Advancing housing market statistics in Europe using web-scraped data

Rui Evangelista, Eurostat([1]) ([2])

Abstract

This paper examines the feasibility of producing experimental monthly residential house price indices and price level indicators using web-scraped real estate listings. Building on a previous Eurostat-supported study, the paper draws lessons from the construction of asking price-based indices, the temporal disaggregation of quarterly transaction-based indices, and best practices in web-scraping strategies.

The study focuses on country case-studies, where data coverage and portal representativeness are strongest. Using hedonic regression methods, monthly indices are constructed and, where possible, compared with official quarterly transaction-based indices to assess their potential as early indicators of housing market developments. Other indicators, such as price levels are also derived to provide complementary spatially and temporally granular information.

Results demonstrate that high-quality web-scraped data can produce experimental statistics that are broadly consistent with transaction-based statistics while offering greater timeliness and detail. The paper concludes with recommendations for improving web-scraping practices, ensuring data representativeness, and integrating such experimental statistics into Eurostat's statistical innovation framework.

JEL Codes: C43, C81, R31

Keywords: house price indices, web-scraping, experimental statistics.

([1]) The views expressed in this paper are solely those of the author and do not necessarily reflect the views of Eurostat, the European Commission, or any other institution. the author before citing.

([2]) Draft. Please contact the author before citing. *Comments are welcome. rui.evangelista@ec.europa.eu*

1. Introduction

Housing market statistics are essential for monitoring macroeconomic conditions, informing housing and social policy, and supporting financial-stability analysis. In the European Union (EU), official house price indicators are well by limitations in timeliness, frequency, granularity, and territorial detail.

Recent policy developments, including the European Affordable Housing Plan ([3]), stress that EU-wide indices remain indispensable but that important gaps persist in higher-frequency and more detailed market metrics (European Commission, 2025). These include the lack of consistent statistics on absolute housing prices and rents, limited segmentation, and insufficient detail by property type and small-area geographies.

Against this background, web-scraped real estate listings can complement transaction-based statistics. Listings data containing asking prices are available at high frequency and include rich property characteristics, enabling quality adjustment through hedonic methods. Although they are not a like-for-like substitute for transaction prices, carefully cleaned and standardised listings data can underpin experimental indicators that are more timely and granular than many official outputs.

This paper builds on a Eurostat-supported feasibility study commissioned by Eurostat (KR&A and MMIT, 2025), which covered 13 European countries and focuses on settings where portal coverage and representativeness are strongest. It develops two use cases that could expand the range of housing market statistics in EU countries, namely monthly price level indicators relevant for affordability analysis and asking-price indices as earlier signals of transaction-based measures, including as a potential input to experimental flash euro area and EU aggregates.

This paper is organised as follows. Section 2 summarises key findings from the feasibility work. Section 3 reviews the relevant literature. Section 4 presents the data and methodology. Section 5 reports empirical results for selected countries. Section 6 concludes by synthesising the main findings and outlining lessons learned and next steps for the further development of experimental housing market statistic.

([3]) The European Affordable Housing Plan - Housing - European Commission.

2. Study on the monthly house price indices from web-scraped data

This section summarises the main findings of a cross-country feasibility study commission by Eurostat (KR&A and MMIT 2025) that assessed whether housing listings scraped from online real estate portals can complement and strengthen official house price statistics.

The study evaluated the feasibility of producing monthly house price indices and their potential to increase frequency and improve timeliness relative to existing quarterly outputs. Currently, in the European Statistical System, official HPIs and OOHPIs are compiled quarterly and are typically disseminated around three months after the end of the reference quarter ([4]).

2.1. Data generated by the feasibility study

Data were obtained from real estate portals either through web scraping or via direct provision by portal owners. In total, 13 portals from the same number of EU countries ([5]) were covered.

Data collection began in March 2023 and reached full coverage by May 2023, comprising four file deliveries from portal owners and nine web scrapers. Collection continued through March/ April 2025, ensuring at least two full years of observations for each of the 13 portals.

The choice of the portals was guided by both statistical considerations and practical constraints. Portals were first shortlisted from a longlist of 104 sites and ranked using criteria such as market leadership and 'scrapability', while also aiming for a broad geographical spread and minimum content in primary attributes such as asking price, size, dwelling type, new or existing status and location ([6]).

([4]) The European Statistical System a partnership between Eurostat and the national statistical authorities of EU Member States, with participation from EFTA countries, that develops, produces and disseminates European statistics. The legal basis underpinning the calculation and dissemination of the HPIs and OOHPIs is available online from Legislation - Housing price statistics - Eurostat .

([5]) The countries covered by the study were Austria, Belgium, Cyprus, Czechia, Denmark, Finland, Hungary, Lithuania, Luxembourg, Malta, Romania, Slovakia and Spain.

([6]) Market leadership was assessed using web traffic proxies such as visitor counts, which informed the country rankings of portals. 'Scrapability' was evaluated through a combined legal and technical screening, including terms of use and robots.txt restrictions, and practical feasibility based on page structure, JavaScript dependence and anti-bot measures.

Across the collection window, the feasibility study assembled approximately 34.8 million offers from the selected portals (around 1.4 million offers per month), with within-run duplicates removed where the same property ID appeared more than once.

A first cleaning stage then excluded offers deemed unusable for index construction, namely foreign properties (about 900 thousand), non-residential listings (about 1.2 million), rental offers (about 1.2 million) and records without an asking price (about 1.2 million), leaving around 30.2 million usable offers (relating to 6.7 million distinct properties, observed on average 4.5 times). The following table summarises the outcome of the study.

Table 1: Portal coverage and dataset volumes in the Eurostat-supported feasibility study

Country	Portal	Period covered	No. of obs. (offers)	Distinct properties	Average no. of appearances
Austria	FindMyHome	May 2023–Apr 2025	165 284	59 465	2.8
Belgium	Immovlan	May 2023–Apr 2025	1 169 702	394 257	3.0
Cyprus	Bazaraki	May 2023–Apr 2025	527 681	134 749	3.9
Czechia	Sreality	May 2023–Apr 2025	970 936	255 953	3.8
Denmark	Boliga	May 2023–Apr 2025	1 058 194	204 273	5.2
Finland	Oikotie	May 2023–Apr 2025	1 117 515	247 040	2.8
Hungary	Ingatlan	Mar 2023–Apr 2025	5 071 064	991 323	5.1
Lithuania	Kampas	May 2023–Apr 2025	279 171	73 466	3.8
Luxembourg	AtHome	May 2023–Apr 2025	244 975	84 881	2.9
Malta	Remax	Apr 2023–Apr 2025	156 623	22 572	6.9
Romania	Imobiliare	Mar 2023–Apr 2025	4 435 197	1 213 974	3.7
Slovakia	Nehnutelnosti	Apr 2023–Apr 2025	833 235	262 550	3.2
Spain	Habitaclia	Mar 2023–Apr 2025	14 249 227	2 785 831	5.1
TOTAL			**30 278 804**	**6 730 334**	**4.5**

Source: KR&A and MMIT (2025).

All 13 portals covered listings for apartments, houses and plots of land, except in Spain where data for plots of land were not available from the portal.

2.2. Practical lessons from the study on web-scraping strategies

Two main lessons on web-scraping strategies emerge from the feasibility study. The first concerns data accessibility and, by extension, source selection. Portals were initially shortlisted using statistical criteria such as market leadership, likely coverage of the advertised market and the availability of key attributes, but these criteria could not be applied rigidly in practice.

In line with web-scraping best practice (Eurostat, 2020), portal owners were approached to seek permission, yet many did not respond, refused cooperation, prohibited scraping in their terms and conditions, or requested remuneration or offered only aggregated outputs rather than offer-level data ([7]). The resulting collection strategy was therefore pragmatic, combining nine operational scrapers with four stable monthly file deliveries based on direct portal agreements ([8]).

This pragmatism affects how representative the data are and helps explain why results differ across countries in the next sections. Where the study could rely on a dominant portal, the listings are more likely to reflect the broader advertised market.

This is the case in Czechia, where Sreality is described as the biggest real estate website in the country, and, to a lesser extent, in Romania, where Imobiliare is a major portal with wide reach and substantial listing volumes. In Austria, by contrast, evidence on FindMyHome suggests narrower market coverage, which may limit how well the data capture overall market activity.

The second lesson is the trade-off between stability and data richness when comparing direct file deliveries with web scraping. Direct portal agreements can improve continuity and basic data quality through stable, guaranteed feeds, but they often provide only core variables, limiting the scope for richer modelling and quality adjustment.

Web scraping can capture additional attributes, yet it is operationally fragile, as website changes and anti-bot measures require frequent maintenance and can create gaps or discontinuities. In practice, however, the study did not find systematic differences in index performance between the two transmission modes beyond the reduced modelling flexibility implied by variable-poor file deliveries.

The study also highlighted practical engineering and data management requirements. A mixed architecture was typically needed, combining simple HTTP collection for static pages with browser automation for dynamic sites, supported by safeguards such as rate limiting to reduce

([7]) It should be noted that the revision of the EU Statistical Law, Regulation (EU) 2024/3018, which entered into force on 26 December 2024 and was not applied in this study, may improve access conditions. It strengthens statistical authorities' access to administrative data and, for the first time, introduces an obligation for private data holders to make relevant data available for official statistics, subject to clear safeguards, proportionality requirements and enforcement provisions.

([8]) There was no financial compensation involved in the four situations in which there was a direct agreement with the portal for the transmission of data files.

blocking risk, checkpointing to restart from the last completed page after failures, and routine volume checks to detect incomplete runs. Downstream, quality depends on robust pipelines that stored raw data, transformed them into a harmonised model, standardised heterogeneous formats, and implemented de-duplication as a core statistical step. Privacy by design is essential, notably by minimising or excluding personal data and by handling location via privacy-preserving links to statistical area classifications that prevent identification while preserving analytical utility for later modelling.

Taken together, these lessons point to implementation priorities for the European Statistical System. These include strengthening relationships with major portals to secure access, deploying monitoring to detect breaks in collection and variable coverage, and investing in scalable infrastructure and modern deployment practices to minimise downtime when sites change.

These priorities align with the aim of this paper, which is to translate operational experience into guidance for producing experimental indicators from web-based housing data.

2.3. Construction of monthly asking price indices

In the assessment study, the construction of quality-adjusted monthly price indices new and existing dwellings and plots of land was based on a standard pooled hedonic time-dummy regression ([9]):

$$Ln(A_{it}) = \beta_0 + \sum_{k=1}^{K} \beta_k \, x_{k,it} + \sum_{\tau=2}^{T} \delta_\tau D_{\tau,it} + \varepsilon_{it} \qquad (1)$$

Where:

A_{it} = asking price of dwelling *i* observed in period *t*

x_{kit} = characteristic *k* of dwelling i (e.g. floor area, location dummies, etc)

$D_{\tau,it}$ = 1 if observation *i* is in period τ, otherwise *0*

In (1), period 1 is the base period and is omitted to avoid perfect multicollinearity.

([9]) For more information on this hedonic method and its use in the compilation of European statistics, see Linz et al. (2009).

Quality-adjusted price indices are obtained by exponentiating the estimated time-dummy coefficients and scaling the results so that the base period equals 100 ([10]):

$$I_\tau = 100 \times exp(\delta_\tau), \tau = 1, \dots, T;\ \delta_1 = 0 \qquad (2)$$

The set of characteristics in (1) was selected using an automatic forward stepwise procedure based on the Bayesian information criterion. To improve comparability and maximise overlap across country datasets over time, only observations from May 2023 through March 2025 (23 months) were used to estimate the monthly price indices.

The derivation of the indices included systematic treatment of outliers and inconsistencies (including reused identifiers, ambiguous property type coding and extreme values in price and size variables), and a selection strategy to approximate market activity by using the latest observation of each property within a month.

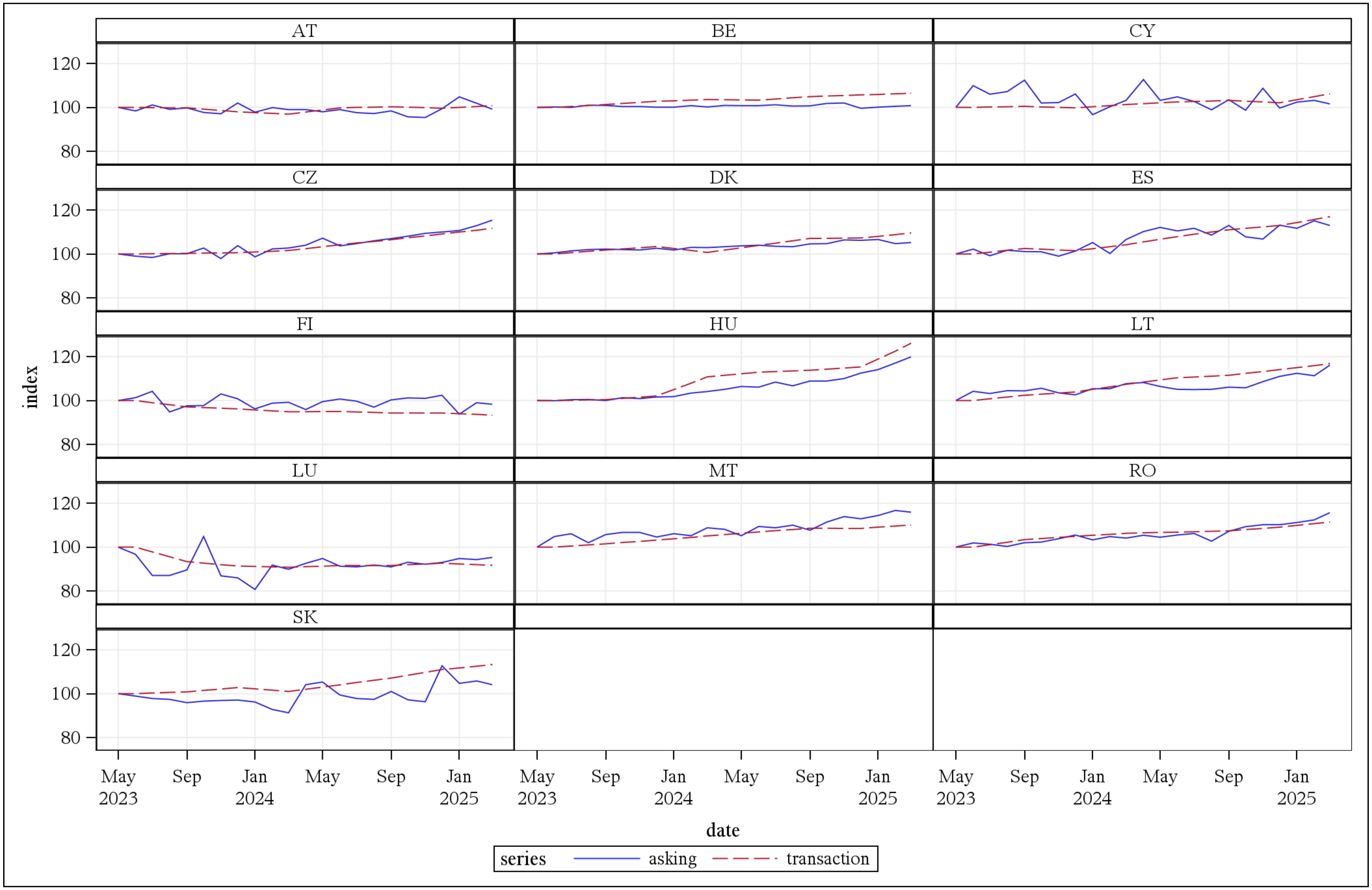
This 'latest offer' design aimed to reduce the influence of long time-on-market listings in monthly comparisons and to bring the timing of asking prices closer to the point at which a listing leaves the market.

Core hedonic covariates were generally available in sufficient detail for modelling, notably asking price, floor area, dwelling type, a location proxy and an indicator distinguishing new from existing properties. By contrast, secondary characteristics such as year of construction and energy performance were inconsistently available across sources, and the depth of attribute information varied markedly by portal.

Figure 1 below provides an overview of the results. For benchmarking, linearly interpolated quarterly HPIs are used as the reference series for the overall HPI.

([10]) Since the index is obtained by exponentiating the estimated time-dummy coefficients, the non-linear back-transformation can induce a small finite-sample bias (Kennedy, 1981). This correction was not applied because it is negligible here given the small standard errors of the time-dummy estimates in our large sample

Figure 1: Monthly web-scraped asking-price indices and interpolated transaction-based HPIs, by country (May 2023 = 100)



Source: KR&A and MMIT (2025), complemented by the author's calculations on the interpolated HPI.

When the resulting asking price indices were compared with quarterly transaction-based indices, outcomes were mixed but informative. Table 2 provides the Mean Error (ME) and Mean Absolute Error (MAE) measures for the differences found between the asking price and transaction-based indices.

Table 2: Bias and precision

Country	Portal	ME	MAE
Austria	FindMyHome	0.3	1.8
Belgium	Immovlan	2.8	2.8
Cyprus	Bazaraki	-2.0	3.9
Czechia	Sreality	-0.6	1.4
Denmark	Boliga	0.6	1.4
Finland	Oikotie	-3.5	3.8
Hungary	Ingatlan	3.5	3.6
Lithuania	Kampas	1.5	2.8
Luxembourg	AtHome	0.9	3.4
Malta	Remax	-3.0	3.2
Romania	Imobiliare	0.3	1.5
Slovakia	Nehnutelnosti	5.2	5.8
Spain	Habitaclia	0.1	2.3

Source: KR&A and MMIT (2025). Own calculations.

Visual inspection of the graphs alongside ME and MAE values suggests that alignment is strongest in markets where the selected portal is perceived as a market leader or where listing coverage is sufficiently broad and dynamic. In particular, Czechia (Sreality) and Denmark (Boliga) show the smallest typical deviations (MAE equal to 1.4 in both cases) with limited average bias (−0.6 and 0.6, respectively).

Romania (Imobiliare) also performs well (MAE iqual to 1.5 and ME equal to 0.3), consistent with earlier evidence of good market dynamism. Spain (Habitaclia) combines near-zero average bias (ME equal to 0.1) with large listing volumes. Although its MAE is somewhat higher (2.3), the series broadly tracks the quarterly transaction-based profile and provides a useful high-frequency signal, albeit with greater intra-quarter volatility than the benchmark.

In the remaining markets, larger typical deviations and, in some cases, sizeable mean biases point to weaker alignment. This is most pronounced in Slovakia (MAE equal to 5.8 and ME equal to 5.2), and also in Cyprus (MAE equal to 3.9 and ME equal to −2.0), Finland (MAE equal to 3.8 and ME equal to −3.5) and Hungary (MAE equal to 3.6; and ME equal to 3.5).

For residential building plots, however, results were consistent across countries in showing that index production for this property type was substantially more challenging. Low volumes of data, missing key attributes such as plot size in some sources, and sizeable unobserved heterogeneity resulted in more volatile series, making these plot indices less credible as experimental statistics under current conditions. Figure 2 below exemplifies this situation, where the case for Finland is shown (the Annex provides examples for other countries).

Figure 2: Monthly asking-price indices for building plots for Finland (May 2023 = 100)

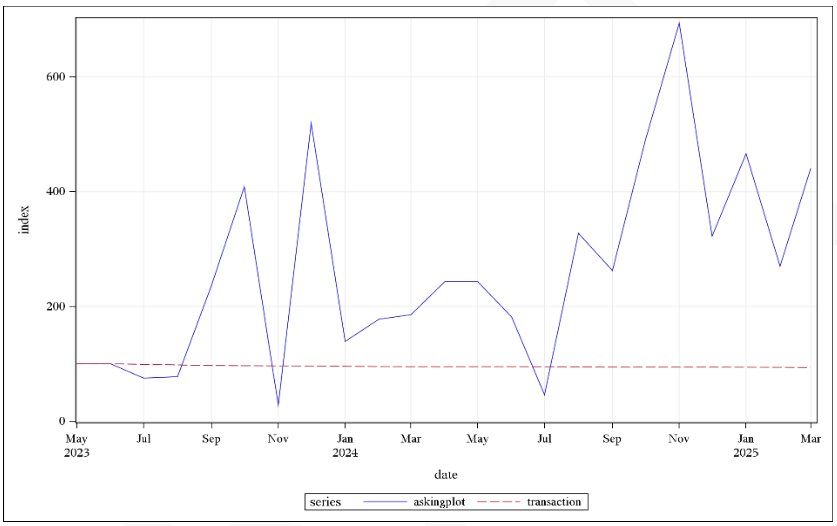


Figure 2 illustrates how plot indices can become extremely volatile when monthly samples are small and key plot attributes are not available or not used in modelling (see also the annex, where the cases of Denmark and Malta are also evident of extremely volatile series). In Finland, the experimental plot index swings sharply over the period, which the feasibility study associates with low plot volumes and the absence of plot size, a central determinant of land value.

Overall, the results suggest that compiling experimental monthly indices for dwellings from online listings is feasible when data collection is stable and the selected portal provides a sufficiently large and broadly representative basket.

By contrast, indices for residential building plots are less robust. Monthly volumes are often thin, plots are highly heterogeneous, and listings frequently lack key quality drivers. In this setting, a hedonic time-dummy model can misattribute changes in composition to changes in prices. If important plot attributes that materially affect value, such as zoning constraints or buildability, are unobserved and their distribution changes over time, the estimated time effects will partly capture these shifts. The resulting index can be biased and may display spurious jumps or declines that reflect changes in the mix of plots advertised rather than genuine market movements. The conclusions for plots should therefore be interpreted with caution.

These findings suggest a targeted improvement agenda for indices on plots of land, but also a clear prioritisation for experimental outputs. Priority actions include improving the availability and treatment of plot area, including imputation where partially missing, and enriching listings with proxy variables for zoning and buildability where feasible. Further steps are to estimate indices within more homogeneous plot strata, for example by size bands and location, and to increase effective sample sizes by pooling months or combining sources in countries where a single portal has limited coverage.

Even with these enhancements, plot indices would still require more extensive development and validation than dwelling indices, because portals typically provide fewer and less standardised characteristics for plots than for dwellings. As a result, they should not be among the first candidates for dissemination as experimental statistics.

2.4. Temporal disaggregation of quarterly HPIs

The previous section examined whether web-scraped listings can support the direct production of quality-adjusted monthly asking-price indices that track the movements of transaction-based HPIs. Building on this, a related question is whether those monthly asking-price indices can be used as an indicator series to derive a benchmark-consistent monthly profile of the official quarterly, transaction-based HPI.

The motivation is primarily improved short-term monitoring and earlier detection of turning points, either within the historical benchmarked window or, where the indicator is timely, as an experimental nowcast subject to revision when new quarterly benchmarks become available.

Temporal disaggregation provides a pragmatic route to a monthly series that remains fully aligned with the official quarterly benchmark. In practical terms, the aim is to assess whether monthly asking prices are sufficiently informative to shape a plausible within-quarter trajectory for the official series, rather than filling monthly gaps through purely mechanical interpolation.

In the commissioned feasibility study (KR&A and MMIT, 2025), temporal disaggregation was implemented using a regression-based approach, treating the quarterly transaction-based HPI as the benchmark series and the monthly asking-price index as the indicator. The Chow and Lin (1971) procedure, which is widely used in official statistics to distribute low-frequency series using related higher-frequency information, was the method applied to the series ([11]).

The study indicates that disaggregation performance depends critically on the strength and stability of the relationship between the asking-price indicator and the transaction-based benchmark. This was most evident in Czechia, Denmark and Spain, where the hedonic asking-price index tracked the official quarterly HPI closely. In these countries, the Chow–Lin disaggregation produced a monthly profile that preserved the official quarterly trajectory while incorporating plausible intra-quarter movements signalled by listings.

Conversely, Cyprus and Finland showed that when alignment is weak or unstable, most of the month-to-month volatility in the asking-price indicator is filtered out by the Chow–Lin procedure, so the disaggregated monthly series ends up looking little different from a simple interpolation of the quarterly HPI. Similar situations arose in Slovakia and Lithuania, and also in Austria and Malta, where the indicator–benchmark relationship was weak, drifting, or low-correlation, so temporal disaggregation transmitted little credible high-frequency movement and could not compensate for deficiencies in the asking-price series.

Overall, the evidence supports temporal disaggregation as a practical route to benchmarked monthly HPIs when web-scraped asking-price indices are demonstrably reliable and broadly representative of market dynamics captured by transactions. Where these conditions hold, the approach can enrich short-term monitoring without undermining the quarterly official benchmark.

([11]) A practical implication is that the method produces a benchmark-consistent monthly path only up to the latest available quarterly HPI: if the monthly asking-price indicator extends beyond the last published quarter, Chow–Lin can be used to generate an experimental forward extension (a nowcast), but these months are not yet constrained to match the as-yet-unknown quarterly benchmark and will be revised once the next quarterly HPI is released.

Where they do not, the resulting monthly profile adds little beyond interpolation and may convey a misleading sense of precision. This motivates focusing subsequent methodological development and empirical illustration on countries and segments where portal coverage and asking-price index performance are among the strongest.

3. Literature review

This section briefly reviews recent literature on the use of asking prices to enhance existing transaction-based measures or to develop alternative indicators. The review focuses on recent evidence, published from 2015 onwards, and considers two main use cases. The first is deriving early signals for the evolution of transaction-based house price indices. The second is constructing price levels indicators.

3.1. Early signals from asking prices literature

A central motivation for using web-scraped housing listings in official statistics is to reduce the information lag inherent in transaction-based HPIs compiled from administrative sources and released with delay.

Asking prices are observed at the start of the buying-selling process and are therefore available earlier than recorded sale prices, while also tending to adjust more rapidly to changing market conditions as sellers revise list prices. This timing advantage has motivated a growing empirical literature examining whether quality-adjusted growth rates from listings co-move with official indices and can provide earlier signals of turning points.

Using rich micro data for Japan, Shimizu et al. (2016) compare price distributions across successive stages of the selling process and show systematic differences between initial list prices, prices at the point of offer, contract prices and registry prices. This evidence underlines why a stable lead relationship cannot be taken for granted, since the size of the list to sale discount can vary over time and can be dependent on the state of the market (e.g. boom vs downturn). Similar concerns are raised in work that compares indices built from listings and transactions, where divergence is often linked to changing bargaining power, shifts in liquidity and compositional differences between what is listed and what is ultimately sold.

Against this backdrop, several studies find that listing based indicators can nevertheless provide useful early information when the data are cleaned and the index construction controls for quality and composition. Lyons (2019) shows for Ireland that a listings based hedonic index captures broad price movements, while also documenting circumstances in which listing indicators can drift when the discount between asking and achieved prices changes.

More recent work has expanded from monthly indicators to higher frequency monitoring and nowcasting. Bricongne et al. (2023) use daily web scraped listings for the United Kingdom to produce high frequency measures of market activity and asking price dynamics, illustrating their value for real time monitoring during the Covid 19 shock.

Pfeifer and Steurer (2022) similarly construct early indicators from daily list price data and show how high frequency information can track rapid changes before conventional releases. Trojanek et al. (2025) move from monitoring to forecasting and demonstrate that quality adjusted monthly list price indices can nowcast quarterly transaction-based indices using MIDAS regressions, with the strongest gains when early within quarter information is used.

A consistent lesson across these studies is that timeliness gains come with greater exposure to noise and to compositional change, which increases the premium on robust methods. More generally, the literature suggests that listing-based early indicators are most informative when they combine strong data processing, clear timing definitions, explicit quality adjustment, and evaluation metrics that focus on whether the indicator improves the early reading of quarter-on-quarter movements rather than only long run co-movement.

3.2. Price levels as housing market indicators

Price indices measure pure price change over time, but they do not answer level questions that matter for policy, such as how expensive housing is in euro terms, how affordability differs across places, or how price levels compare across regions and countries.

For these purposes, indicators expressed in national currency are widely used alongside indices because they are easy to interpret, can be related directly to income measures, and support spatial comparisons that an index alone cannot provide.

For instance, Pionnier and Schuffels (2021) highlight that consistent regional price level statistics are fundamental for assessing affordability and potential barriers to labour mobility, precisely because indices do not provide a level anchor.

Price level indicators are not automatically comparable over time or across space, because they are subject to compositional and quality mix effects in much the same way as simple unadjusted price measures. If the mix of dwellings changes, a raw average price per square metre can move even when constant quality prices are stable.

An implication is that a meaningful price level series should be tied to a clearly defined target concept. One option is a constant quality level defined as the predicted price per square metre for a fixed reference dwelling, holding its characteristics constant and using a hedonic model to adjust for quality, so that changes over time reflect price movements rather than shifts in the mix of properties. Another option is a descriptive level statistic intended to represent a typical market outcome, such as the median price per square metre within well-defined strata.

Pionnier and Schuffels (2021) formalise the need to make these concepts explicit and propose a bridge between levels and indices, so that levels are interpretable while remaining consistent with the temporal patterns captured by quality adjusted indices.

Work on cross country comparisons reinforces why levels can complement indices. Bricongne et al. (2019) show that assessments relying on index numbers alone face limitations for international surveillance, and they construct a dataset of house price level estimates in euro per square metre to support comparisons and affordability style metrics such as price to income expressed in levels. Their discussion also highlights that different data sources imply different biases, and that asking price based methods can be informative but require adjustment for the upward bias of list prices relative to transactions.

Statistical practice provides additional lessons for defining and communicating these indicators. Statistics Finland (2025) publishes prices per square metre for the whole country and regional breakdowns as weighted arithmetic averages of square metre prices and, separately, price development through indices, which helps users distinguish level from change concepts. Notaires de France (2026) disseminates median prices per square metre for existing apartments and houses, illustrating the communication value of a robust level statistic that is easy to understand.

For the construction of price level statistics based on web scraped listings, these strands of evidence suggest that it is useful to define a clear target concept, for example by estimating a constant quality level for a reference dwelling using hedonic modelling, while recognising that asking price levels may be systematically higher than transaction based levels and may therefore require benchmarking against transaction data, an explicit adjustment for the asking price premium or a clear indication that what is being given are asking, not transaction, prices.

4. Data and methodology

The data used in this section is the same e web-scraped listings datasets generated through the commissioned feasibility study addressed in this paper.

The analysis is used to assess two use cases, deriving earlier signals of transaction-based indices and compiling price levels. The methodology for both of these use cases is described below.

4.1. Early-signal assessment method

Building on the evidence reviewed in Section 3.1, this use case evaluates whether quality-adjusted asking-price indices derived from web-scraped listings can provide early information on quarter-on-quarter movements in official transaction-based HPIs.

Monthly asking-price indices are available for 13 EU countries over a common window spanning six complete quarters (2023Q4 to 2025Q1). In addition to country results, an experimental aggregate is computed for the 13 countries, which together account for around 27% of the weight of the EU HPI aggregate. This provides an initial indication of whether the approach may add value for European aggregates even when country-level performance differs.

To ensure comparability with the official HPI, the monthly asking-price index levels are converted to quarterly levels by averaging the three monthly observations within each quarter.

Let $I_{q,1}$, $I_{q,2}$, $I_{q,3}$, denote the monthly indices in quarter q (as taken from equation (2) above), then the quarterly level is defined as

$$I_q = \frac{I_{q,1}+I_{q,2}+I_{q,3}}{3} \qquad (3)$$

Quarter-on-quarter growth is then computed from successive quarterly indices as

$$g_q = 100 \times \left(\frac{I_q}{I_{q-1}} - 1\right) \qquad (4)$$

For each country and quarter, g_q is compared with the official quarter-on-quarter HPI growth rate. Performance is summarised by the mean error (bias, in percentage points) and the mean absolute error (accuracy, in percentage points). Directional performance is measured as the share of quarters for which asking-price and official HPI growth have the same sign.

Co-movement is assessed using Pearson correlations, computed both for the same quarter and with asking-price growth shifted one quarter earlier relative to the official HPI. The latter provides a direct test of whether asking-price growth contains usable information about the next official HPI release.

This evaluation is aligned with ongoing work in Eurostat on flash estimates of euro area and EU HPIs. In current transmission arrangements, national HPI data are delivered to Eurostat within 85 days after the quarter ends, while EU aggregates are typically disseminated later in the production cycle. When available, asking-price indicators can be produced shortly after the end of the month and may therefore support earlier estimates or nowcasts of quarter-on-quarter movements.

The flash HPI work plan foresees quarterly exchanges of test results starting in 2025Q1 and structured assessments once 8, 12 and 16 quarters of test results are available. Guidance documents for national and for EU and euro area flash HPIs are targeted for completion by December 2026. The empirical strategy adopted here provides a consistent set of evaluation metrics that can complement such assessments.

4.2. Price level indicators

Alongside price indices, experimental price level indicators provide an interpretable anchor for spatial comparisons and affordability analysis. This section elaborates on constant-quality asking-price levels from the same cleaned web-scraped listings used elsewhere, but with the objective of estimating levels rather than growth rates.

To limit compositional effects, price levels are defined for a fixed reference dwelling and adjusted using a hedonic framework. We follow a reference-dwelling approach (pricing a dwelling with fixed characteristics) rather than a market-basket approach, because the latter would require stable, comparable weights by dwelling type and location that are not consistently available across countries and small areas for the purposes of this paper.

The target statistic is the monthly price level of a 100 m² reference apartment at NUTS 3 level. Spatial price differences are captured through region fixed effects, while monthly movements are captured through time dummies:

$$Ln(A_{it}) = \beta_0 + \sum_{k=1}^{K} \beta_k \, x_{k,it} + \sum_{\tau=2}^{T} \delta_\tau D_{\tau,it} + \sum_{r=2}^{M} \gamma_r R_{r,i} + \varepsilon_{it} \quad (5)$$

Where:

$R_{r,i}$ = 1 if dwelling *i* is located in NUTS 3 ([12]) region *r*.

Because the model is estimated in logs, fitted values are retransformed to euros using the smearing estimator to avoid bias (Duan, 1983). Let $\hat{\varepsilon}_{it}$ be the regression residuals and define the smearing factor $\hat{S} = \frac{1}{n}\sum_{i=1}^{N} exp(\hat{\varepsilon}_{it})$. The quality-adjusted asking price level for the 100 m² reference apartment in region *r*, month *t* is:

$$\hat{A}_{r,t} = exp\left(\hat{\beta}_0 + \sum_{k=1}^{K} \hat{\beta}_k \, x_k^{ref} + \delta_t + \hat{\gamma}_r\right) \times \hat{S} \quad (6)$$

Where:

x_k^{ref} = denotes the characteristics of the reference apartment

and the corresponding price per square metre could be computed as $\hat{p}_{r,t}^{m^2} = \hat{A}_{r,t} / 100$.

The annual NUTS 3 price level for the 100 m² reference apartment can be used to derive an additional affordability indicator. Using Eurostat national accounts adjusted gross disposable income as an income concept and dividing by population to obtain income per capita, an annual level-based price-to-income indicator for country *c* and region *r* in 2024 is:

([12]) The Nomenclature of Territorial Units for Statistics, or NUTS, it’s the European Union’s standard regional classification used to produce and compare regional statistics, maintained by Eurostat.

$$Af_{c,r} = \frac{\bar{A}_{c,r,2024}}{ADI_{c,2024}/_{POP_{2024}}} \qquad (7)$$

Where:

$\bar{A}_{c,r,2024}$ = 2024 annual average asking-price level of the reference apartment

$ADI_{c,2024}$ = gross adjusted disposable income in 20204

POP_{2024}= population in 2024.

The indicator given in (7) is complementary to index-based affordability measures as it anchors affordability in *levels* and is particularly informative for within-country spatial comparisons across NUTS 3 regions ([13]).

5. Use cases empirical results

This section reports the empirical results for the two use cases described in the previous section. It first assesses whether web-scraped asking-price indices can provide early signals of quarter-on-quarter movements in official transaction-based HPIs and presents experimental price level indicators and an illustrative level-based price-to-income measure.

5.1. Early-signal assessment results

Table 3 summarises the performance of the asking-price indicators over six quarters (2023Q4 to 2025Q1), both for individual countries and for the experimental 13-country aggregate (around 27% of the EU HPI weight).

Overall, the results do not provide clear evidence that quarter-on-quarter asking-price growth rates move closely with transaction-based HPIs in a consistent way across countries. Same-quarter correlations vary markedly, with the strongest positive values observed for Romania

([13]) Please note that Eurostat already provides information on house price-to-income ratios. These are index-based measures, with a house price index in the numerator and an income-per-capita index in the denominator. The level-based indicator proposed here is complementary: by anchoring the comparison in price and income levels, it is better suited to spatial comparisons and to communicating the magnitude of affordability gaps, while the Eurostat index-based ratio remains more appropriate for monitoring short-run dynamics and long-run trends.

(0.70), Spain (0.63), Lithuania (0.59) and Czechia (0.59), while several countries show weak or negative correlations (for example Belgium -0.45 and Malta -0.87).

There is also little evidence that asking-price growth rates lead official HPI growth by one quarter. Most one-quarter lead correlations are close to zero or negative. Austria (0.76) and Romania (0.69) are the largest positive lead correlations in this sample, but the overall pattern suggests that any lead relationship is country- and data-source-specific rather than general.

Table 3: Asking-price indices as early signals of quarter-on-quarter HPI changes, by country

	Mean error (p.p.)	Mean absolute error (p.p.)	Directional performance (%)	Correlation (-1, 1)	
				Same quarter	One quarter lead
Austria	0.16	1.83	67	0.33	0.76
Belgium	-0.85	1.11	50	-0.45	-0.85
Cyprus	-1.84	4.26	50	0.19	0.23
Czechia	0.35	1.18	83	0.59	-0.25
Denmark	-0.67	2.25	67	-0.19	0.49
Finland	0.37	1.86	50	0.61	-0.84
Hungary	-1.32	2.32	100	0.6	-0.45
Lithuania	-0.78	1.64	67	0.59	0.04
Luxembourg	1.65	3.87	50	0.00	-0.68
Malta	0.32	1.28	83	-0.87	0.04
Romania	0.62	1.10	100	0.70	0.69
Slovakia	-0.56	3.17	67	0.50	0.00
Spain	-0.23	1.83	83	0.63	0.06
Aggregate	**-0.19**	**0.74**	**100**	**0.67**	**0.10**

Source: KR&A and MMIT (2025). Own calculations.

Country-level accuracy is modest. Mean absolute errors range from 1.10 percentage points in Romania to 4.26 percentage points in Cyprus, with most countries falling between 1 and 3 percentage points. Directional performance ranges from 50% to 100%, but these rates should be interpreted cautiously given that the evaluation window covers only six quarters.

A key result, however, is that performance improves when countries are pooled into an aggregate. The 13-country aggregate shows a mean absolute error of 0.74 percentage points and a same-quarter correlation of 0.67, with perfect directional agreement over the six quarters. This suggests that aggregation can attenuate idiosyncratic country-level noise and deliver a more stable early reading of quarter-on-quarter movements.

These findings are relevant for ongoing work on flash HPIs. Even if asking-price indicators are not sufficiently precise to serve as a stand-alone proxy for every country, they may still contribute as an auxiliary input to earlier estimates of EU aggregates. They may also be informative as a contingency proxy if a country cannot deliver transaction-based HPI data in time for the compilation of European aggregates (transmission is due within 85 days after the quarter ends).

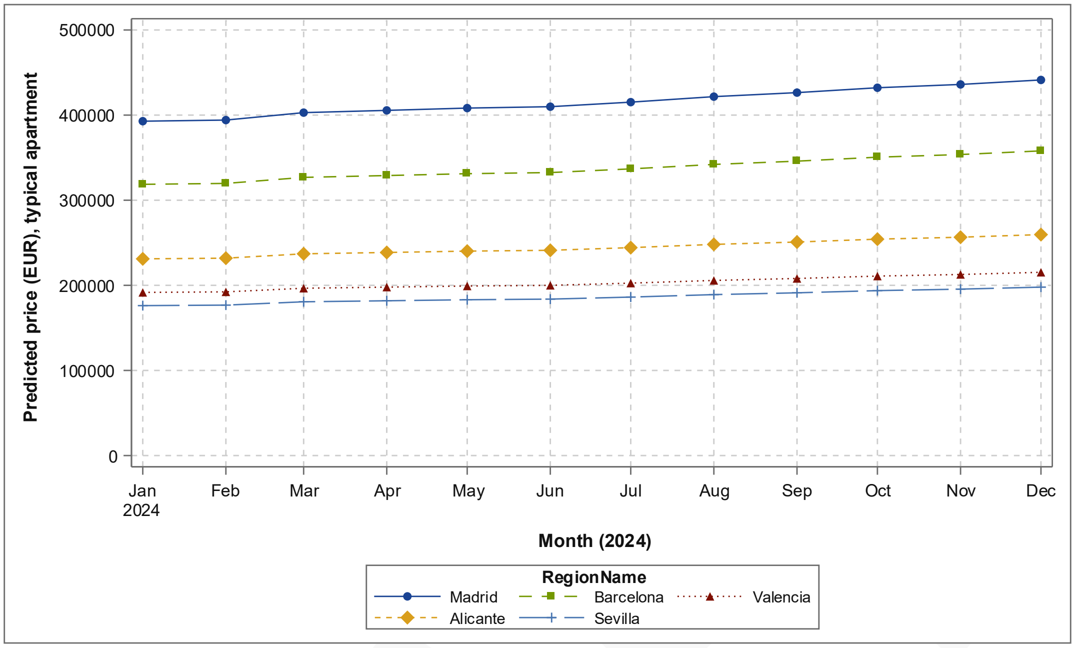
The results should be treated as indicative because they are based on only six quarters. Future work could also test alternative modelling choices that may be more robust to omitted variable bias (e.g. double imputation hedonic approach), for example by enriching the set of characteristics, allowing more flexible interactions, or combining listings with external information where feasible.

5.2. Results on price level experimental series

This section reports the empirical results of the price-level indicators defined in Section 4.2, focusing on three datasets covering Denmark, Romania, and Spain.

In line with the framework set out in Section 4.2, results are presented for a 100 m² reference apartment in five of the most populous NUTS 3 regions within each country. Basic summary statistics and model diagnostics for the constant-quality asking-price estimates for this dwelling type are reported in the annex. Figure 3 summarises the estimated monthly average price levels in Spain in 2024.

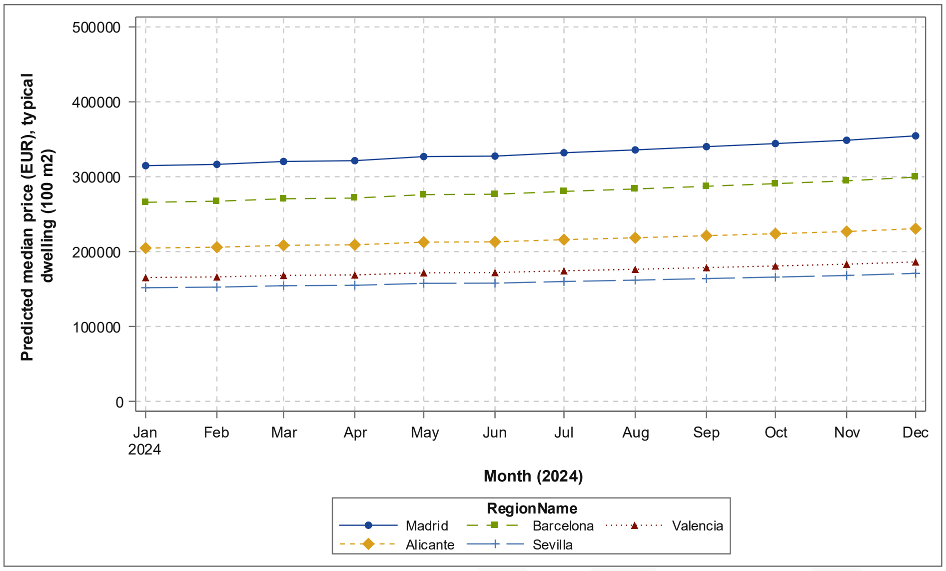
Figure 3: Monthly asking-price levels for an apartment in selected regions, 2024 (€)



The results show a clear and plausible ranking across regions. In particular, Madrid and Barcelona exhibit consistently higher price levels throughout the year, while Seville is systematically lower. This stable ordering is encouraging, as it suggests that the approach yields interpretable spatial differences in levels once compositional effects are controlled for.

To test the robustness of the results (in terms of both ordering and levels) to outlying listings, the same statistic was also computed using a median regression estimator instead of the ordinary least squares (OLS) estimator. The next figure presents the results.

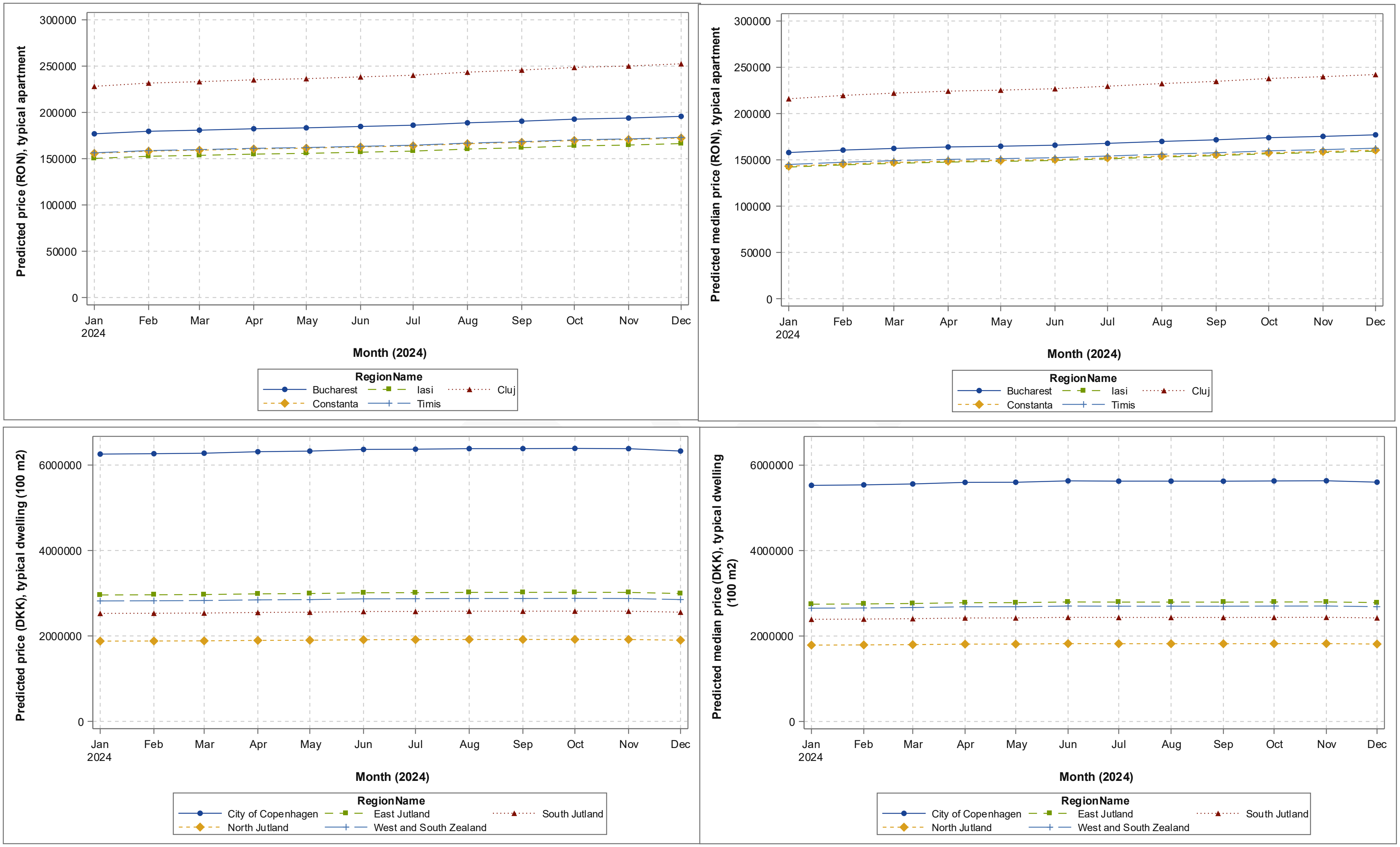
Figure 4: Monthly asking-price levels for an apartment in selected regions, 2024 (€) (median regression)



As can be seen, the ordering of regions does not change when the median regression estimator is used. However, the estimator has a sizable impact on levels, particularly for Madrid, for which the estimated median monthly price level is around €100,00**0** lower than the mean-based estimate shown in the previous figure.

At this stage, it is useful to assess whether the same pattern holds for the other datasets. The next figure presents the results.

Figure 5: Mean (left panel) and median (right panel) asking prices in Romania (top) and Denmark (bottom), 2024 (€)



For the other two cases, the estimates appear more robust to the presence of outlying observations. While the most expensive regions still seem to be influenced the most, the results for these two countries are noticeably less sensitive than those obtained from the Spanish portal data. Importantly, the regional ordering is unchanged when the median estimator is applied.

The Danish results are particularly robust, with the Copenhagen area (as expected) recording the highest mean and median asking-price levels, followed by the region including Aarhus (the second biggest city in Denmark). In Romania, the most expensive region is not the capital (Bucharest) but Cluj, which is consistent with external evidence reported elsewhere (Deloitte, n.d.).

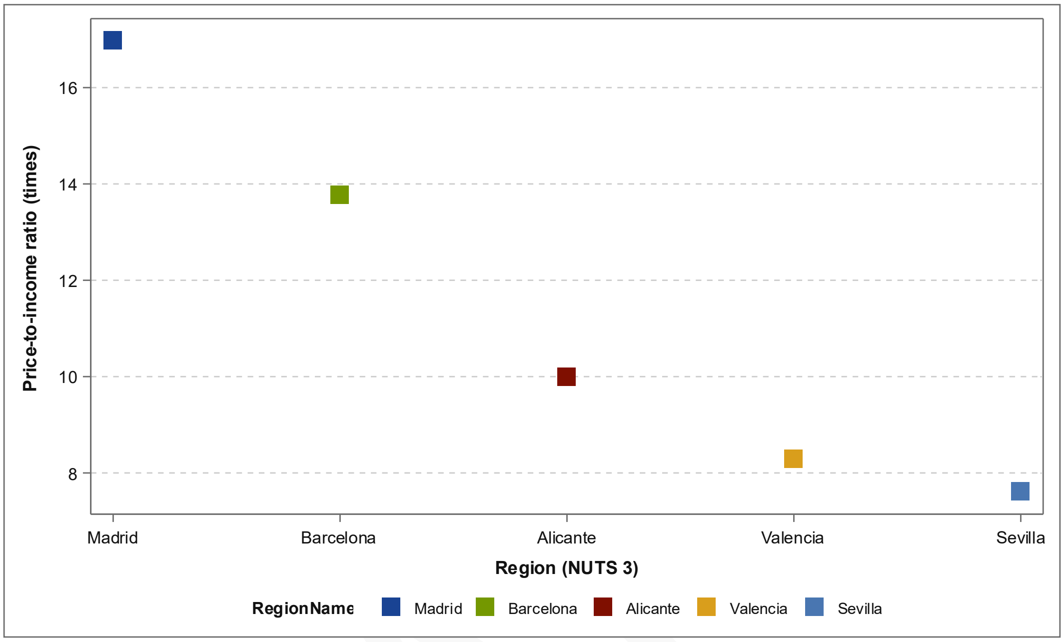
The greater sensitivity to outliers in the Spanish data may reflect genuine market features (e.g. a highly more demanded market), but it also raises the possibility of platform-related effects. For example, the portal may differ in how quickly listings are updated or removed, which could increase the prevalence of stale or atypical observations and, in turn, affect mean-based estimates.

Finally, note that the price levels are estimated using all available observations, but listings with missing key characteristics are excluded from the regression estimation (i.e. the effective sample varies depending on item non-response). This is an important practical consideration for operational use, as systematic missingness in particular attributes may affect regional comparability and precision.

As described earlier, a second application of the estimated price levels is the construction of a level-based price-to-income indicator. This is obtained by comparing the annual average of the monthly price levels for the 100 m^2 reference apartment with national per-capita income in Denmark, Romania, and Spain in 2024.

The resulting ratio provides an intuitive measure of how many annual per-capita incomes would be required to purchase the reference dwelling in each region. The next figure shows the results for Spain; results for the other two datasets are reported in the annex

Figure 6: Implied price-to-income ratios for purchasing a 100 m² reference apartment, selected regions, 2024 (based on asking-price levels)



As the results across the three datasets illustrate, even though the measure is based on asking prices, it provides a clear and communicable picture of relative regional affordability differences. The same approach could also be applied to other dwelling types (e.g. houses) to support policy-oriented analysis of spatial affordability pressures.

The indicator should be interpreted with caution because asking-price-based levels may be upward biased relative to transaction prices, and any asking-price premium may vary across regions. Further methodological work is therefore needed to assess sensitivity to outliers, missing characteristics, and model-specification issues.
This highlights the importance of communicating the indicator explicitly as an asking-price measure and, where feasible, benchmarking it against transaction data.

A further avenue for improvement is to adopt more systematic approaches to handling missing characteristics (e.g. imputation methods) to reduce potential selection effects from complete-case estimation.

6. Conclusions and the way forward

This paper assesses whether web-scraped real estate listings from online portals can support the development of experimental housing market statistics for EU countries. It does so in two ways. First, it revisits the main findings of a recent feasibility study commissioned by Eurostat (KR&A and MMIT, 2025). Second, it extends those findings by developing and testing two use cases based on asking prices, namely monthly asking-price indices as earlier signals of official transaction-based HPIs and constant-quality price level indicators.

Regarding the feasibility study, the main conclusion is that producing higher-frequency experimental indicators from web-scraped asking prices is achievable, but only under specific statistical and operational conditions.

Statistically, the underlying portal data must provide sufficient and stable volumes and include a minimum set of key variables (notably asking price, floor area, property type and a usable location proxy). Equally important, the listings need to be broadly representative of the market segment being modelled; otherwise, apparent movements may mainly reflect shifts in the composition of listings rather than genuine price change.

Operationally, sustained access to portal data, resilient collection pipelines and systematic production monitoring are essential. Without these elements, breaks in collection, changes in website structures, or shifts in variable availability can create discontinuities that distort high-frequency measures.

The feasibility results therefore support further development, while also pointing to the need for strong quality assurance and careful source selection. The feasibility evidence also reinforces that residential building plots may be a more demanding segment because volumes are often thin and heterogeneity is high, so plot indices are typically less robust and should not be prioritised for early experimental dissemination without further methodological and data enrichment work.

Regarding the extensions developed in this paper, the empirical assessment of asking-price indices as early signals suggests that performance varies substantially by country and source. Across countries, there is limited evidence of a stable one-quarter lead relationship with official HPI growth.

However, a key result is that performance may improve when countries are pooled into an aggregate because aggregation attenuates idiosyncratic country-level noise and can deliver a more stable early reading of quarter-on-quarter movements. In this sense, asking-price indicators appear most promising as an auxiliary input for European aggregates and as a practical empirical contribution to ongoing work on experimental flash estimates of the HPI.

The second use case shows that constant-quality price level indicators can be derived from asking prices using a reference-dwelling approach within a hedonic framework. The regional patterns obtained are interpretable and, in the examples assessed, produce plausible and stable spatial rankings, which is particularly relevant for territorial affordability analysis.

At the same time, the results highlight the importance of robustness choices, notably sensitivity to outliers in some portal datasets, and of clear communication because these are asking-price-based levels which may be systematically above transaction-based levels, and any asking-price premium may vary across markets and regions. Where feasible, benchmarking to transaction data or using transparent adjustment strategies should therefore form part of the methodological roadmap before wide use.

Beyond the specific results, the harmonised microdata assembled for the feasibility study are valuable in their own right. They provide a practical basis for further experimentation (including alternative hedonic specifications, improved handling of missing characteristics, and additional segmentation), for trial production exercises, and for developing shared training and guidance material within the European Statistical System on the production of experimental real-estate indicators from web data.

Building on these conclusions, the way forward could take the form of an implementation-oriented project that uses the results described in this paper as a starting point for developing a small number of experimental outputs (notably: benchmark-consistent monthly profiles of quarterly HPIs using suitable indicator series; and constant-quality price level indicators with regional breakdowns).

A first step would be a structured landscaping exercise to identify and assess relevant real estate portals across EU Member States. This would document market coverage, update frequency, data accessibility constraints and business models, and would help determine where multi-portal strategies, negotiated data provision, or a combination of scraping and direct feeds are necessary.

In parallel, representativeness assessment should be treated as a core statistical workstream rather than an afterthought. Practical options include systematic comparisons of listings against housing stock and transaction information by region and dwelling type, and the routine monitoring of shifts in portal coverage, property mix and key-variable missingness, so that experimental series are accompanied by transparent quality indicators.

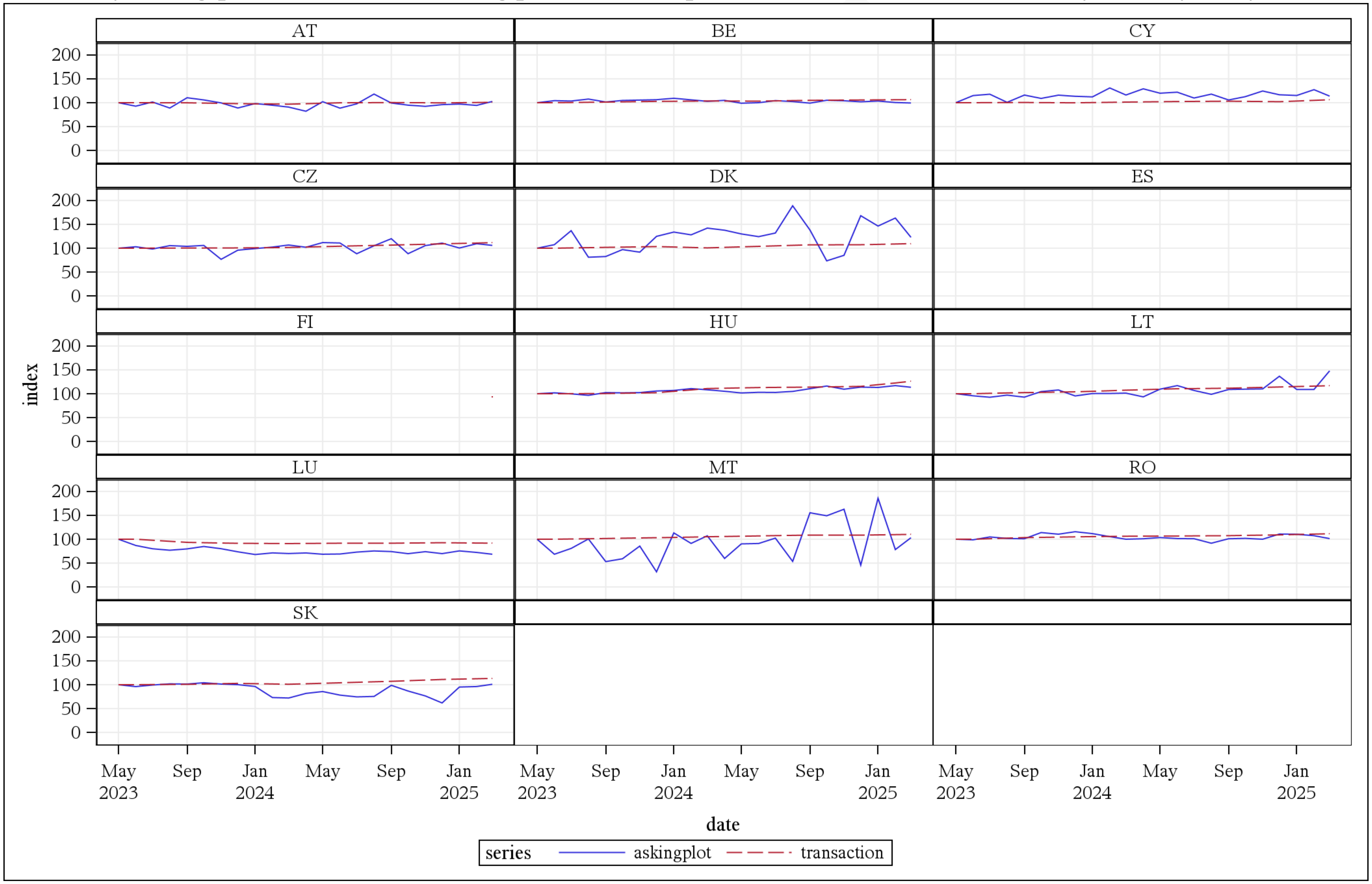
The policy context is favourable for moving from feasibility to implementation. Developments in the European framework for accessing privately held data, together with the prioritisation of price statistics within the European Statistical System, strengthen the case for piloting a limited set of experimental housing indicators with clear governance, quality reporting and revision policies.

In practical terms, this argues for integrating these outputs into Eurostat's statistical innovation framework as experimental statistics that are published with transparent metadata, quality indicators and clear labelling on scope and limitations, while enabling iterative methodological improvement as coverage expands and longer time series become available.

In this context, Eurostat and Member States are considering a limited number of concrete use cases for collaborative development. These could include experimental monthly indices benchmarked to quarterly HPIs, price levels with regional breakdowns, and, subject to further methodological progress, indicators for residential building plots derived from web portals. The prioritisation exercise is expected to be completed in 2026.

Annexes

A.1: Monthly asking-price indices for building plots and interpolated transaction-based HPIs, by country (May 2023 = 100)



Source: KR&A and MMIT (2025). Notes: No data are shown for plots of land in Spain because the portal did not provide this data. For Finland, the plot is reproduced in the main text.

A.2: Regression outputs

Dependent variable: *Natural logarithm of asking price*

Estimator: OLS

Independent variables (estimated coefficients):

	Int	*LArea*	*DRg2*	*DRg3*	*DRg4*	*DRg5*	*DUr2*	*DUr3*	*Dold2*	*Dold3*	*Dold4*	*Dold5*	*DEnerg AB*	*DEnergy CD*	*DEnergy E*	*DApart*	*DRooms*
DK	*10.362*	*.877*	*-.749*	*-.906*	*-1.203*	*-.797*	*.122*	*.774*	*.038*	*.185*	*.261*	*.494*	*.476*	*.384*	*.184*	*-.081*	*0.016*

	D2	*D3*	*D4*	*D5*	*D6*	*D7*	*D8*	*D9*	*D10*	*D11*	*D12*
DK	*.001*	*.003*	*.009*	*.001*	*.002*	*.018*	*.020*	*.020*	*.021*	*.020*	*.011*

Note: The coefficient on Larea (log area) is the (partial) elasticity of price with respect to floor area.

Dummy variables that are taken out from the model:

- *DRg1:* Copenhagen
- *DUr1:* Degree of urbanisation 3 (less urbanised)
- *Dold1:* Construction year <=1950 (oldest category)
- *DEnergFG:* Energy certificates F and G (worse energy efficiency labels)
- *D1:* If the listing is from January 2024

Basic model stats:

N: *267 232*
Adj R^2: *.550*

Dependent variable: *Natural logarithm of asking price*

Estimator: OLS

Independent variables (estimated coefficients):

	Int	LArea	DRg2	DRg3	DRg4	DRg5	DUr2	DUr3	Dold2	Dold3	Dold4	Dold5	DEnergAB	DEnergyCD	DEnergyE	DApart	Dattic	DGfloor
ES	8.313	.835	-.209	-.718	-.530	-.802	.320	.693	-.239	-.138	-.113	.196	.258	.167	.104	-.026	.293	-.100

	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
ES	.003	.025	.032	.038	.042	.055	.071	.082	.095	.104	.116

Note: The coefficient on Larea (log area) is the (partial) elasticity of price with respect to floor area.

Dummy variables that are taken out from the model:

- *DRg1:* Madrid
- *DUr1:* Degree of urbanisation 3 (less urbanised)
- *Dold1:* Construction year <=1950 (oldest category)
- *DEnergF:* Energy certificates F and G (worse energy efficiency label)
- *D1:* If the listing is from January 2024

Basic model stats:

N: *566 736*
Adj R^2: *.588*

Dependent variable: *Natural logarithm of asking price*

Estimator: OLS

Independent variables (estimated coefficients):

	Int	*LArea*	*DRg2*	*DRg3*	*DRg4*	*DRg5*	*DUr2*	*DUr3*	*Dold2*	*Dold3*	*Dold4*	*Dold5*	*DApart*
RO	*7.040*	*.995*	*-.163*	*.255*	*-.127*	*-.123*	*.132*	*.571*	*-.223*	*-.020*	*.029*	*.075*	*.050*

	D2	*D3*	*D4*	*D5*	*D6*	*D7*	*D8*	*D9*	*D10*	*D11*	*D12*
RO	*.015*	*.022*	*.030*	*.036*	*.044*	*.051*	*.065*	*.074*	*.086*	*.092*	*.101*

Note: The coefficient on Larea (log area) is the (partial) elasticity of price with respect to floor area.

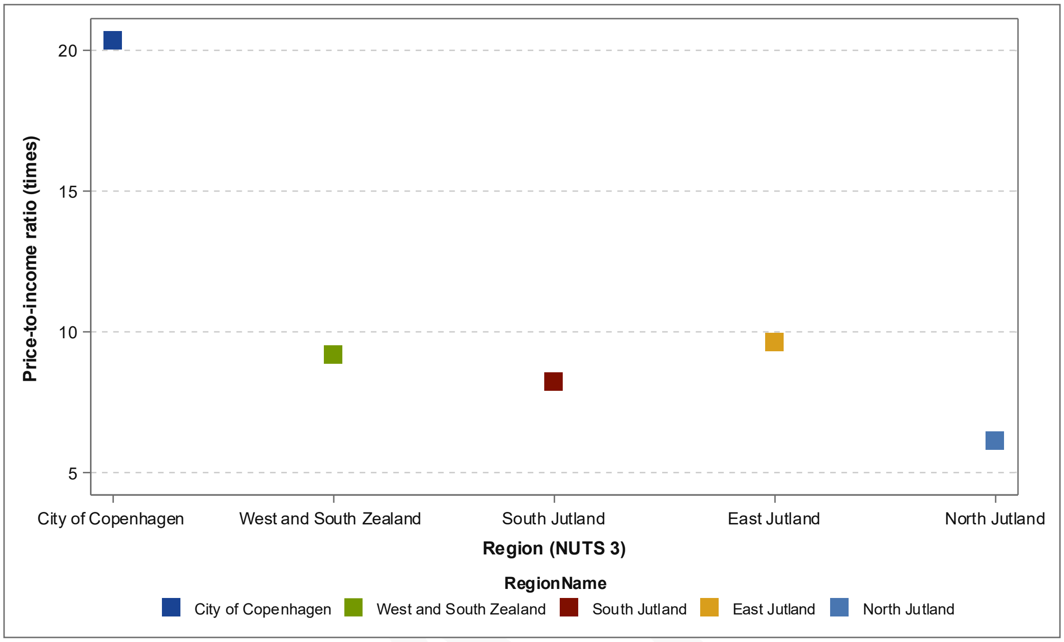
Dummy variables that are taken out from the model:

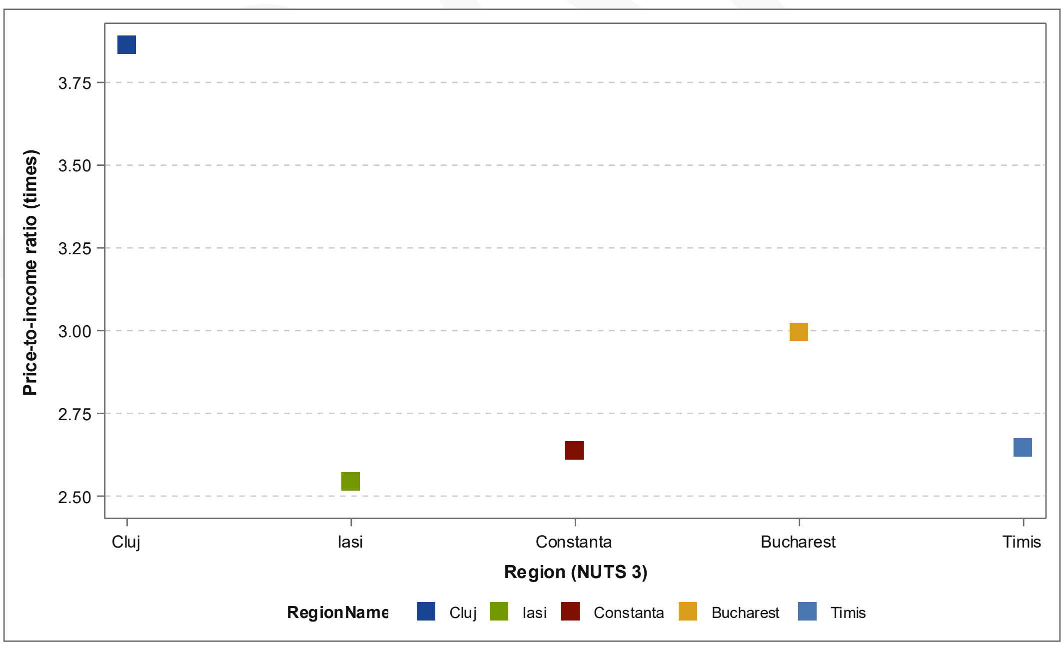
- *DRg1:* Bucharest
- *DUr1:* Degree of urbanisation 3 (less urbanised)
- *Dold1:* Construction year <=1950 (oldest category)
- *D1:* If the listing is from January 2024

Basic model stats:

N: *543 788*
Adj R^2: *.765*

A.3: Price-to-income ratio plots for Romania (bottom) and Denmark (top)





References

- Bricongne, J. C., Turrini, A., and Pontuch, P. (2019). *Assessing house prices: Insights from "Houselev", a dataset of price level estimates* (European Economy Discussion Paper No. 101). European Commission, Directorate General for Economic and Financial Affairs

- Bricongne, J. C., Meunier, B., and Pouget, J. (2023). Web scraping housing prices in real time. The Covid 19 crisis in the UK. *Journal of Housing Economics*, 59, 101906. https://doi.org/10.1016/j.jhe.2022.101906

- Chow, G. C., and Lin, A.-L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics, 53*(4), 372–375. https://doi.org/10.2307/1928739

- Deloitte. (n.d.). *Deloitte study: Romania maintains third position in the top of the European countries with the cheapest new housing. Cluj-Napoca is the most expensive city for both buyers and tenants*. https://www.reff-associates.ro/re/en/about/press-room/deloitte-study-romania-maintains-third-position-top-european-countries-cheapest-new-housing-cluj-napoca-expensive-city-buyers-tenants.html

- Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association, 78*(383), 605–610. https://doi.org/10.1080/01621459.1983.10478017

- European Commission (2025). *Commission staff working document: Understanding the housing crisis. Accompanying the European Affordable Housing Plan* (SWD(2025) 1053/2, Part 1/2). Brussels.

- Eurostat (2017). *Technical manual on owner-occupied housing and house price indices*. Publications Office of the European Union. https://ec.europa.eu/eurostat/documents/7590317/0/Technical-Manual-OOH-HPI-2017/

- Eurostat (2020). *Practical guidelines on web scraping for the HICP*. European Commission. https://ec.europa.eu/eurostat/documents/272892/12032198/Guidelines-web-scraping-HICP-11-2020.pdf

- KR&A and MMIT (2025). *Producing monthly house price indices from web-scraped data* (Final report/Deliverable D7). [Unpublished report prepared for Eurostat].

- Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review, 71*(4), 801. https://doi.org/10.2307/1806207

- Linz, S., Beisteiner, A., Böttcher, M., Dorka, K., Evangelista, R., de Haan, J., Handmann, P., Lohn, A., Mota, D., van Mulligen, P. H., Murray, A., Ribe, M., Taschowsky, P., & Vos, M. (2009). *Handbook on the application of quality adjustment methods in the Harmonised Index of Consumer Prices*: Developed within the European project "CENEX HICP Quality Adjustment" (Statistics and Science, Vol. 13). Statistisches Bundesamt (Federal Statistical Office)

- Lyons, R. C. (2019). Can list prices accurately capture housing price trends. Evidence from Ireland. *Finance Research Letters*, 30, 228–232. https://doi.org/10.1016/j.frl.2018.10.004

- Notaires de France (2026). *Marché de l'immobilier: Indices et cartes de prix au m²*. Conseil supérieur du notariat.

- Pionnier, P. A., and Schuffels, J. (2021). *Estimating regional house price levels: Methodology and results of a pilot project with Spain* (OECD Statistics Working Papers 2021/03). OECD Publishing. doi:10.1787/b9fec1b2-en

- Pfeifer, N., & Steurer, M. (2022). Early real estate indicators during the Covid 19 crisis. Journal of Official Statistics, 38(1), 319–351. https://doi.org/10.2478/jos-2022-0017

- Shimizu, C., Nishimura, K. G., and Watanabe, T. (2016). House prices at different stages of the buying and selling process. *Regional Science and Urban Economics*, 59, 37–53. https://doi.org/10.1016/j.regsciurbeco.2016.04.006

- Statistics Finland (2025). *Prices of dwellings in housing companies (OSF): Statistics page and tables*. Statistics Finland.

- Trojanek, R., Hartigan, L., Pfeifer, N., and Steurer, M. (2025). *Nowcasting transaction based house price indices using web scraped listings and MIDAS regression* (CAMA Working Paper 45/2025). Centre for Applied Macroeconomic Analysis. https://ideas.repec.org/p/een/camaaa/2025-45.html