



Advancing Housing Market Statistics in Europe Using Web-scraped data

Session 6: Data Sources, Big Data and Innovation

Eurostat: Rui Evangelista (*)

*International Conference on Real Estate Statistics
19 and 20 February 2026*

() This presentation reflects the author's views and not necessarily those of Eurostat, the European Commission, or other institutions.*

Outline

- Motivation: gaps in frequency, timeliness and territorial detail
- Data foundation: 13-country portal dataset and what we learned
- Methods: hedonic time-dummy indices and Chow–Lin disaggregation
- Results: where asking-price indices align (and where they do not)
- Use cases:
 - Early signals for quarterly HPIs
 - Price levels & affordability
- Conclusions and way forward

Why more and better housing indicators?

Policy and analytical needs

- Monitoring macroeconomic conditions and financial stability
- Understanding affordability pressures and spatial disparities
- Need for timelier 'market temperature' signals (turning points)

Limitations of many official outputs

- Quarterly frequency and publication lag
- Limited segmentation by dwelling type and small areas
- Few 'level' statistics (price levels) that users can interpret directly

Listings data can complement transactions (not replace them)!

Feasibility-study data at a glance

13 / 13

Portals / countries covered

~30.3M

Usable offers after cleaning

~6.7M

Distinct properties

Collection window and architecture

- Collection began Mar–May 2023; continued through Mar/Apr 2025 (≥ 2 years per portal)
- Mixed approach: 9 operational web scrapers + 4 monthly file deliveries from portal owners
- After filtering foreign/non-residential/rentals/missing price: ~30.3M offers remain
- Repeated observations: listings observed ~4.5 times on average

Lessons learned: access, stability, representativeness

Data access

- Permission requests often fail (no reply / refusal / paywall)
- Need pragmatic sourcing strategy (multi-portal, fallbacks)
- Relationships with market leaders matter for continuity

Stability vs richness

- Direct feeds: stable but typically fewer variables
- Scraping: richer attributes but fragile (site changes, blocking)
- Monitoring is essential (volume checks, break detection)

Representativeness

- Indices work best when portals cover is broad and they show dynamism (go for market leaders)
- List of core variables to be covered (price, size, type, location, new/existing)
- Quality reporting should include coverage and missingness

Takeaway: operational feasibility is as important as modelling choices.

Method: hedonic time-dummy monthly indices

Core idea

- Use asking prices from online listings (monthly frequency)
- Control for changing 'mix' of properties via hedonic regression
- Time-dummy coefficients capture quality-adjusted price change

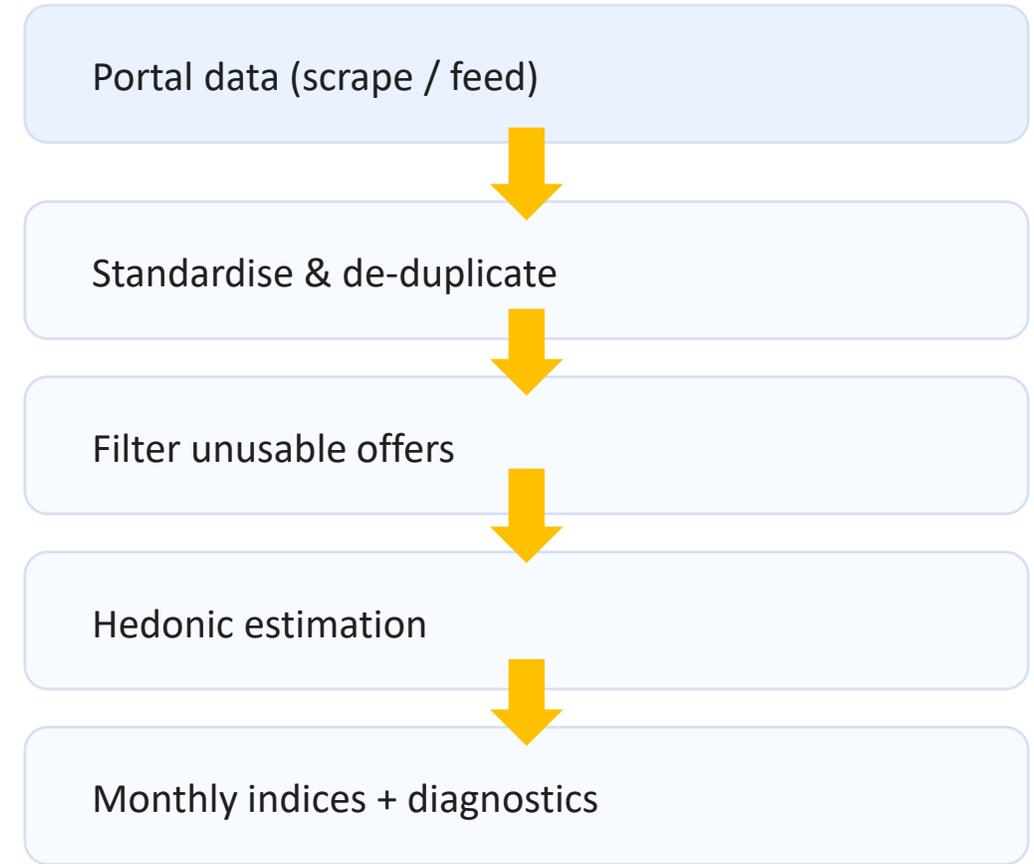
Log-price model (pooled, time dummies)

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

Design choices (for comparability)

- Outlier and inconsistency treatment
- 'Latest offer' per property per month to reduce time-on-market bias
- Core covariates: price, floor area, type, location proxy, new/existing

End-to-end pipeline

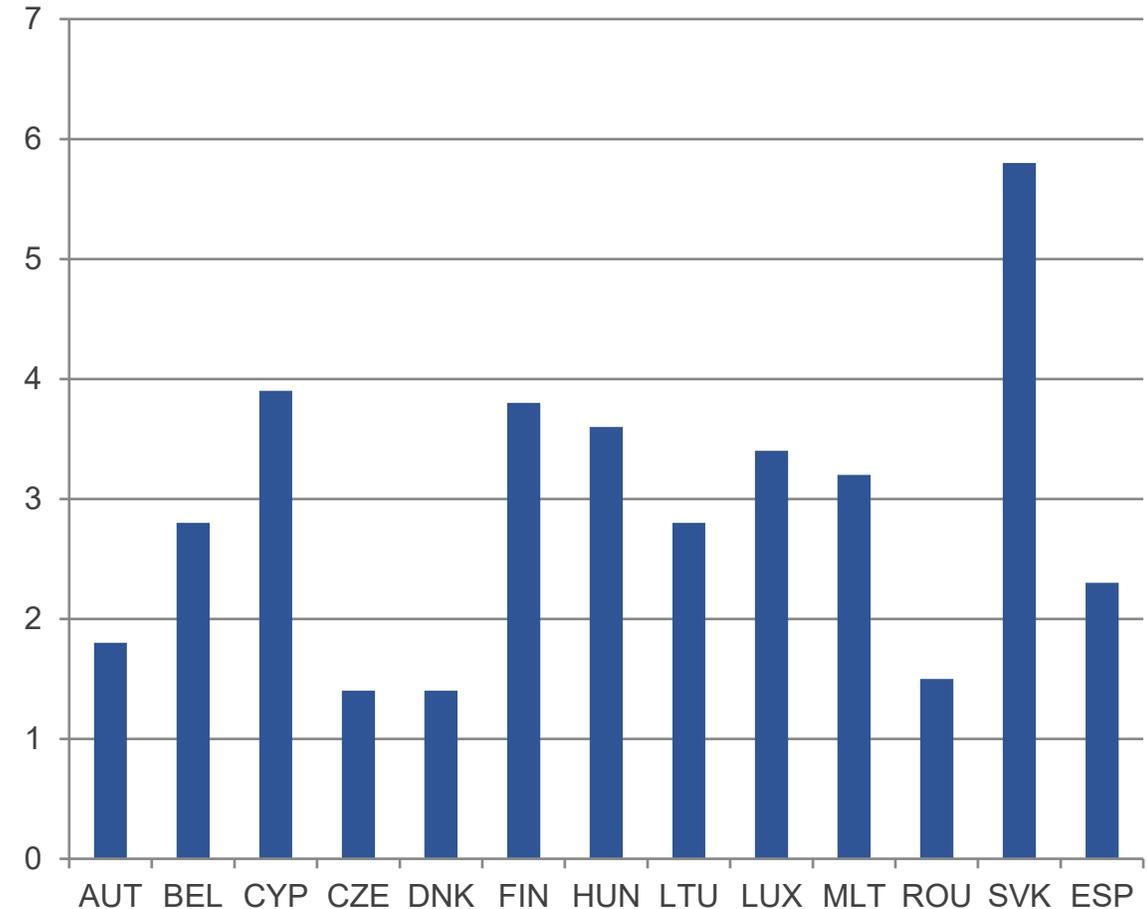


Results: asking-price indices vs transaction HPIs

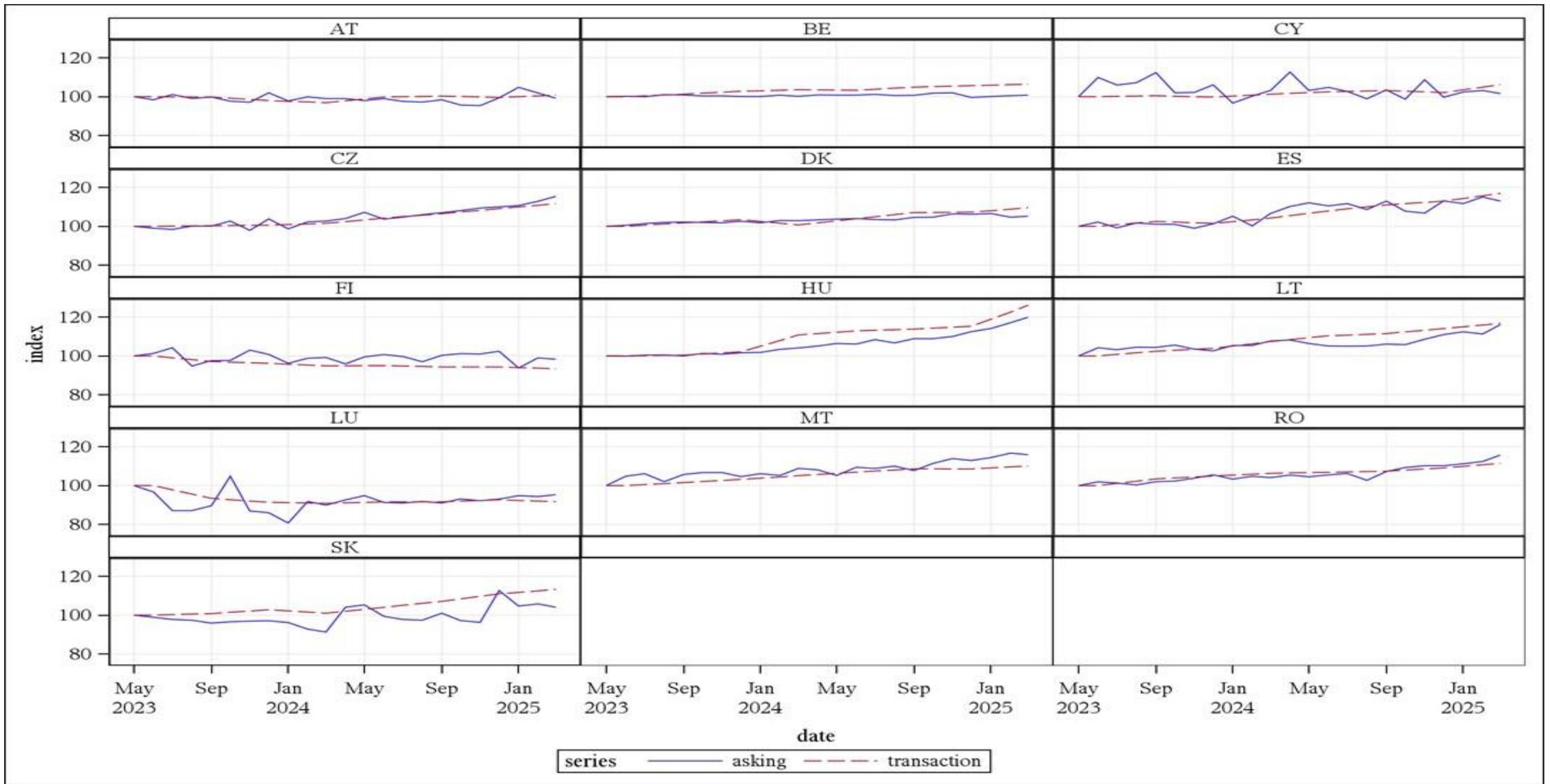
Accuracy varies strongly by portal and market coverage

- Best alignment in Czechia and Denmark (lowest Mean Absolute Error)
- Good performance also in Romania; Spain broadly tracks with larger intra-quarter volatility
- Weak alignment in Slovakia, Cyprus and Finland (higher Mean Absolute Error)
- Building plots: typically volatile due to low volumes and missing key attributes

Mean Absolute Error (in index points)

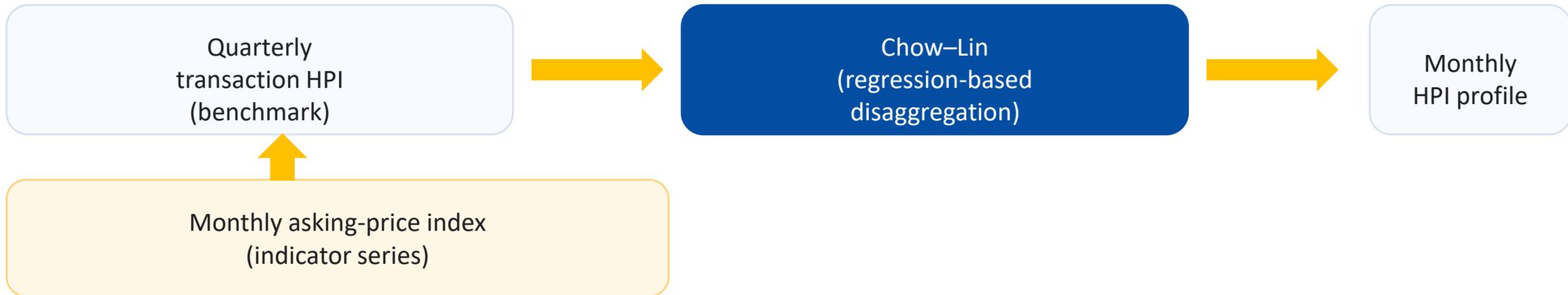


Note: lower MAE indicates closer tracking of official quarterly HPIs.



From quarterly to monthly: Chow–Lin disaggregation

Goal: benchmark-consistent monthly profiles of the official quarterly HPI



When does this add value?

- Strong indicator–benchmark relationship \Rightarrow plausible intra-quarter movements (e.g. Czechia, Denmark, Spain)
- Weak or unstable relationship \Rightarrow disaggregation collapses to near-interpolation (risk of false precision)
- Could be useful for short-term monitoring, nowcast monthly movements, while preserving the official quarterly benchmark

Use case 1 — Early signals for quarterly HPIs

Evaluation setup

- Convert monthly asking-price indices to quarterly levels (average of 3 months)
- Compare quarter-on-quarter growth with official transaction HPI
- Summarise bias (Mean Error), accuracy (Mean Absolute Error), direction-of-change and correlations

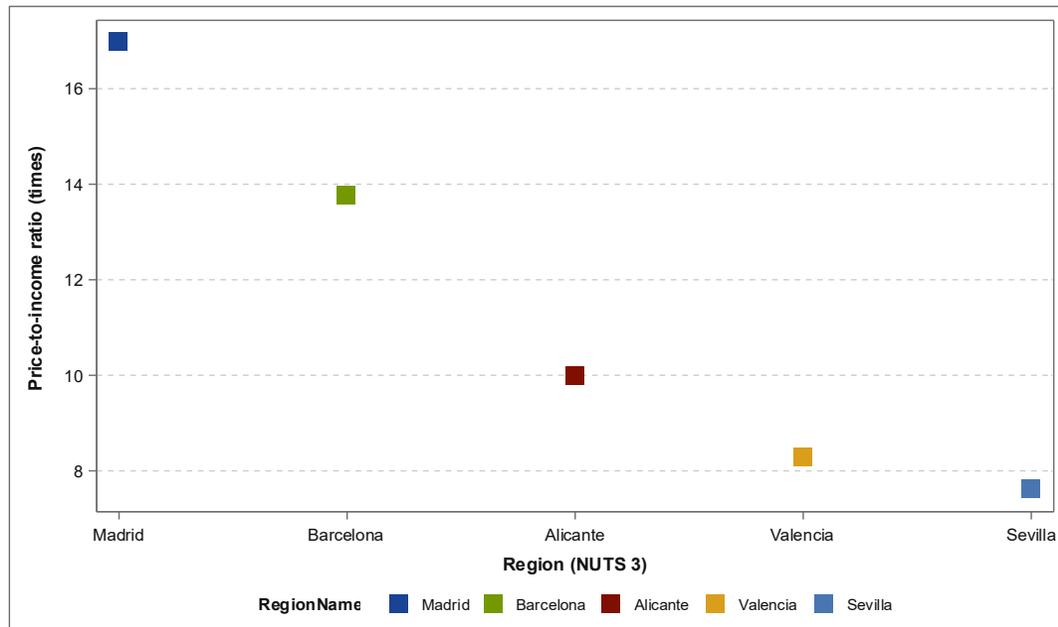
Key findings (6 quarters)

- Limited evidence of a stable one-quarter lead relationship at country level
- Performance improves when pooling: 13-country aggregate MAE \approx 0.74 p.p.
- Aggregate shows high co-movement (corr. \approx 0.67) and correct direction in all evaluated quarters
- Most promising as an auxiliary input for EU/euro area monitoring and flash work

Use case 2 — Price levels and affordability

Why levels?

- Indices tell us ‘change’, not ‘how expensive’
- Levels support spatial comparisons and affordability analysis
- Listings offer rich characteristics for quality adjustment



Reference-dwelling approach

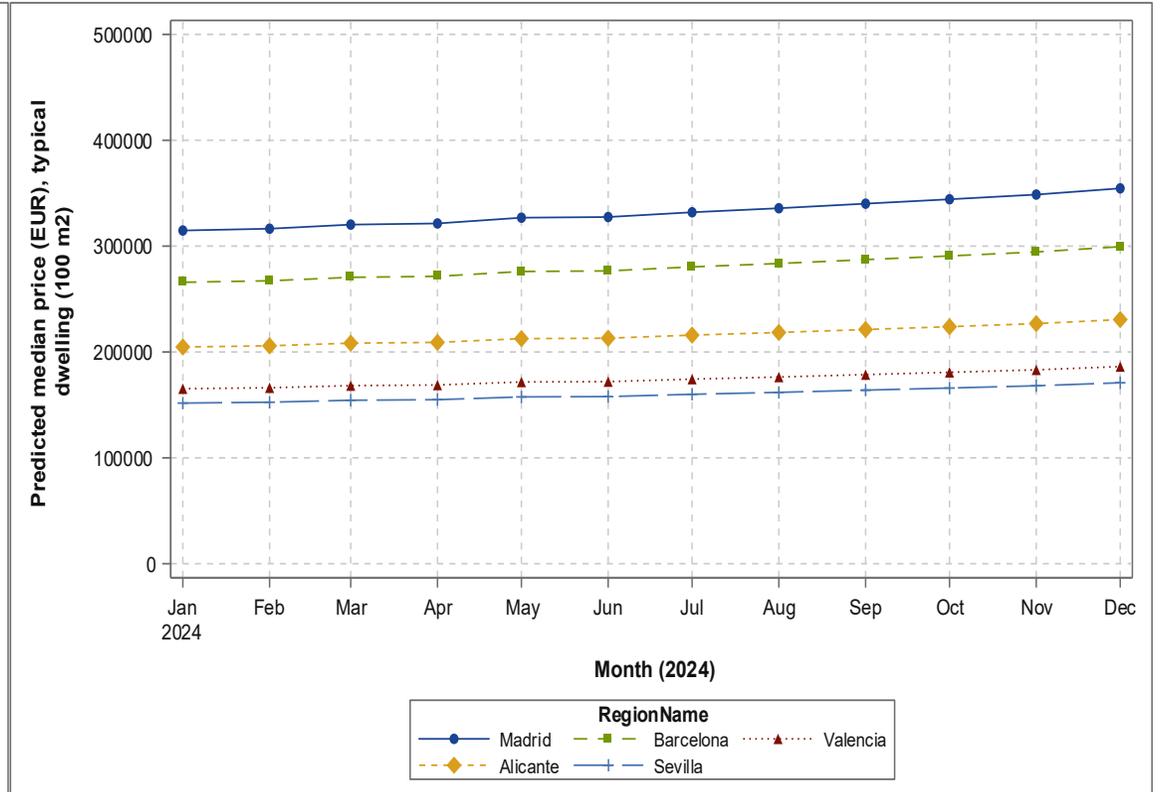
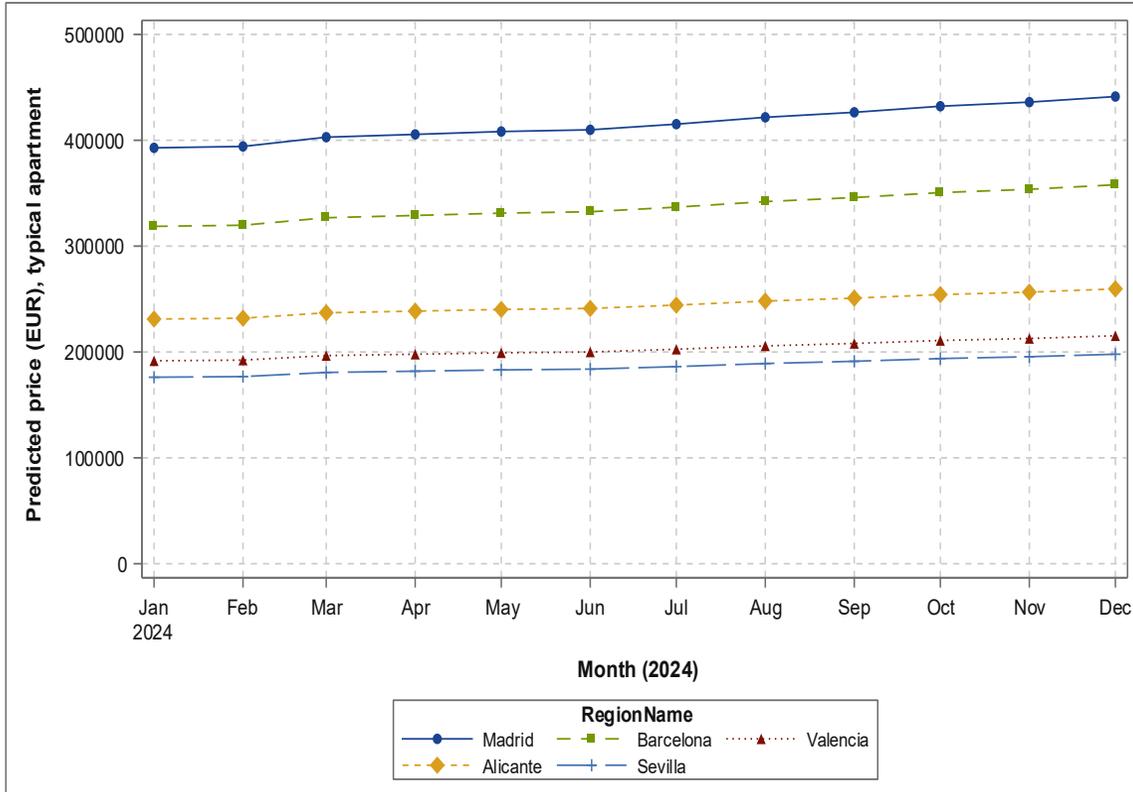
Example output: monthly asking-price level of a 100 m² apartment at NUTS 3

Quality adjustment via region fixed effects + monthly time dummies; retransform using smearing.

Robustness & communication

- Clearly label as asking-price levels (potential premium vs transactions)
- Derive level-based price-to-income ratios using disposable income per capita
- Mean vs median: some portals show outlier sensitivity

Use case 2 — Price levels and affordability



Conclusions and way forward

What the evidence supports

- Monthly experimental indices are feasible when portal coverage is broad and stable
- Core variables (price, size, type, location proxy, new/existing) are essential; extra attributes help but are not decisive
- Benchmark-consistent monthly profiles via disaggregation are useful where indicator quality is proven
- Price-level indicators can add strong value for territorial affordability analysis

Next steps for implementation

- Revisit landscape exercise and assess portals (coverage, update frequency, accessibility, business models)
- Treat representativeness as a core workstream: monitor coverage shifts and missingness
- Pilot a small set of experimental outputs: monthly benchmarked HPI profiles + regional price levels
- Experimental statistics, with transparent metadata and quality indicators

Thank you

Explore official European housing and commercial property price statistics:

[Overview - Housing price statistics – Eurostat](#)

 ESTAT-REAL-ESTATE@ec.europa.eu