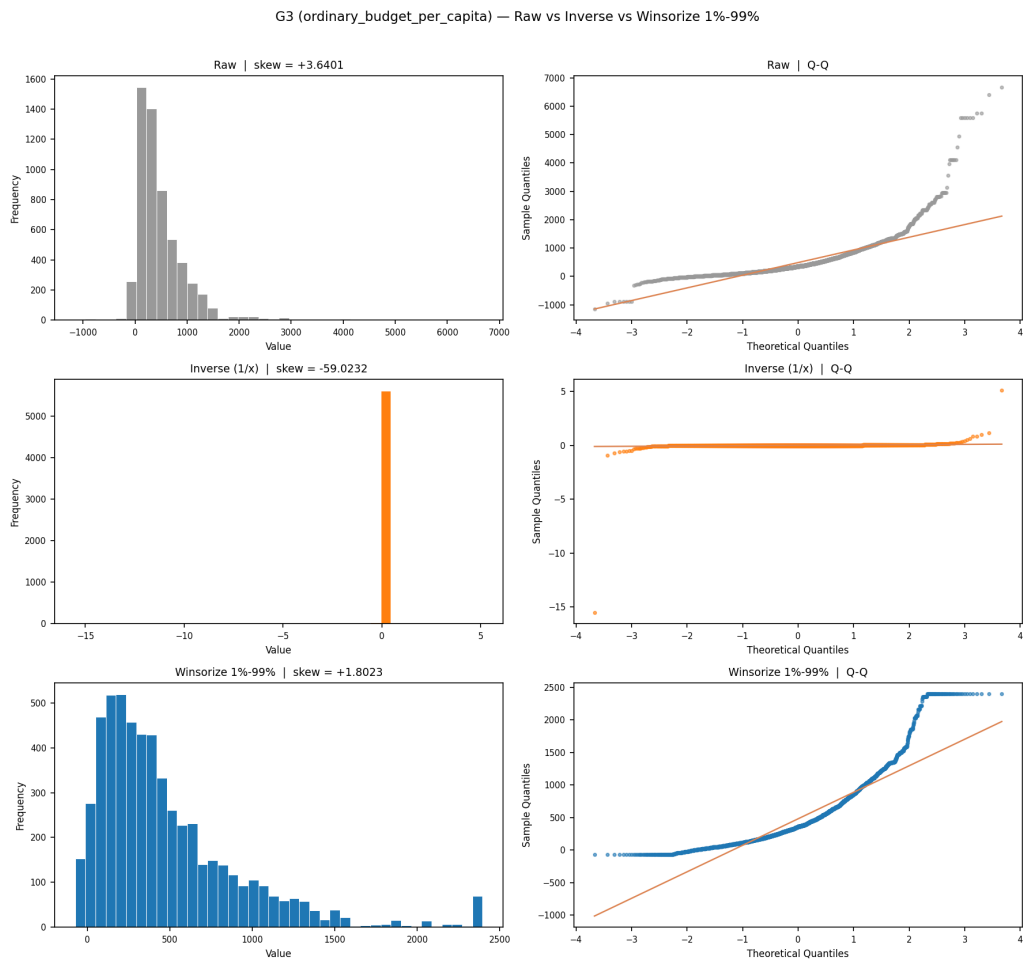


Figure 8: Distributional diagnostics and winsorisation for indicator G3

Indicator G1 exhibited pronounced negative skewness. Polynomial transformations were evaluated, with the square transformation retained as the best compromise between skewness reduction and distributional preservation.

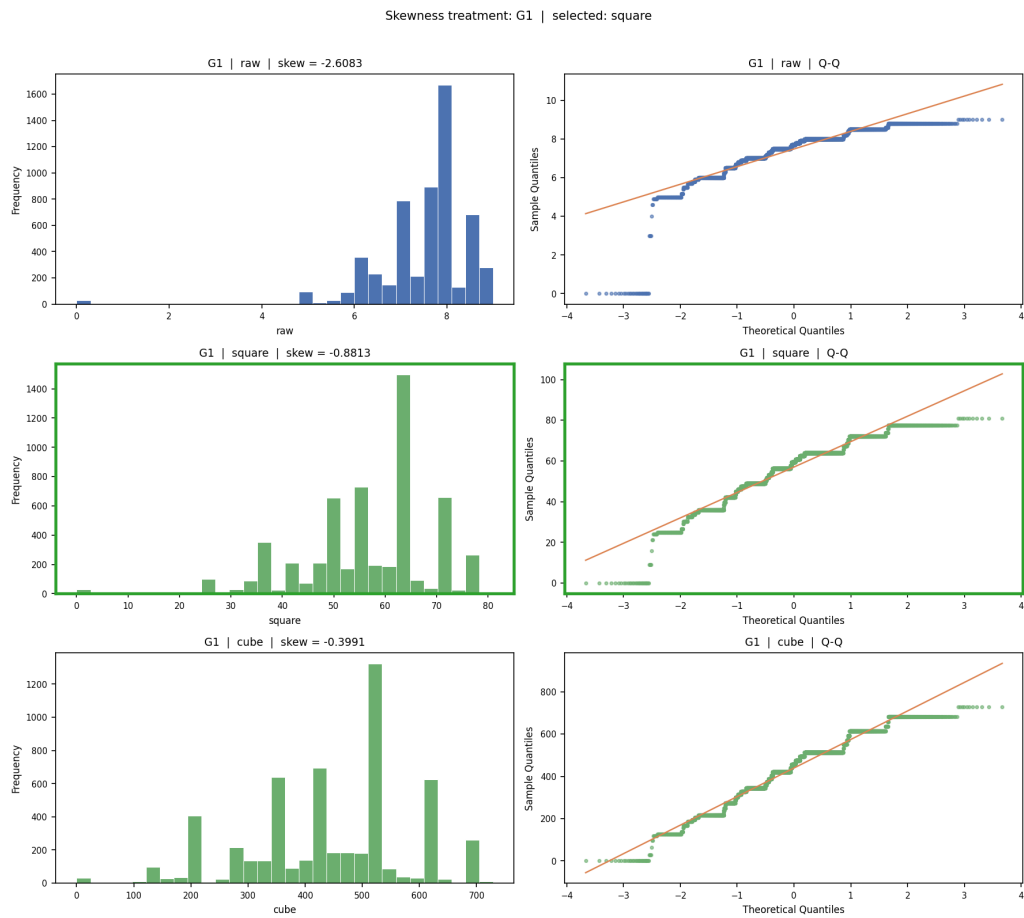


Figure 9: Distributional diagnostics and candidate transformations for indicator G1

Indicator A2 remained moderately skewed after transformation. The square-root transformation was nevertheless retained as the most appropriate compromise among the candidate treatments considered.

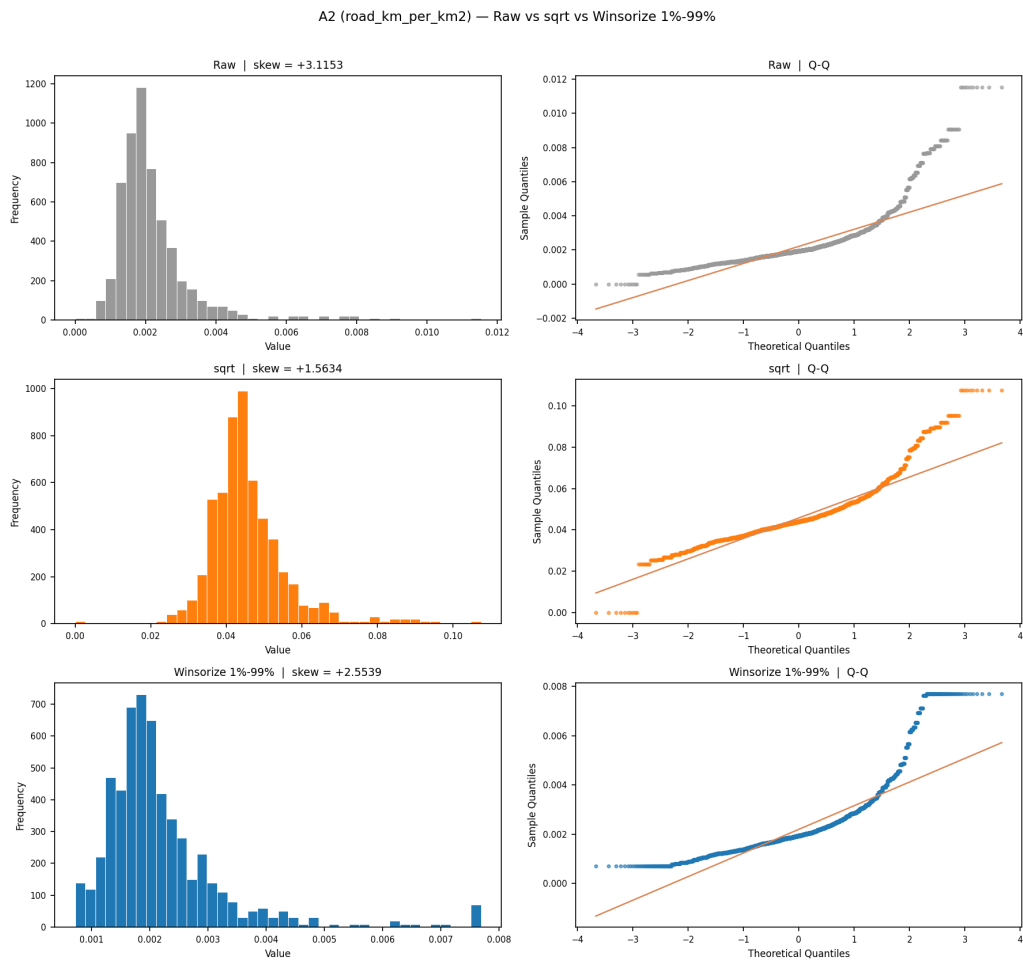


Figure 10: Distributional diagnostics and candidate transformations for indicator A2

©

Appendix C Correlation Matrix

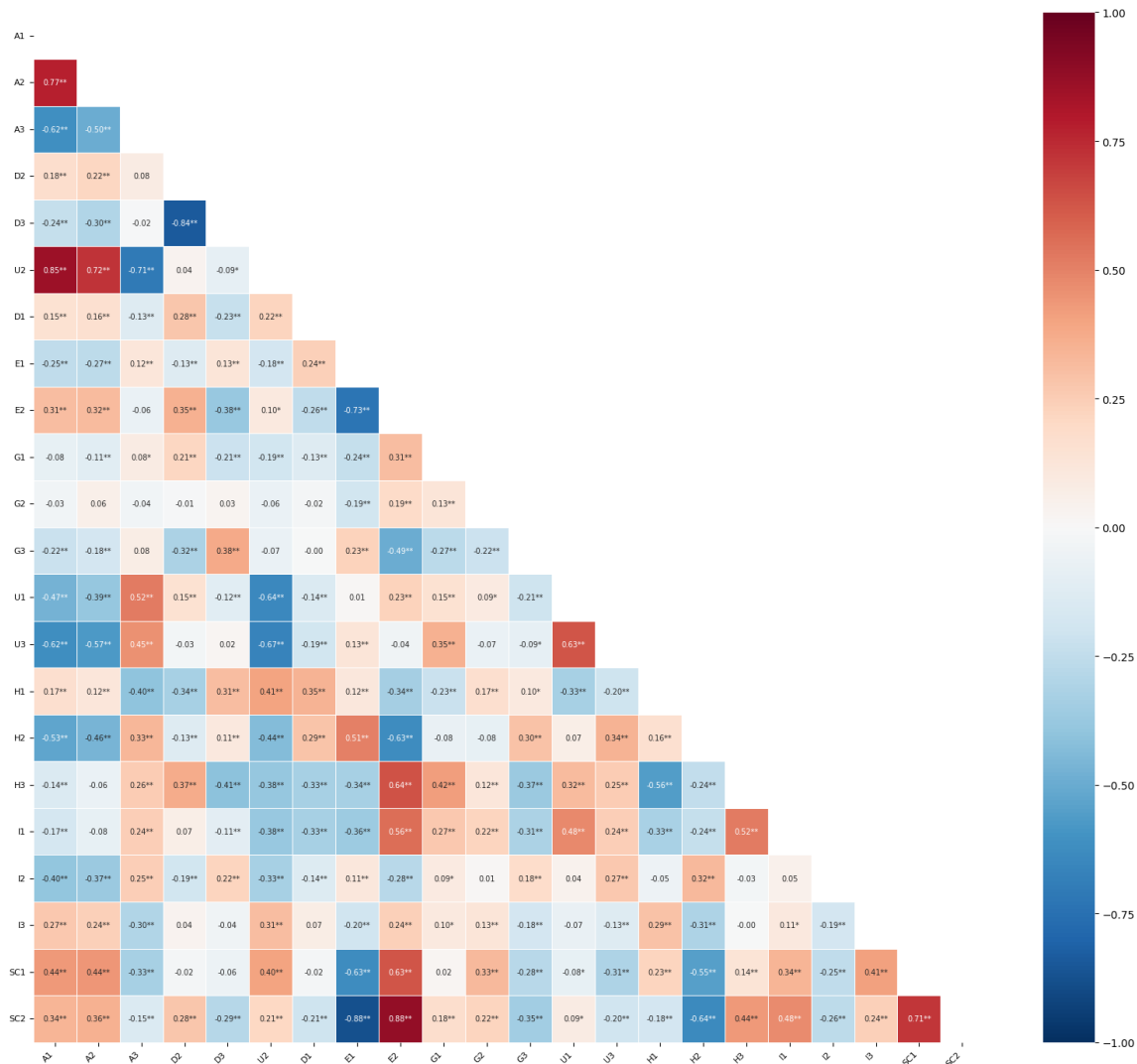


Figure 11: Correlation matrix of the candidate indicators.

Appendix D Principal Component Analysis

This appendix reports the complete set of Principal Component Analysis (PCA) diagnostics, including eigenvalues, explained variance, and graphical representations.

D.1 Eigenvalues and Explained Variance

Table 19: Eigenvalues and Explained Variance

Component	Eigenvalue	Variance Explained (%)	Cumulative Variance (%)
PC1	6.08	28.92	28.92
PC2	3.93	18.68	47.60
PC3	1.78	8.46	56.06
PC4	1.51	7.18	63.24

Component	Eigenvalue	Variance Explained (%)	Cumulative Variance (%)
PC5	1.27	6.02	69.26
PC6	0.96	4.54	73.80
PC7	0.78	3.70	77.50
PC8	0.72	3.44	80.94
PC9	0.63	2.99	83.93
PC10	0.55	2.62	86.55
PC11	0.49	2.33	88.88
PC12	0.45	2.16	91.04
PC13	0.41	1.95	92.99
PC14	0.40	1.90	94.89
PC15	0.30	1.43	96.31
PC16	0.23	1.09	97.40
PC17	0.20	0.94	98.34
PC18	0.17	0.80	99.15
PC19	0.11	0.52	99.67
PC20	0.04	0.19	99.86
PC21	0.03	0.14	100.00

D.2 Scree Plot

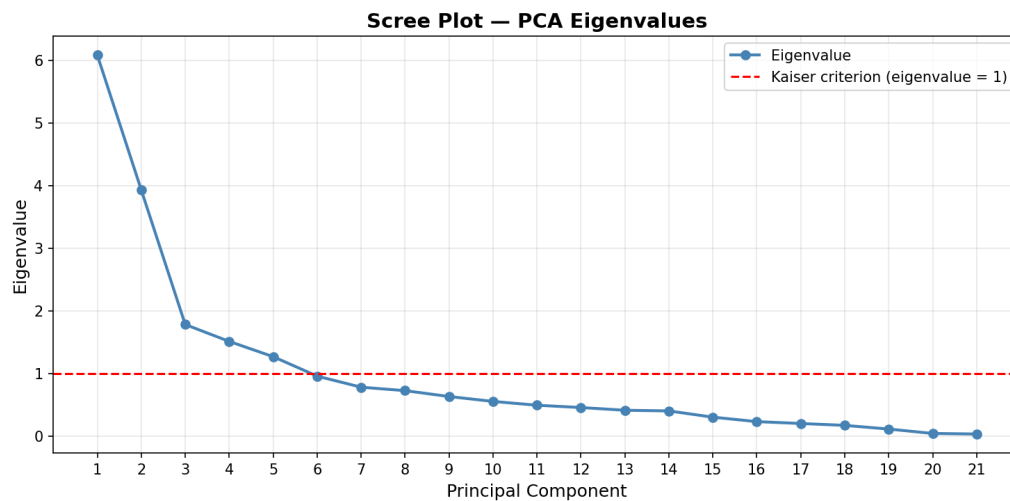


Figure 12: Scree plot of PCA eigenvalues.

D.3 Cumulative Explained Variance

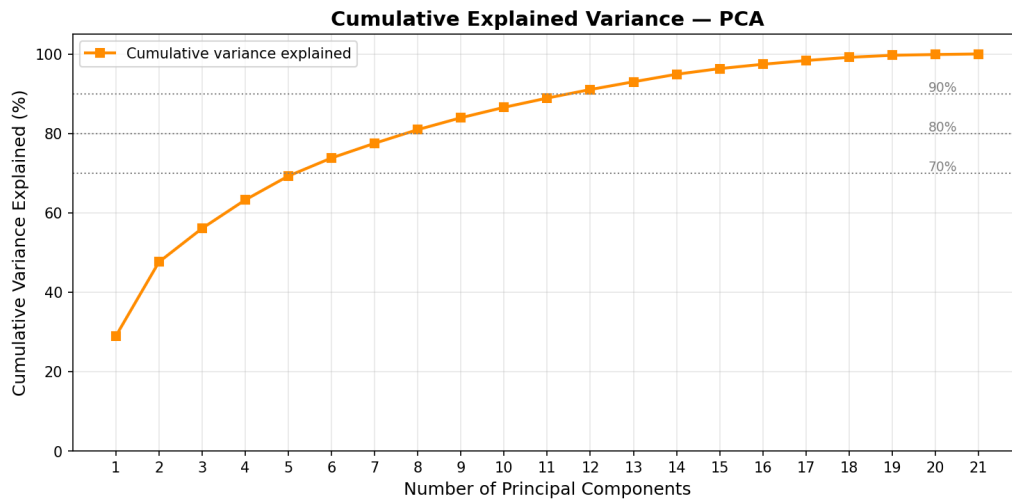


Figure 13: Cumulative explained variance across principal components.

Appendix E Missing data treatment for housing prices

Table 20: Missingness patterns and imputation methods for median housing prices

Variable	Pattern	Method	Count	% of observations
Median price	Gap	Spatial median	71	1.26
	Persistent	Spatial median	11	0.19
	Right truncation	Spatial median	21	0.37
	Left truncation	Spatial median	34	0.60
	Total	–	137	2.42

Appendix F Descriptive Diagnostics

F.1 Box–Cox Transformation

Table 21: Box–Cox estimation results

Statistic	Value
Number of observations	5,650
Optimal λ^*	-0.028
95% CI for λ	[-0.080, 0.020]
$\lambda = 0$ in CI	Yes

Note: The estimated confidence interval includes $\lambda = 0$, supporting the use of the logarithmic transformation (Box and Cox, 1964).

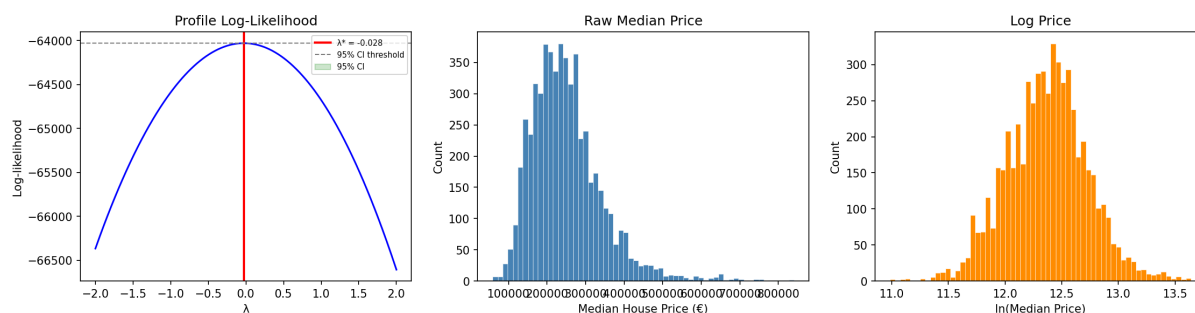


Figure 14: Box–Cox profile likelihood and distribution of housing prices

Note: The profile likelihood is maximized near $\lambda = 0$, confirming the appropriateness of the log specification.

F.2 Variance Decomposition

Table 22: Between–within variance decomposition

Variable	Total SD	Between SD	Within SD	Within/Total
RAI	0.293	0.282	0.082	0.281
ln(Price)	0.355	0.323	0.146	0.413

Note: SD = standard deviation. The within/total ratio measures the share of variation coming from changes over time within municipalities.

F.3 Cross-Sectional and Within Associations

Table 23: Correlation between RAI and ln(Price)

Type	Correlation (r)	p-value	N
Cross-sectional (between)	0.638	< 0.001	565
Within (demeaned)	0.580	< 0.001	5,650

RAI $\times$ ln(House Price): Cross-Sectional vs Within

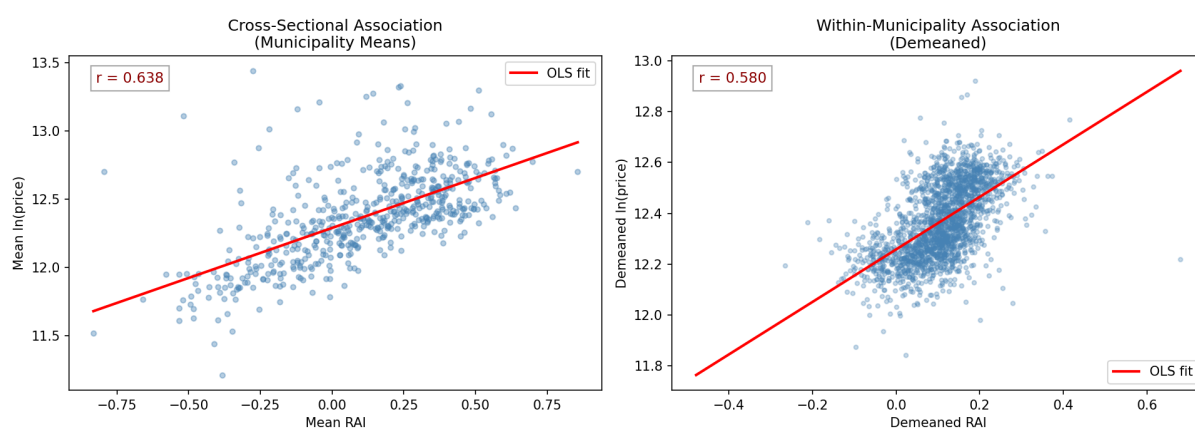


Figure 15: Associations between RAI and ln(Price)

Appendix G Econometric Robustness

G.1 Econometric Robustness Checks

Table 24: Summary of robustness checks

Specification	$\hat{\beta}$	SE	p-val.	Sig.
<i>Panel A — Additional tests</i>				
Granger ($k = 1$)	F = 0.222		0.638	
Granger ($k = 2$)	F = 0.194		0.824	
Granger ($k = 3$)	F = 0.805		0.491	
<i>Panel B — Within</i>				
Baseline	0.016	0.023	0.494	
Min-max	-0.003	0.172	0.988	
PCA	-0.023	0.025	0.347	
Geometric	-0.0003	0.018	0.988	
MPI	0.0000	0.0000	0.003	**
<i>Panel C — Between</i>				
Baseline	0.733	0.095	<0.001	***
Min-max	3.948	0.619	<0.001	***
PCA	0.682	0.088	<0.001	***
Geometric	0.506	0.175	0.004	**
MPI	-0.0000	0.0001	0.965	
<i>Panel D — Levels</i>				
Between	160,530	21,562	<0.001	***
TWFE	-4,593	8,544	0.591	

Note: Dependent variable: $\ln(\text{median housing price})$ unless stated. Between-model standard errors are Conley (1999) spatial standard errors

with a 25 km cutoff. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.10$.

G.2 Spatial Dependence

Table 25: Moran's I test for spatial autocorrelation

Residual type	Moran's I	z-score	p-val.	Sig.
Between-municipality residuals	0.6887	26.94	0.001	***
Within-municipality residuals	-0.0040	-0.09	0.485	

Note: Positive values indicate positive spatial autocorrelation, values close to zero indicate no spatial pattern, and negative values indicate

negative spatial autocorrelation. *** $p < 0.001$.

G.3 Alternative Conley Distance Cutoffs

Table 26: Conley spatial standard errors under alternative distance cutoffs

Specification	Cutoff	$\hat{\beta}$	SE	Sig.
Baseline	10 km	0.733	0.092	***
Baseline	25 km	0.733	0.095	***
Baseline	50 km	0.733	0.112	***
Province FE	10 km	0.597	0.053	***
Province FE	25 km	0.597	0.069	***
Province FE	50 km	0.597	0.087	***
Region FE	10 km	0.738	0.076	***
Region FE	25 km	0.738	0.107	***
Region FE	50 km	0.738	0.117	***

Note: Conley (1999) spatial standard errors computed using alternative distance cutoffs. *** $p < 0.001$.

G.4 Rank-Based Within Specifications

To assess whether the within-municipality results are sensitive to the scale of the Residential Attractiveness Index, the main TWFE specifications are replicated using the within-year percentile

rank of the RAI instead of its continuous score.

Table 27: Rank-based robustness checks for within estimates

Specification	Score			Rank		
	$\hat{\beta}$	p-val.	Sig.	$\hat{\beta}$	p-val.	Sig.
Baseline, lag = 0	0.016	0.494		-0.005	0.819	
Lag = 1	0.010	0.638		0.006	0.769	
Lag = 2	0.013	0.560		0.012	0.614	
Lag = 3	0.047	0.093	.	0.068	0.016	*

Note: Rank specifications use within-year percentile rankings of the RAI instead of the continuous index values. *** $p < 0.001$, ** $p < 0.01$,

* $p < 0.05$, . $p < 0.10$.

Appendix H Hyperparameter search space

Model	Hyperparameter	Search space
Ridge	α	Log-uniform [10^{-5} , 10^3]
Randomn_estimators Forest	max_depth	{50, 100, 200}
	min_samples_leaf	{3, 5, 7, 10, None}
	min_samples_split	Integer [2, 30]
	max_features	{sqrt, log2, 0.5, 0.8}
XGBoostn_estimators	max_depth	{50, 100, 200}
	min_child_weight	Integer [2, 8]
	learning_rate	Log-uniform [0.01, 0.3]
	subsample	Uniform [0.6, 1.0]
	colsample_bytree	Uniform [0.5, 1.0]
	min_child_weight	Integer [1, 10]
	gamma	Uniform [0, 5]
	reg_alpha	Log-uniform [10^{-4} , 10]
	reg_lambda	Log-uniform [10^{-4} , 10]
	lambda1	Log-uniform [10^{-4} , 10]

Appendix I Best Performing Ridge Specifications

I.1 Estimated coefficients

Table 28 reports the estimated coefficients of the best-performing Ridge specifications. All predictors were standardized prior to estimation within each horizon-specific training sample. Comparisons across horizons should therefore be interpreted cautiously: coefficients reflect standardized units that are not directly comparable across specifications, and the models differ in predictor set, sample size, and regularization intensity (Table 29).

The price variable remains the dominant predictor across all specifications by a substantial margin, confirming the strong autoregressive nature of municipal housing prices. Its coefficient declines slightly from $h = 0$ to $h = 3$, though this comparison should be interpreted with caution as the predictor differs between specifications: M1 uses lagged prices while M2 uses contemporaneous prices.

The RAI displays the most stable pattern in the table, with coefficients remaining between 0.016 and 0.017 across all three horizons despite changes in specification, sample size, and predictor set. This consistency is notable given that the two specifications differ in their feature composition.

At $h = 1$, mortgage rates rank above the RAI in absolute coefficient magnitude. At $h = 3$, their coefficient more than doubles relative to $h = 1$, while the RAI remains virtually unchanged. This suggests that mortgage rates capture an effect that strengthens at longer horizons, whereas the RAI contribution is invariant to horizon.

Economic sentiment is included in M2 at both $h = 1$ and $h = 3$. At $h = 1$ its coefficient is negligible. At $h = 3$ it changes sign and reaches a magnitude comparable to the RAI, despite all three models operating under near-zero regularization. This sign instability across horizons suggests that economic sentiment captures no robust predictive signal in this setting.

Table 28: Estimated Ridge coefficients by forecasting horizon

Horizon	Predictor	Coefficient
$h = 0$	$\ln(\text{Price}_{i,t-1})$	0.3258
	$\text{RAI}_{i,t}$	0.0160
$h = 1$	$\ln(\text{Price}_{i,t})$	0.3224
	mortgage_rate_t	-0.0169
	$\text{RAI}_{i,t}$	0.0161
	econ_sentiment_t	-0.0024
$h = 3$	$\ln(\text{Price}_{i,t})$	0.3131
	mortgage_rate_t	-0.0390
	$\text{RAI}_{i,t}$	0.0166
	econ_sentiment_t	0.0151

I.2 Model fit statistics

The regularization parameter α is taken from the last backtest fold for each horizon

Table 29: regularization parameters of the selected Ridge models

Horizon	α	n
$h = 0$	1.03×10^{-5}	4,520
$h = 1$	1.00×10^{-5}	4,520
$h = 3$	3.19×10^{-5}	3,390

Appendix J GitHub Repository

The complete codebase developed for this thesis, including data preprocessing, construction of the Residential Attractiveness Index, econometric analyses, predictive modelling, is available at: https://github.com/lisadounia/Master_Thesis_Lisa_Dounia

