A Timely House Price Index for Belgium:

Integrating Listings and Transactions for Improved Forecasting

Peter Reusens & Joren Vandenbergh[*]

This version: 13 February 2026

ABSTRACT

Timely house price indicators are essential for market participants and policymakers, yet official indices suffer from long publication lags because they rely on notarial transaction records. This paper introduces a novel house price index for Belgium covering January 2016 to March 2025, constructed from initial asking prices of newly listed properties on online platforms. To address challenges such as unsold listings, unknown time-until-sale and price negotiation, we adjust for expected probability of sale, time-until-sale and sales-to-listing price ratio using a merged dataset of listings, transactions, and energy performance certificates. Our adjusted listing index anticipates transaction-based indices by several months and forecasts price changes with an average six-month error of about 1 percentage point, outperforming benchmark models. This approach improves timeliness and forecasting accuracy, offering policymakers earlier insights into housing market dynamics and supporting international efforts to leverage big data for real estate statistics.

---
[*]Peter Reusens: National Bank of Belgium - peter.reusens@nbb.be
Joren Vandenbergh: KU Leuven - joren.vandenbergh@kuleuven.be

1.   Introduction

Timely house price data is vital for housing market participants to make informed decisions. It enables the early detection of house price trends, turning points, and bubbles, which is crucial for policymakers due to the major impact of house prices on housing affordability, household wealth, economic activity, and financial stability. However, official house price indices are recorded with a substantial delay, because they are based on the official recording of transactions by the notaries.

This study addresses this data need by constructing a timely house price index for Belgium for the period between January 2016 and March 2025, based on the initial asking prices of newly posted listings, and demonstrating its performance in forecasting a standard transaction price index. The main advantage of using data on asking prices is that they are available several months earlier than data on transactions due to two time lags: (1) the interval between the initial posting of the listing and the initial sales agreement, which has a median duration of about 2 months; (2) the period between the date of the initial sales agreement and the notarial deed, with a median duration of about 3 months. These time differences in data availability, and the fact that initial asking prices are strongly correlated with the final sales prices, motivate the use of house price indices based on listings as a predictor for the transaction price index. An additional timing advantage is that the official house price index has a publication lag and, in Belgium, is only released about 3 months after the end of the quarter.

The main challenge of using initial asking prices as a predictor for the sales price index lies in the fact that a subset of listings do not result in a sale, and that for those that do, the time-until-sale is unknown in advance, and the final selling price may differ from the original asking price. We develop a methodology to address these challenges by taking into account the expected probability of sale, time-until-sale, and sales-to-listing price ratio of each listing in the construction of the listing price index. Our model allows these estimates to depend on dwelling characteristics, and it also takes into account the fact that overpriced listings (i.e. listings with an asking price above the expected value) are less likely to sell, tend to have lower sales-to-listing price ratios, and remain on the market longer. We use a unique dataset that merges sales data for all Belgian dwelling

1

transactions with a dataset of listings from the most important online real estate platforms in Belgium, and with datasets on energy performance certificates (EPC).

We find that by taking into account the sales-to-listing price ratio, time-until-sale, and probability of sale, we can construct a listing index that more closely aligns with the sales price index compared to a listing price index without adjustments, both regarding timing and price developments. This adjusted listing price index also allows us to forecast a transaction-based index up to 8 months ahead, with an average six-month prediction error of 1 percentage point.

As we do not observe transactions for recently posted listings, our method requires extrapolating trends in sales-to-listing price ratio, time-until-sale, and probability of sale. The accuracy of a price forecast using the adjusted listing price index strongly depends on this extrapolation exercise, especially at turning points in the price index. In this way, the method presented in this study shifts the challenge from predicting price changes to predicting evolutions in secondary market indicators.

This article builds on the expanding research finding predictive value of listings-based data for sales-based indices across various countries. Kolbe et al. (2021) show that their online listings price index for single-family houses in Berlin effectively helps predict the next quarter's sales price growth rate. Likewise, Lyons (2019) finds that their price index based on first-time listings significantly Granger cause the sales price index in Ireland. Wang, Li, and Wu (2020) observe that a price index based on online listings closely follows both the long-term trends and short-term dynamics of the sales price index in China. Lastly, Anenberg and Laufer (2017) demonstrate that their listing price index accurately forecasts the U.S. Case-Shiller index several months ahead.

Our main contribution with respect to the above-discussed existing literature is that we take into account the expected probability of sale, time-until-sale, and sales-to-listing price ratio of each listing. The likely reason why other studies overlook these factors is the need to merge sales and listing datasets, which researchers often lack access to (Gluszak and Trojanek 2024). To our knowledge, the only other study to have performed a similar analysis is Anenberg and Laufer (2017). This study shows that these corrections improve the forecasting performance of a listing price index.

We differ from Anenberg and Laufer (2017) in several key aspects. Most notably, we construct a listing index based on the initial asking price of newly posted listings, rather than using the last observed listing price. This allows us to gain approximately two months of timeliness due to the period during which the property was advertised. While the final list price at delisting does incorporate listing price revisions, such revisions occur only for a small fraction (13.7%) of listings, and therefore capture only a limited portion of deviations between the initial asking price and the final sales price. Additionally, since a delisting does not indicate whether it results from a sale or a withdrawal, it offers no extra insight compared to using the initial listing in determining the reason for a property's removal. Finally, although the delisting date can provide some insight into the timing of the initial sales agreement, it remains an imprecise measure. Sold listings often remain online for weeks, as marketplaces typically charge a fixed fee for ad durations, giving sellers little incentive to remove their ad promptly after a sale.

A second contribution relative to Anenberg and Laufer (2017) is our novel methodological approach that takes into account the projected sales-to-listing price ratio, time-until-sale, and probability of sale. Finally, by controlling for a very extensive set of dwelling characteristics - including the improving energy efficiency of sold dwellings - we reduce the omitted variables problem that can distort the estimation of quality-adjusted house price indices Reusens, Vastmans, and Damen (2023).

This article also contributes to the literature on economic nowcasting—a term that combines "now" and "forecasting" to describe the prediction of the present, the very near future, or the very recent past of an economic indicator. A prominent application involves estimating current-quarter economic activity using timely indicators such as survey-based sentiment measures and financial market data, which is particularly valuable given that official statistics are typically released with a delay of several months. Seminal contributions by Giannone, Reichlin, and Small (2008) and Bańbura et al. (2013) introduced dynamic factor models that integrate large and heterogeneous datasets for real-time macroeconomic analysis.

The paper is structured as follows. Section 2 introduces the data, Section 3 details the methodology, Section 4 discusses the results, and Section 5 provides the conclusion.

2. Data

The analysis in this paper requires both a listing and a transaction dataset. The former is provided by Realo, which collects the data from multiple online platforms and real estate agents. It provides information on the first and last observed asking price, the date on which the advertisement was posted and removed, and a rich set of dwelling characteristics for advertisements listed between 2015 and the first quarter of 2025. Duplicate observations are removed from the data by Realo based on the address, characteristics, and timing. Such observations are either the result of a property being advertised on multiple platforms, or of sellers removing an advertisement and reposting it online again later, possibly because newly placed advertisements generate more search activity. Note that in the case of duplicates, the initial listing price is the one reported in the duplicate that was posted first. The listing time-on-market is measured as the time between the first duplicate being posted and the last duplicate being removed.

The transaction dataset is provided by the Federal Public Service Finance, and covers the universe of real estate transactions in Belgium, excluding new builds. Aside from various property characteristics, it includes the address, the cadastral parcel key (or CaPaKey), the sales price, the date of the notarial deed, and the type of transaction (e.g., private sale, public sale, or inheritance). In contrast to the listing dataset, the transaction dataset does not contain information about the energy performance of the dwellings. For this reason, we match the transaction dataset with data on energy performance certificates (EPC's) using the address and CaPaKey. This data was provided by the three regional energy agencies: VEKA (Flemish Region), SPW énergie (Walloon Region) and Brussels Environment. Each transaction is linked to the last issued EPC prior to the date of sale.[1] For the cases where we do not observe an EPC value for a transaction but are able to match the transaction to a listing, the EPC value reported in the listing is used.

In order to obtain a fair comparison between a listing and transaction index, it is important to remove outliers in both datasets in a similar way. As observations with extreme values were already removed from the listing dataset by Realo, we repeat this process using the same thresholds for

---

[1]We are not able to match the certificates to specific units in a building, as there is no consistent system of box numbers across databases. This results in some error on the energy performance variable for apartments.

the transaction dataset. Furthermore, we remove advertisements/transactions concerning holiday homes and public sales. Finally, following Eurostat's guidelines for outlier removal (Eurostat 2017), we drop observations with studentized residuals beyond the range of [-2, 2].

2.1. Matching the data

We link the listing data and the transaction data to each other in two steps. First, we match all plausible combinations of listings and transactions. A match is considered "plausible" if the following five conditions are met: (1) the street name[2] and the municipality match, (2) the house number matches or is unknown for the listing, the transaction, or both, (3) the timing between the market time of the advertisement and the sale date is plausible[3], (4) the characteristics of the property match within a certain margin, and (5) the asking price is no more than 30% higher or 20% lower than the sales price.

Using the above method, we already obtain a one-on-one match for the majority of listings/transactions that can be matched. However, some advertisements are linked to multiple transactions, or vice versa, either because of observations where the house number is missing, because multiple apartments with the same house number were offered/sold during a similar period, or because of rapid successive sales of the same property. To obtain a one-on-one match for these cases, we select the most optimal match for each transaction and for each listing. A match becomes definite if a listing and a transaction are each other's most optimal match. We repeat this process with the remaining listings and transactions until all plausible matches are allocated.

The most optimal match in this second step is determined using the following priority rules. We first give priority to matches where the house number for both the transaction and the listing is known. Then, priority is given to the match for which the characteristics of the property are most similar. If this similarity is the same for multiple matches with the same advertisement/transaction (often due to missing information in the ad), priority is given to the match with the smallest difference between the transaction and asking price.

---

[2]Fuzzy string matching techniques are used to correct for different spellings of street names across databases.
[3]That is, the transaction date is at least five days after the listing first appeared online, and no more than 300 days after the listing disappeared from the website.

The advantage of this approach is that we can match listings and transactions even when addresses are incomplete. In many Belgian streets, real estate transactions are infrequent, and we often observe only one listing and transaction in a given period. In this case, our method matches the observations even if the house number or dwelling characteristics are missing in one of the datasets. In more dynamic streets, our method also uses the dwelling characteristics to identify the most probable combinations of listings and transactions. While our approach also allows us to match data on apartments - which is not possible when matching only by address, as box numbers are inconsistent across datasets - the matching will likely be less accurate. This is because apartments within the same building often have similar characteristics.

2.2. Data description and comparison

Table I displays the average values of the price and dwelling characteristics in the listing dataset (columns 1 & 2) and the transaction dataset (columns 3 & 4). Columns 1 and 3 show the averages for the entire sample of listings/transactions, while columns 2 and 4 show the averages for the listings (transactions) that were matched to a transaction (listing). Note that columns 2 and 4 concern the same sample, and that the differences between the averages in these columns are the result of variables being defined differently in the two datasets. For example, there are different ways to measure the habitable area of a property. Measures might (partly) include or exclude a garage, cellar, attic, or areas covered by a sloped roof. The average suggests that in case of houses, sellers choose a more optimistic value to report in their advertisement, making their house about 10 m^2 larger on average. The number of bathrooms is likely under-reported in the administrative transaction dataset, as this information is used to calculate the cadastral income and property tax, while advertisers might use a favorable definition of "bathrooms". Furthermore, advertisers often report the year in which the dwelling underwent structural renovations as the building year, resulting in a lower average dwelling age in the listing dataset. Renovations are in turn rarely reported in an advertisement, especially if the renovation is not recent. Despite these differences in the definitions of variables, values are highly correlated for most variables.

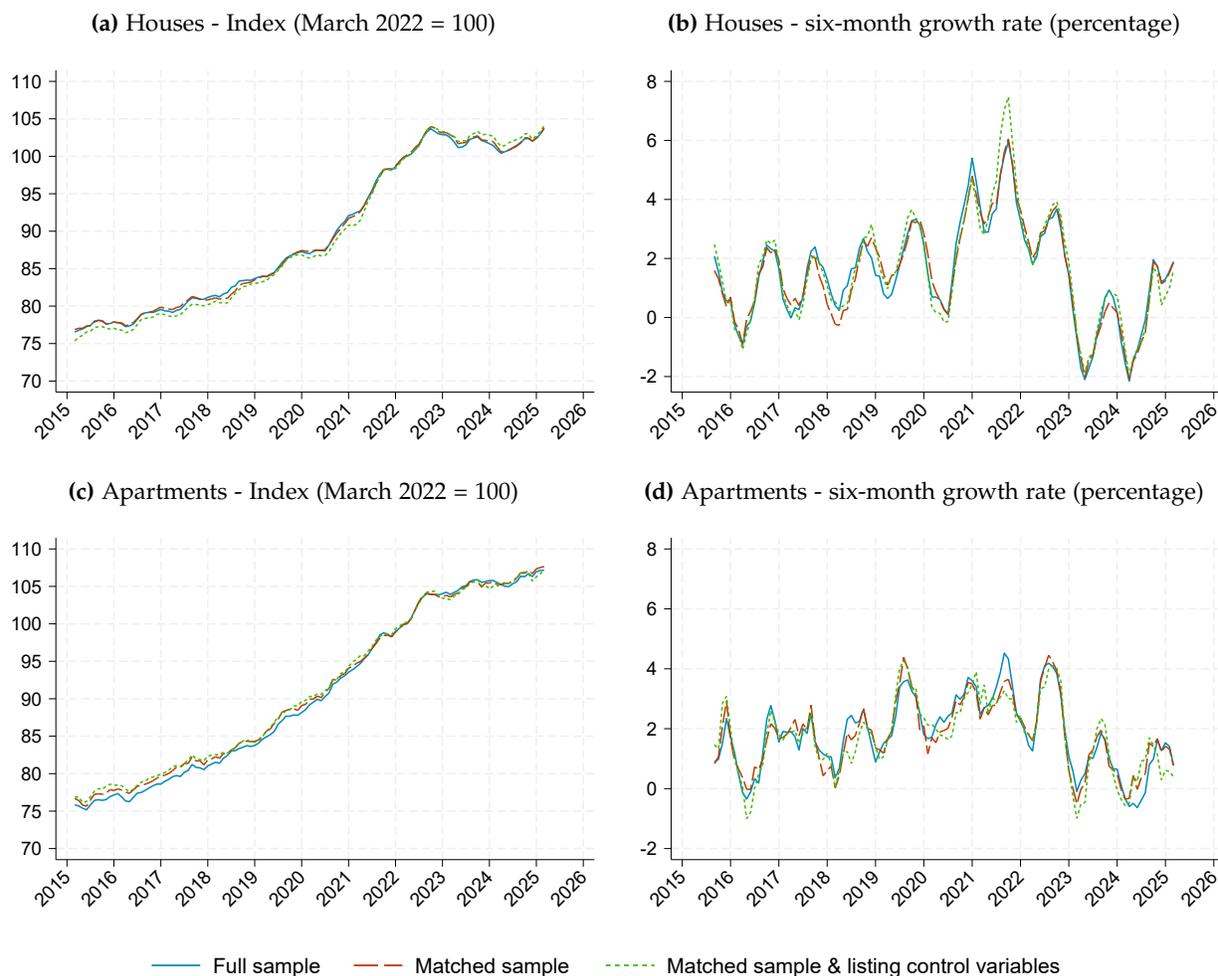**Table I:** Average price and dwelling characteristics across samples

| | Listings | | | Transactions | |
| | (1) | (2) | | (3) | (4) |
	Full sample	Matched		Full sample	Matched
Houses					
Initial listing price	312,703	302,003	Transaction price	280,739	286,476
Living area (m²)	184	177	Living area (m²)	169	167
Bedrooms	3.27	3.21	Rooms	5.58	5.54
Bathrooms	1.32	1.29	Bathrooms	0.89	0.89
Energy consumption (kWh/m²)	413	408	Energy consumption (kWh/m²)	424	410
Age	59	57	Age	75	74
Renovated (share)	0.03	0.03	Renovated (share)	0.31	0.31
Years since last renovation	10.55	10.15	Years since last renovation	16.95	16.95
Garden area (m²)	575	512	Garden area (m²)	522	503
One or two facades	0.28	0.30	Terraced	0.42	0.42
Three facades	0.23	0.24	Semi-detached	0.27	0.27
Four facades	0.29	0.28	Detached	0.31	0.31
Facades missing	0.20	0.18	Type missing	0.00	0.00
# Observations	781,937	457,017	# Observations	764,304	457,017
Apartments					
Initial listing price	257,691	245,790	Transaction price	223,976	235,852
Living area (m²)	96	92	Living area (m²)	90	92
Bedrooms	2.02	1.99	Rooms	3.78	3.82
Bathrooms	1.19	1.18	Bathrooms	1.04	1.04
Energy consumption (kWh/m²)	230	228	Energy consumption (kWh/m²)	251	239
Age	40	39	Age	47	46
Renovated (share)	0.02	0.02	Renovated (share)	0.18	0.17
Years since last renovation	8.02	7.74	Years since last renovation	9.86	10.67
# Observations	254,546	137,935	# Observations	315,292	137,935

Note: The table reports averages for variables in the listing dataset (columns 1 & 2) and the transaction dataset (columns 3 & 4). Columns 1 & 3 report the averages calculated on the full sample of listings/transactions, columns 2 & 4 the averages calculated on the sample of listings/transactions matched to a transaction/listing.

Aside from the differences in variables across the two datasets, Table I also informs about the representativeness of the listing data. Comparing columns 3 and 4 indicates that the transactions that are matched to listings have similar average values for characteristics compared to the full sample of transactions. Matched transactions have a somewhat higher sales price (2% higher for houses, 5.3% higher for apartments), suggesting that the listings database captures a slightly higher segment of the Belgian real estate market. Comparing columns 1 and 2 provides information about the type of dwellings that are sold. In particular, properties with a lower listing price and smaller properties have a higher likelihood of being sold.

Figure 1 compares a hedonic sales price index - in levels as well as six-month growth rates - based on the full sample of transactions (the blue line) with an index based on those transactions

Figure 1. Sales price index computed using the full sample of transactions, matched transactions, and matched transactions and listing variables

(a) Houses - Index (March 2022 = 100)

(b) Houses - six-month growth rate (percentage)

(c) Apartments - Index (March 2022 = 100)

(d) Apartments - six-month growth rate (percentage)

Full sample — — Matched sample ····· Matched sample & listing control variables

that are matched to a listing (the orange line) and an index computed using information on dwelling characteristics originating from the listing dataset (the green line). The indexes are estimated using the model described by Equation 1 in Section 3 (using sales prices and transaction dates). Indexes are smoothed by calculating a moving average of the current and past two months. The graphs show that for both housing types, the indexes are overall quite similar and mostly coincide. This suggests that transactions for which an advertisement is included in the listing data are representative of the universe of sales. Furthermore, the variables on property characteristics available in the listing dataset appear to control for heterogeneity between properties in a similar way to the variables available in the transaction dataset.

3. Methodology

In this section we discuss the methodology for the estimation of the hedonic price indexes. We estimate an index based on listing data as well as an index based on transaction data, so that we can analyze the differences by means of comparison (Section 3.1). Next, we will make adjustments to the estimation of the listing index with the goal of approximating and forecasting the transaction index (Section 3.2).

3.1. The listing and transaction price index

The primary difference between indexes based on listings and transactions is timing. In our transaction database, a sale is linked to the date on which the deed is signed. This date is not observed in data from real estate platforms. Instead, we observe the date on which the property appears and disappears from the website, which can be an indication of the listing time-on-market. It is this difference in timing that we exploit in order to forecast price evolutions, as for the most recent advertisements a transaction has yet to be observed.

Figure 2 indicates the definitions regarding the time-on-market. The total time-until-sale, which we define as the number of days between a dwelling first appearing online and the date the deed is signed, can be divided into two parts. First, there is the period during which the listing is advertised online. Second, there is a period of on average 90 days between delisting and the signing of the deed. The delisting date might indicate the time at which the seller and buyer reach an agreement. However, sellers often do not immediately remove the advertisement after finding a buyer.

Figure 2. Timeline
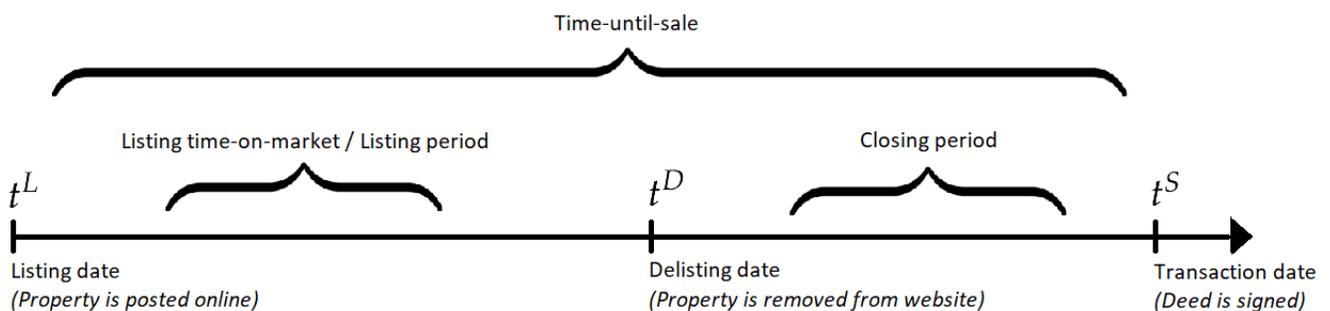
For the estimation the transaction price index, a standard log-linear hedonic regression model is used.

$$\log(p_i^S) = \delta_{t^S} + \beta x_i^S + \epsilon_i \qquad for\ i = 1, ..., N^S \qquad (1)$$

The dependent variable $(\log(p_i^S))$ is the log of the sales price of dwelling i, and the explanatory variables are fixed effects indicating the month in which the dwelling is sold (δ_{t^S}) and a vector of dwelling characteristics observed in the transaction dataset (x_i^S). ϵ_i indicates the error term. We estimate the model using the full sample of transactions (as indicated by $i = 1, ..., N^S$). The index is calculated using the time dummy method $(exp(\hat{\delta}_{t^S}))$, following Eurostat (2017).

The listing index is estimated in quite a similar way:

$$\log(p_i^L) = \delta_{t^L} + \beta x_i^L + \epsilon_i \qquad for\ i = 1, ..., N^L \qquad (2)$$

The difference with the transaction index is that we now control for the month in which the advertisement for dwelling i first appears online (δ_{t^L}) instead of the time at which the dwelling is sold. Furthermore, instead of the transaction price, the dependent variable (p_i^L) is the first observed listing price, and the model is estimated using the variables and observations from the listings dataset.

Recording listings in the index at the time they are posted allows us to make use of all listing observations, including listings not matched to a sale and recently posted listings for which the delisting date is not yet observed. In other words, it makes the index more "timely", and enables us to capture price trends earlier. A drawback is that we cannot use information that is only available at the end of the listing period, such as the final observed listing price, which might be a better indicator of the eventual sales price. However, only 13.7% of advertisements know a change in listing price in our sample.

The dwelling characteristics in the transaction dataset controlled for in Equation 1 (x_i^S) and those in the listing dataset controlled for in Equation 2 (x_i^L), are municipality indicators and the characteristics listed in table I. Continuous variables are modeled using piecewise linear functions,

as described in Reusens, Vastmans, and Damen (2023). We choose to control only for characteristics for which a variable is available in both the transaction and the listing dataset to obtain a comparable index. However, the definition of variables differs between datasets, as described in section 2.2.

Following the methodology of Anglin, Rutherford, and Springer (2003), we estimate the residual of Equation 2 for each listing. This estimated residual is referred to as the degree of overpricing (DOP), which can be interpreted as the percentage difference between the asking price and the expected asking price based on the listing characteristics and the listing date. The variable will serve as an explanatory factor in the following section.

3.2. Adjusting the listing index

In this section, we make three adjustments to the estimation of the listing index with the goal of approximating the transaction index, while still retaining the advantage of the timeliness of listing data. First, we estimate the time-until-sale and add it to the observed listing date to predict the transaction date for each listing (Section 3.2.1). Second, as not all dwellings listed online are sold, we weight the observations in the regression with the estimated probability of sale (Section 3.2.2). Third, we multiply the listing price of each listing by an estimated sales-to-listing price ratio (Section 3.2.3). The adjustments to the estimation of the listing price index are further discussed in Section 3.2.4.

3.2.1. Estimating the time-until-sale

In order to estimate a time-until-sale for each listing, we estimate the listing time-on-market and the closing period separately (see Figure 2 for the definitions). While the listing time-on-market can (partly) be explained by the characteristics of the property, the location, and the asking price, the variation in the closing period is mostly random as it depends on the time required by the notary to complete administrative procedures.

As a first step, we follow the methodology of Anenberg and Laufer (2017) by simulating a number of fictitious closing periods for each listing. This is achieved by randomly drawing R times

from the distribution of closing periods of listings matched to a sale (see Figure 3a).[4] Using this approach instead of simply assigning the average closing duration to each listing results in a more stable adjusted listing index for two reasons. First, the randomness will spread a peak or a trough in asking prices due to an economic shock in a given month over multiple months. This is also reflected in the transaction index, as the price is determined prior to signing the deed. Second, because we draw from the distribution of closing periods R times, a single listing will be duplicated R times, each duplicate with a different estimated time of sale. This results in a smoother index. We choose to set R=10, as increasing R further only has a negligible impact on the index.

As a second step, we estimate the following regression model predicting the listing time-on-market:

$$\log(TOM_i) = \delta_{t^L} + \gamma CLOSING_i + \theta DOP_i + \beta x_i^L + \epsilon_i \qquad for\ i = 1, ..., N^M \qquad (3)$$

Where TOM_i indicates the listing time-on-market for dwelling i. As before, x_i^L represents a vector of k dwelling characteristics from the listing dataset, and δ_{t^L} are fixed effects indicating the month in which the advertisement first appears online. We also control for the duration of the closing period ($CLOSING_i$), as sellers do not always immediately remove their advertisement after finding a buyer.[5] As such, sellers that take longer to remove their listing will appear to have a shorter closing period. This leads to a negative correlation between the observed closing period and the listing time-on-market. Finally, we control for the Degree Of Overpricing (DOP), as described in section 3.1. The model is estimated using the sample of listings matched to a sale ($i = 1, ..., N^M$).

After estimating the parameters of the model, we use the model to estimate a listing time-on-market for all listings (including the R duplicates). For this purpose, we use the simulated closing duration described earlier, as a real closing duration is not observed for listings not tied to a sale. The model only explains 13.62% (9.95% for apartments) of the variation in the listing time-on-market. This suggests that much of this variation is random, or related to variables not

---

[4]We allow this distribution to vary across years.

[5]Real estate agencies representing the seller might use advertisements for already sold homes to promote their business. This phenomenon also means that we may overestimate the true time-on-market of a dwelling. However, it could also be that a real estate agent first uses other media (a sign on the dwelling, own website, etc.) before advertising the dwelling on a real estate platform. In the latter case, we might underestimate the true time-on-market.

included in our dataset. To simulate this variation, we randomly draw R times from the distribution of estimated error terms ($\hat{\epsilon}_i$) of the model described by Equation 3 for each listing, and add this simulated error term to the predicted listing time-on-market.

The estimated time-until-sale is calculated by adding the estimated listing time-on-market, a simulated error term for this estimation, and a simulated closing period. By adding the estimated/simulated time-until-sale to the listing date (t^L) we obtain R predicted sales dates for each listing, as described in Equation 4.
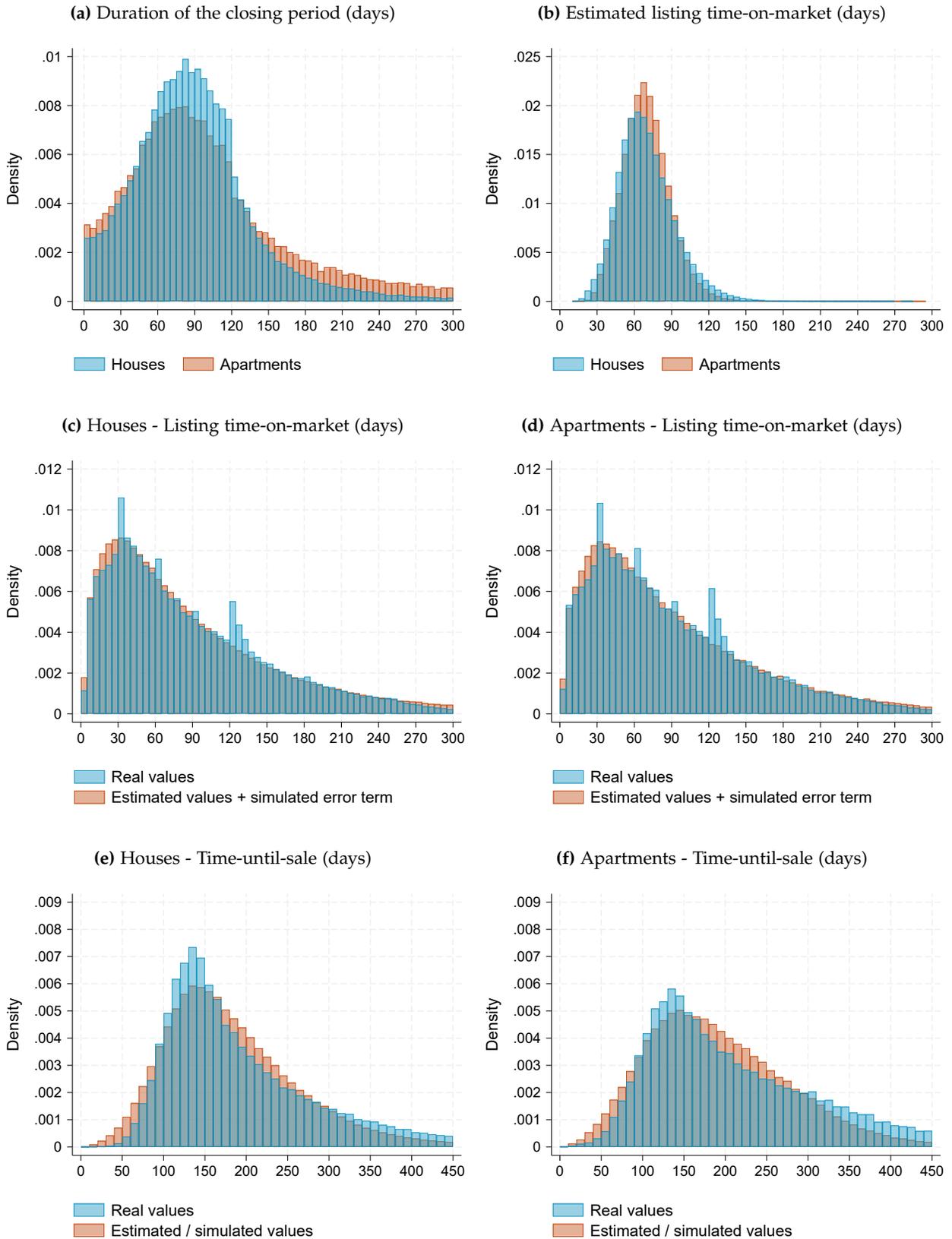
$$\widehat{t_{ir}^S} = t_i^L + exp(\log\widehat{(TOM_{ir})} + \widehat{\epsilon_{ir}}) + CL\widehat{OSIN}G_{ir} \qquad for\ i = 1, ..., N^L\ and\ r = 1, ..., R \qquad (4)$$

Because Equation 3 is estimated using the sample of listings matched to a sale with a total time-until-sale of less than 365 days, we can only estimate $\hat{\delta}_{t^L}$ up until one year before the last observed transaction. In other words, the time-on-market cannot be accurately predicted for the most recent listings, as we do not yet observe the entire listing period for these listings. Therefore, we extrapolate the trend of $\hat{\delta}_{t^L}$ in Equation 3 by taking the value of the coefficient for the most recent month for which it can be estimated, and adding average monthly seasonal effects.[6]

Figure 3a shows the distribution of the duration of the closing period, conditional on this duration being positive. A negative closing period occurs when a seller removes their advertisement from the real estate platform only after the deed is signed. We avoid assigning listings a negative closing period as this could result in a negative estimated/simulated time-until-sale. Furthermore, an advertisement is not matched to a transaction if this would result in a closing period longer than 300 days. Given this lower and upper limit, the average contract period is 89 days for houses and 98 days for apartments.

---

[6]Seasonal effects are calculated as the difference between the monthly average of $\hat{\delta}_{t^L}$ and its overall average. We remove seasonality before extrapolation, and reapply it after.

Figure 3. Distribution of the time-until-sale

(a) Duration of the closing period (days)

(b) Estimated listing time-on-market (days)

(c) Houses - Listing time-on-market (days)

(d) Apartments - Listing time-on-market (days)

(e) Houses - Time-until-sale (days)

(f) Apartments - Time-until-sale (days)

Note: The distributions represent the sample of listings that first appeared online between 2015 and the first quarter of 2024.

15

The distribution of the listing period estimated using the model specified by Equation 3, is shown on Figure 3b. By adding random draws from the distribution of estimated error terms of the model, we obtain the distributions indicated in orange on Figures 3c and 3d. These distributions are similar to the distributions of the real listing time-on-market shown in blue, except that the latter shows spikes at certain durations. Such peaks are likely the result of options for subscription periods that advertisers/sellers choose ex-ante on many real estate platforms. The height of these peaks gives insight into the share of advertisers who do not remove their ad after finding a buyer and keep it online until the end of the subscription period. For the observations with a total time-until-sale of less than one year and a positive closing period, the average (real) listing period is 88 days for houses, and 89 days for apartments.

Finally, Figures 3e and 3f show the distribution of the total time-until-sale for houses and apartments, respectively. The blue distributions represent the real time-until-sale observed in the dataset, while the distributions of the estimated/simulated time-until-sale are shown in orange. Simulated values exceeding 450 days are set to missing, which is the case for 1.4% / 1.1% of observations for houses/apartments.

While the estimated and real distributions have a similar shape, the average absolute prediction error is 79 days for houses and 88 days for apartments. This is because much of the variation in the time-until-sale is random, or at least related to factors not available in our dataset. However, for the purpose of adjusting the listing index, it suffices to explain the variation linked to the price or characteristics controlled for in the estimation of the price indexes.

3.2.2. Estimating the probability of sale

Next, we estimate the probability of sale for each listing. This is done using the following regression model:

$$SOLD_i = \delta_{t^L} + \theta DOP_i + \beta x_i^L + \epsilon_i \qquad for\ i = 1, ..., N^L \tag{5}$$

The dependent variable ($SOLD_i$) is a dummy indicating whether or not a listing is matched

16

to a transaction. Listings with a total time-until-sale longer than one year are also treated as "not sold". Note that this is a rough indication of whether a property is actually sold. It could be that a listing is not matched to a transaction because of misreported information in the listing (such as a wrong address), or missing or misreported information in the transaction data. Therefore, we likely underestimate the true probability of sale, especially for apartments as we cannot use box numbers to match datasets (they are not consistent across datasets). The regression controls for a range of dwelling characteristics (x_i^L), as well as the Degree Of Overpricing (DOP).

We use the model described by Equation 5 to estimate a probability of sale for all listings. Extrapolating $\hat{\delta}_{t^L}$ allows us to estimate the probability of sale for the most recent advertisements. Because we use a linear model, the estimated probability can be lower than zero or higher than one. It is set to zero if estimated negative. The distribution of the estimated probability of sale for houses and apartments first listed online in the period 2015 to Q1 2024 is shown on Figure 4.

Figure 4. Distribution of the estimated probability of sale
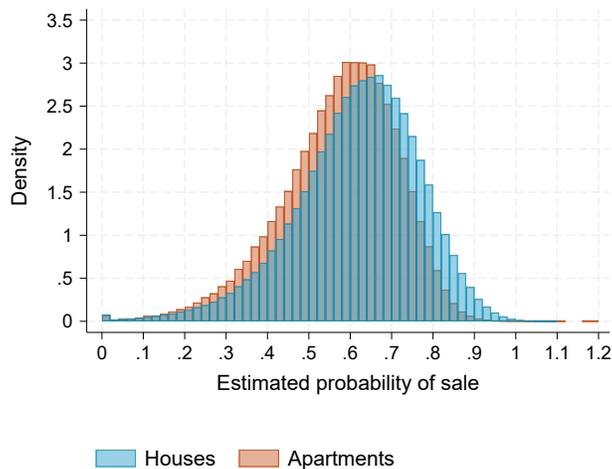
3.2.3. Estimating the sales-to-listing price ratio

Finally, we estimate the relative difference between the sales and the listing price using the following regression model

$$\frac{p_i^S}{p_i^L} = \delta_{t^S} + \theta DOP_i + \beta x_i^L + \epsilon_i \qquad for\ i = 1, ..., N^M \tag{6}$$

where $\frac{p_i^S}{p_i^L}$ indicates the sales-to-listing price ratio, DOP stands for the degree of overpricing, and x_i^L represents a vector of dwelling characteristics in the listing dataset. Unlike in Equations 3 and 5, Equation 6 controls for the month in which the dwelling is sold (δ_{t^S}). We find this to be a better predictor than the month at which a dwelling first appears online, as it is closer to the timing at which the sales-to-listing price ratio is determined. [7]

After estimating the parameters of the model, we estimate the sales-to-listing price ratio for all listings (including those not matched to a transaction). For this purpose, we use the estimated time of sale ($\widehat{t_{ir}^S}$ from Equation 4) instead of the real time of sale used to estimate the parameters of the model. As each listing has R estimated transaction dates, we also estimate R sales-to-listing price ratios for each listing:

$$\widehat{SLPR}_{ir} = \frac{\widehat{p_{ir}^S}}{p_{ir}^L} = \hat{\delta}_{\widehat{t_{ir}^S}} + \hat{\theta} DOP_i + \hat{\beta} x_i^L \qquad for\ i = 1, ..., N^M\ and\ r = 0, ..., R \tag{7}$$

As before, we extrapolate the trend of $\hat{\delta}_{t^S}$, by taking the value of $\hat{\delta}_{t^S}$ for the most recent month for which it can be estimated, and adding average monthly seasonal effects. Although we can estimate this parameter up until the last observed transaction date, the most recent listings will have an estimated date of sale beyond the end of our sample of transactions. The estimated sales-to-listing price ratio has an average absolute prediction error of 4.6 percentage points for houses and 5.7 percentage points for apartments.

The distributions of the estimated and true sales-to-listing price ratio are shown on Figure

---

[7]Note that the transaction date might also be a better predictor for estimating the time-on-market and the probability of sale. However, we cannot estimate a date of sale if doing so requires already having estimated the date of sale. Estimating the parameters of the model for predicting the probability of sale requires the full sample of listings, including those not matched to a transaction and for which we do not observe a transaction date.

5. The high bar with value 1 in the orange distribution indicates that for a significant share of transactions, the sales price is equal to the (initial) listing price. Furthermore, there is a bell shaped distribution on the left side of this bar, reflecting transactions for which the buyer was able to negotiate a "discount" on the asking price, with a reduction of about 5% being most frequent. Only a limited share of transactions have a sales price above the asking price. Although the distribution of the estimated sales-to-listing price ratios differs in shape from that of the real values, our primary objective - constructing a price index - mainly requires an accurate measurement of the ratio's evolution over time.

Figure 5. Distribution of the sale-to-listing price ratio

(a) Houses

(b) Apartments
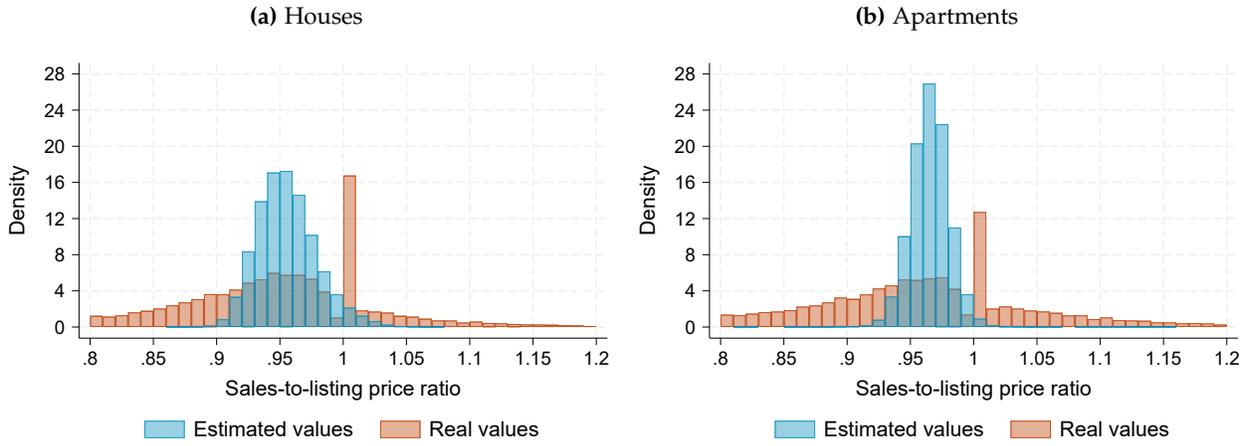
Note: The distributions represent the sample of listings matched to a sale recorded between 2016 and the first quarter of 2025.

3.2.4. Estimating the adjusted listing price index

After estimating the time-until-sale, the probability of sale, and the sales-to-listing price ratio for each listing, we can adjust the estimation of the listing index with the goal of approximating the transaction index. The adjusted listing index is estimated using the following model:

$$\log(p_{ir}^L * \widehat{SLPR_{ir}}) = \delta_{\hat{t}_{ir}^S} + \beta x_i^L + \epsilon_{ir} \qquad for \ i = 1, ..., N^L \ and \ r = 0, ..., R \qquad (8)$$

The first difference compared to the model described by Equation 2 is the timing. The month fixed effects ($\delta_{\hat{t}_{ir}^S}$) indicate the predicted month of sale (see Equation 4) rather than the month in

which the listing first appears online. Because we estimated R=10 possible months of sale for each listing, each listing will enter the model 10 times.

The dependent variable is the log of the first observed listing price multiplied by the estimated sales-to-listing price ratio (see Equation 7). Note that as the ratio is predicted using the predicted month of sale, $\widehat{SLPR_{ir}}$ will vary between the R duplicates of each listing.

Finally, we weight the observations using the probability of sale estimated in Section 3.2.2. In this way, advertised properties with a low likelihood of being sold (for example, because of undesirable characteristics or a high asking price given the characteristics) will have less weight in the regression.

The property characteristics indicated by x_i^L in Equation 8 originate from the listing dataset and are the same as in Equation 2 (municipality indicators and the characteristics listed in Table I). ϵ_{ir} indicates the error term.

4. Results

Our goal is to evaluate the performance of the adjusted listing price index estimated in real time (**Index E**) in forecasting the sales price index (**Index A**). We compare the forecasting ability of **Index E** with that of a simple historical average benchmark model (**Index F**). To better understand the sources of the forecast errors, we also estimate the adjusted listing price index on the semi-real-time sample (**Index D**) and the full sample (**Index C**). We also discuss the unadjusted listing price index (**Index B**). These indexes are defined as follows:

Index A The sales price index (Equation 1), estimated on the full sample of sales.

Index B The unadjusted listing price index (Equation 2), estimated on the full sample of listings.

Index C The adjusted listing price index (Equation 8), estimated on the full sample of listings.

Index D The semi-real-time forecast using the adjusted listing index (Equation 8). This forecast is estimated in real-time using an expanding window, except for the estimation of the sales-to-listing price ratio, the time-until-sale and the probability of sale (which are estimated on the full sample of listings and transactions).

Index E The real-time forecast using the adjusted listing index (Equation 8), estimated in real time using an expanding window.

Index F The real-time forecast using the benchmark model. In this model the forecasted price change is the average monthly price change of the transaction index (**Index A**) over the last observed year, multiplied by the length of the period that is being forecasted.

Section 4.1 analyzes the full sample estimates (**Index A**, **Index B** and **Index C**) and describes the differences between the three estimated indexes. Section 4.2 focuses on the evaluation of the real-time forecast performance. It should be noted that all indices are smoothed using a moving average of the current and past two months.

4.1. Full sample results

Figure 6 shows the evolution of the estimated quality-adjusted sales-to-listing price ratio (Equation 6), time-until-sale (Equation 3) and probability of sale (Equation 5) for a house with average characteristics.[8] The sales-to-listing price ratio and the probability of sale are relatively stable in the short term, but display important trends in the medium term, which are driven by the housing market cycle. In particular, they both strongly increased during the COVID-19 pandemic housing market boom, rising from 95% and 60% in 2020 to 98% and 70% in 2022, respectively, before falling back as a result of the rise in mortgage rates in 2022 and 2023.

The time-until-sale is relatively constant in the long term, fluctuating around 5 months for houses and 5.4 months for apartments. However, it is also affected by the housing cycle, being lower in the COVID-19 period and higher as a result of the rise in mortgage rates, and displays short term fluctuations.[9]
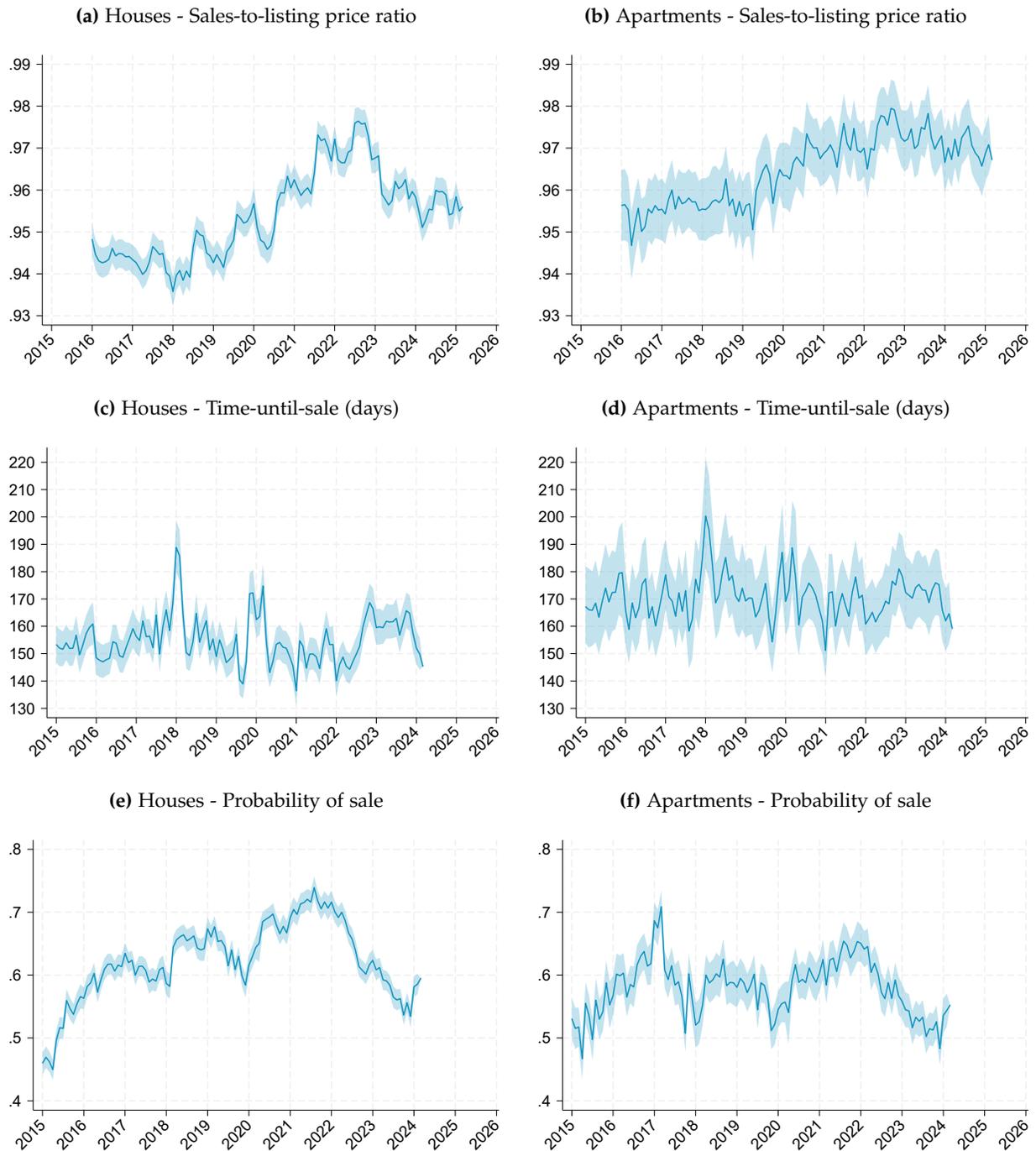
Figure 7 shows the estimated sales price index (**Index A**), unadjusted listing price index (**Index B**) and adjusted listing price index (**Index C**) for houses and apartments, both in levels and in the corresponding six-month growth rate (i.e. growth over the past six-month period). Figure 8 shows the cross correlation function between the six-month growth rate of the sales index on the one hand, and the six-month growth rate of the unadjusted listing price index (left column) or adjusted listing price index (right column) on the other hand.

The adjusted listing price more closely aligns with the sales index compared to the unadjusted listing price index, both regarding timing and price developments. Regarding timing, the time plot and the cross correlation function show that the unadjusted listing price index leads the sales index by about 4 to 5 months, which is consistent with the estimated time-until-sale for the average dwelling. This lead disappears in the adjusted listing index as this index takes into account the

---

[8]Note that unlike the simple average of these indicators, our quality adjusted indicators are not affected by changes in the sample composition of the dwellings sold over time. They can be interpreted as the estimated indicator for a dwelling with each of the characteristics equal to the total sample average, as in Reusens et al. (2025). It should also be noted that the sales-to-listing price ratio is computed at the time of sale, while the time-until-sale and probability of sale are computed at the listing date.

[9]There are several peaks and drops in the time-until-sale as a result of policy changes being implemented on a fixed date (often the 1st of January), which incentivizes buyers and sellers to move forward or delay transactions in order to benefit from more favorable policy conditions.

Figure 6. Evolution of the estimated sales-to-listing price ratio, time-until-sale and probability of sale for a house or apartment with average characteristics

(a) Houses - Sales-to-listing price ratio

(b) Apartments - Sales-to-listing price ratio

(c) Houses - Time-until-sale (days)

(d) Apartments - Time-until-sale (days)

(e) Houses - Probability of sale

(f) Apartments - Probability of sale

Note: The figure shows the sales-to-listing price ratio (in %, top row), time-until-sale (in number of days, middle row) and probability of sale (in %, bottom row) for houses (left column) and apartments (right column) estimated for a dwelling with average characteristics. The margins represent 95% confidence intervals based solely on the sampling uncertainty in the estimated time dummy coefficients. As a result, they capture uncertainty in the time effects only and not in the other regression coefficients.

Figure 7. Comparison of the unadjusted listing price index, adjusted listing price index, and sales price index

(a) Houses - Index (March 2022 = 100)

(b) Houses - six-month growth rate (percentage)

(c) Apartments - Index (March 2022 = 100)

(d) Apartments - six-month growth rate (percentage)

Sales price index (Index A) — — Unadjusted listing price index (Index B) ----- Adjusted listing price index (Index C)
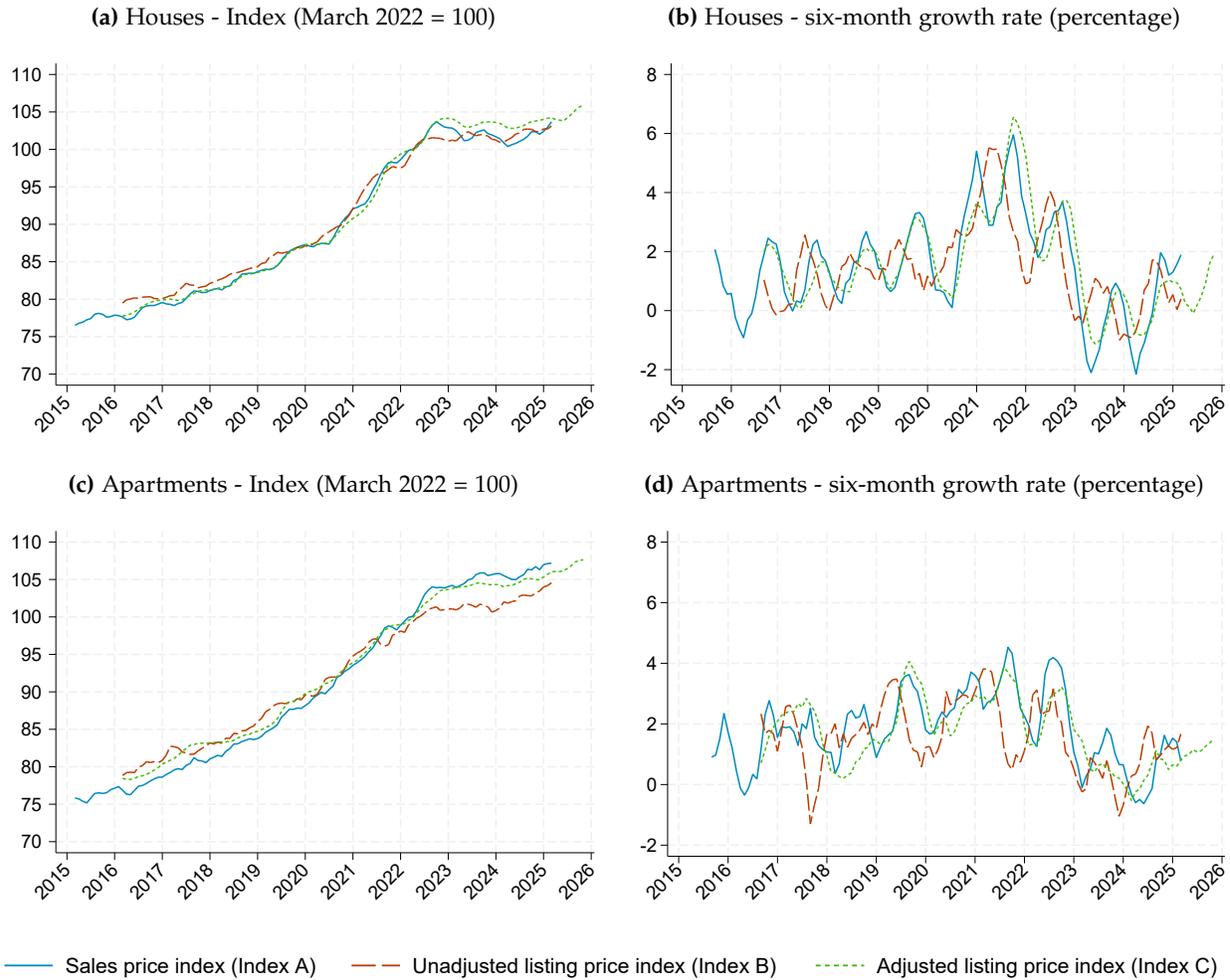
Note: The figure shows the time plot of the price indices (left column) and corresponding six-month growth rate (right column) for houses (top row) and apartments (bottom row).

estimated time-until-sale of the listings, but it should be noted that for houses our current procedure now yields a slight lag of less than one month relative to the sales price index.

In addition, the peak correlation is about 8 percentage points higher for the adjusted listing index (0.96 for houses and 0.86 for apartments) compared to the unadjusted index (0.88 for houses and 0.77 for apartments), which can be attributed to two factors. First, the unadjusted listing price index underestimates the price growth during the period 2019-2022 and overestimates it during the period 2022-2024, as a result of the increase and decrease in the sales-to-listing price ratio in these respective periods. The second reason relates to the mismatch between the start of the listing

Figure 8. Cross correlation function between six-month growth rates

(a) Houses - Unadjusted listing index (Index B)

(b) Houses - Adjusted listing index (Index C)

(c) Apartments - Unadjusted listing index (Index B)

(d) Apartments - Adjusted listing index (Index C)
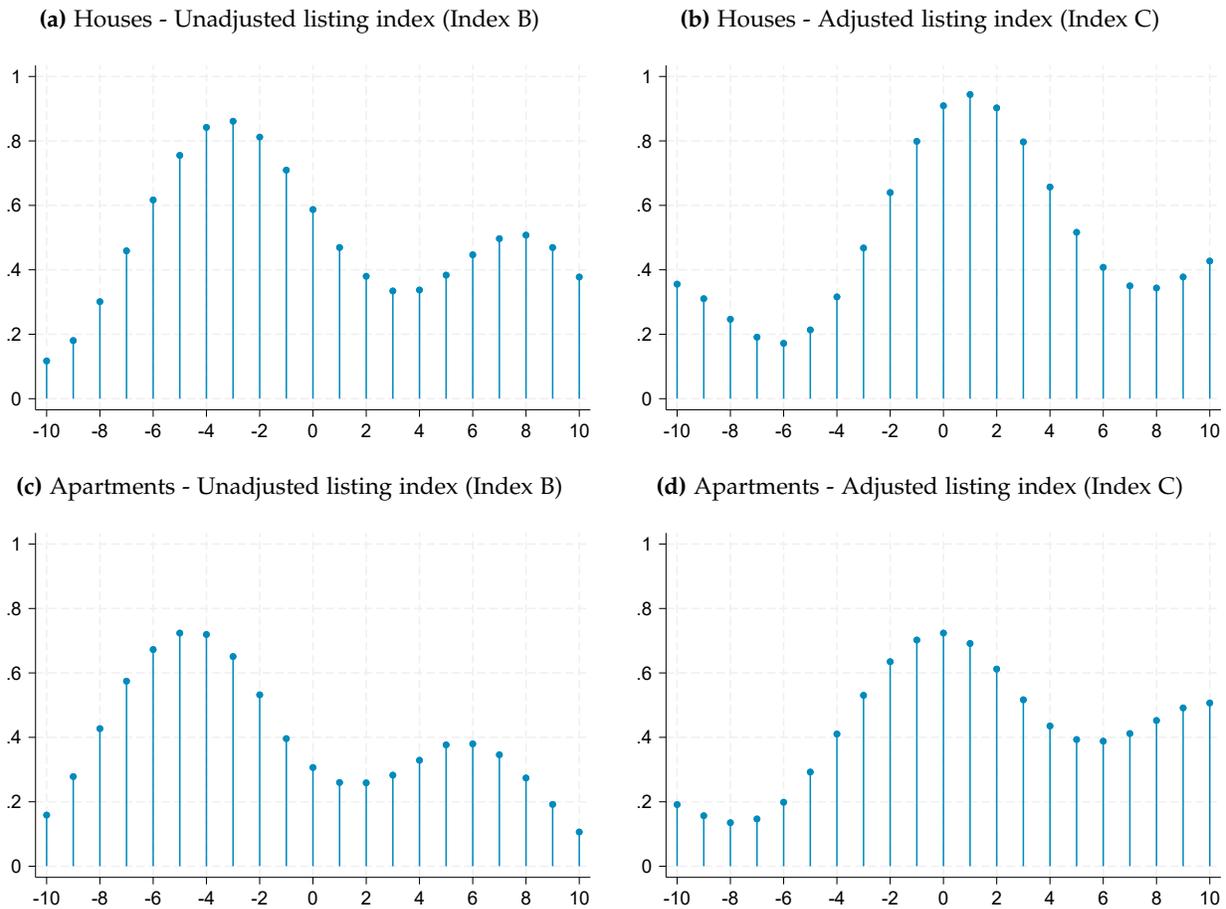
Note: The figure shows the cross correlation function between on the one hand the six-month growth rate of the sales index and, on the other hand, the six-month growth rate of the first listing price index (left column) or adjusted listing price index (right column), for houses (top row) and apartments (bottom row). The vertical axis shows the correlation between listings[t+h] and sales[t], where the lag h is expressed in number of months and is shown on the horizontal axis.

period and the time of sale. Because of the large variation in the time-until-sale of different listings, the dwellings sold in a certain month have listing dates that span multiple months. The adjusted listing index takes this timing difference into account by simulating the time-until-sale for each listing (see Section 3.2.1).

25

4.2. Evaluating the forecasting performance

The adjusted listing index (see Section 3.2) is more timely than the sales price index, as it utilizes listings for which a sale has not yet occurred and projects the prices to the estimated future time of sale. Section 4.2.1 tests the ability of the adjusted listing index to forecast the sales index in this way. In Section 4.2.2, we examine the causes of the forecasting error in more detail. Finally, Section 4.2.3 provides a graphical analysis of the forecast performance.

4.2.1. The real-time forecasting performance

The forecast performance will be evaluated using the following method: First, we estimate the adjusted listing index using the model specified by Equation 8, using all listing and transaction data available up until month τ, for each month τ from January 2017 through July 2023.[10] As such, we estimate the model using an expanding window, i.e. iterating over 84 different samples, each sample containing data for one additional month. We also estimate the time-until-sale, the probability of sale and the sales-to-listing price ratio (see Equations 3, 5 and 6) for each listing using only the data available up until month τ, as they are required for the calculation of the adjusted listing price index. Because we do not observe transactions for listings that are posted in or close to month τ, we extrapolate the trend in the three estimated variables by taking the estimated value for the last month for which we observe the indicator, and adding average monthly seasonal effects (as described in Section 3.2).

Next, for each forecast horizon h between 1 and 8 months, we compare the change in the estimated adjusted listing indexes between month τ and $\tau + h$ (the forecast) with the change for the corresponding period in the transaction index estimated using regression Equation 1 and the full sample of transactions. In other words, we compare the forecasted price changes measured using the adjusted listing index (**Index E**) with the price changes measured ex-post using the transaction index (**Index A**). In order to obtain stable indexes to compare, we calculate a three month average for each index value (the average of the current and past two months).

The results of this real time forecasting exercise are shown in Table II. The root mean square

---

[10]We assume the transaction and listing datasets are available without delay.

error (RMSE) and the mean absolute error (MAE) indicate that the forecasted price changes between τ and $\tau + 6$ (6 months horizon) deviate from the change in the ex-post transaction price index by on average about 1 to 1.2 percentage points. At the same horizon, the correlation coefficient between the forecast for houses/apartments and the six-month price change measured using the transaction index estimated on the full sample, amounts to 0.68/0.48. The mean error (ME) is defined as the mean of the price change according to the transaction index (**Index A**) minus the forecasted price change (**Index E**). The positive mean error for apartments thus suggests that the forecast is downward biased. The forecasted price change is on average about 0.7 percentage points lower than the change in the transaction index for a horizon of 6 months.

Table II: Real-time forecast performance of the adjusted listing index

Horizon (months)	Houses				Apartments			
	RMSE	MAE	ME	Corr.	RMSE	MAE	ME	Corr.
1	0.003	0.002	0.000	0.745	0.004	0.003	0.001	0.367
2	0.005	0.004	0.001	0.790	0.005	0.004	0.002	0.512
3	0.007	0.005	0.002	0.787	0.007	0.006	0.004	0.555
4	0.008	0.007	0.002	0.777	0.009	0.007	0.004	0.556
5	0.010	0.008	0.002	0.743	0.010	0.008	0.005	0.561
6	0.012	0.010	0.001	0.679	0.012	0.010	0.007	0.483
7	0.015	0.012	0.001	0.610	0.014	0.011	0.008	0.427
8	0.017	0.014	0.000	0.540	0.015	0.012	0.009	0.384

Note: The table shows the root mean squared error (RMSE), the mean absolute error (MAE), the mean error (ME) and the correlation (Corr.) comparing price changes forecasted by the adjusted listing index using an expanding window between January 2017 through July 2023 (**Index E**), with the corresponding change in the transaction index estimated using the full sample (**Index A**). The mean error is defined as the mean of the change in the index estimated using the full sample (**Index A**) minus the forecasted price change (**Index E**).

As the horizon becomes longer, the RMSE and MAE increase, but the correlation coefficient also increases at first. This is because at longer horizons, we also compare the price change over a longer time period. While the month-to-month change in the adjusted listing price index and the transaction price index might differ because of statistical noise, the indexes measure the same trend and thus coincide in the long term. The correlation is highest at a 2-month horizon for houses and at a 5-month horizon for apartments. At this point the disadvantage of listings posted before month τ becoming less relevant and needing to extrapolate the trends in the time-on-market, probability

of sale, and sales-to-listing price ratio further begins to outweigh the benefit of comparing a longer time period.

In order to evaluate the results shown in Table II, we compare them with the forecast performance of a benchmark model (**Index F**). In this model the forecasted price change is the average monthly price change of the transaction index of the last observed year multiplied by the length of the forecast horizon. The forecast performance of the benchmark model is reported in Table A.I in the appendix. The benchmark model performs quite well, likely as a result of the relatively stable price trends in Belgium during the analyzed period. The adjusted listing index (Table II) performs better than the benchmark model, with the correlation coefficient being about 0.2 to 0.5 higher, depending on the forecast horizon. For the sample of apartments, the forecast error of the benchmark model is quite similar to that of the adjusted listing index.

4.2.2. Decomposition of the forecast error

In order to identify the cause of the error of the adjusted listing index in forecasting the sales index, we repeat the forecast evaluation in Section 4.2.1, except that we now estimate the time-until-sale, probability of sale, and sales-to-listing price ratio (see Equations 3, 5 and 6) for each listing using the full dataset.

First, we compare this semi-real-time forecast (**Index D**) with the corresponding change in the transaction index (**Index A**). This test allows us to assess the predictive power of listings for future sales, without the results being affected by the extrapolation of the three variables used for the adjustment of the listing index. Second, we compare the semi-real-time forecast (**Index D**) with the corresponding price change measured using the adjusted listing index estimated using the full sample (**Index C**). This test will also not be affected by the difference between the adjusted listing index (**Index C**) and the transaction index (**Index A**). As such, the results reflect the best possible performance that could be achieved by using listings to predict future sales - that is, if the (full sample) adjusted listing index would perfectly coincide with the transaction index and if we were able to perfectly extrapolate the trends in time-until-sale, probability of sale, and sales-to-listing price ratio.

The results of the first test, comparing the price changes measured using **Index D** with those measured using **Index A**, are reported in Table III. Comparing the results in Table III with the main forecast performance results in Table II indicates that having to extrapolate the time-until-sale, probability of sale, and sales-to-listing price ratio increases the RMSE and MAE by up to 0.6 percentage points for houses, and by up to 0.4 percentage points for apartments.[11] As can be expected, this additional error is larger for longer horizons, where we have to extrapolate the trends further. For a horizon of one or two months, the difference is negligible. The method of extrapolation used in this paper (taking the estimated value for the most recent month and adding average monthly seasonal effects) is quite simple, and could be improved in future research.

Table IV compares the semi-real-time forecast (**Index D**) with the price changes measured using

---

[11]Table A.II shows the real-time forecast performance when using alternative (simple) extrapolation methods. The results are similar to those reported in Table II.

Table III: Semi-real-time forecast performance of the adjusted listing index in forecasting the sales index

Horizon (months)	Houses				Apartments			
	RMSE	MAE	ME	Corr.	RMSE	MAE	ME	Corr.
1	0.003	0.002	0.000	0.802	0.003	0.003	0.000	0.430
2	0.004	0.003	0.000	0.850	0.004	0.003	0.001	0.646
3	0.005	0.004	0.000	0.868	0.005	0.004	0.001	0.722
4	0.006	0.005	0.000	0.885	0.005	0.004	0.002	0.827
5	0.007	0.006	0.000	0.883	0.007	0.005	0.003	0.822
6	0.008	0.007	-0.001	0.870	0.008	0.006	0.004	0.776
7	0.010	0.008	-0.001	0.853	0.010	0.008	0.006	0.712
8	0.012	0.009	-0.002	0.837	0.012	0.009	0.007	0.690
Difference with main results reported in Table II								
1	-0.000	-0.000	-0.001	+0.057	-0.000	-0.000	-0.001	+0.063
2	-0.001	-0.001	-0.001	+0.060	-0.001	-0.001	-0.001	+0.134
3	-0.002	-0.001	-0.002	+0.081	-0.002	-0.002	-0.002	+0.167
4	-0.002	-0.002	-0.002	+0.108	-0.003	-0.003	-0.002	+0.272
5	-0.003	-0.003	-0.002	+0.140	-0.003	-0.003	-0.002	+0.261
6	-0.004	-0.004	-0.002	+0.191	-0.004	-0.003	-0.002	+0.294
7	-0.005	-0.004	-0.002	+0.244	-0.003	-0.003	-0.002	+0.285
8	-0.006	-0.005	-0.002	+0.296	-0.004	-0.003	-0.002	+0.306

Note: The table shows the root mean squared error (RMSE), the mean absolute error (MAE), the mean error (ME) and the correlation (Corr.) comparing price changes forecasted by the adjusted listing index using an expanding window between January 2017 through July 2023 (**Index D**) with the transaction index estimated using the full sample (**Index A**). The mean error is defined as the mean of the change in the index estimated using the full sample minus the forecasted price change. The time-until-sale, probability of sale, and sales-to-listing price ratio used to estimate the adjusted listing indexes are estimated using the full sample.

the adjusted listing index estimated using the full sample (**Index C**). For the shorter horizons, the forecast yields close to no error. This is not surprising, as few advertisements have a (predicted) time-until-sale of less than 2 or 3 months. With the listing data available up until month τ, we therefore observe most of the listings with a predicted sales date in the first three months following month τ. [12] Likewise, we only observe about 13% of the listings with a predicted transaction month of $\tau + 8$.

Comparing the results shown in Table IV with those shown in Table III indicates that a considerable share of the real-time forecast error can be explained by the difference between the adjusted listing price index (**Index C**) and the transaction price index (**Index A**). Although the two

---

[12]An error at short horizons in Table IV could also have been the result of coefficients for control variables being estimated on a smaller sample for the forecast. However, this does not seem to impact the forecast.

indexes are highly correlated, especially for houses, some differences remain (see Figure 7 and 8 in Section 4.1). Some of these differences might be the result of the sales and listing datasets not containing the exact same variables (see Section 2.2). Furthermore, by drawing from the distribution of contract periods multiple times (See section 3.2.1), the adjusted listing index is smoother than the sales index. However, we cannot conclude that either index reflects the true underlying movements in real estate prices more accurately.

Table V summarizes the decomposition of the forecast error at a horizon of 6 months, by comparing the changes in the sales price index (**Index A**) with changes in the adjusted listing index (**Index C**) in the first row, the semi-real-time forecast (**Index D**) in the second row, and the real time forecast (**Index E**) in the third row.

Table IV: Semi-real-time forecast performance of the adjusted listing index in forecasting the adjusted listing index

Horizon (months)	Houses				Apartments			
	RMSE	MAE	ME	Corr.	RMSE	MAE	ME	Corr.
1	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.997
2	0.000	0.000	0.000	0.999	0.001	0.000	0.000	0.997
3	0.000	0.000	0.000	0.999	0.001	0.001	0.000	0.996
4	0.001	0.001	0.000	0.998	0.002	0.001	0.001	0.993
5	0.002	0.002	0.000	0.993	0.003	0.002	0.002	0.988
6	0.003	0.003	0.000	0.983	0.004	0.003	0.003	0.980
7	0.005	0.004	-0.001	0.969	0.006	0.004	0.004	0.970
8	0.007	0.006	-0.002	0.952	0.008	0.006	0.005	0.961
Difference with results reported in Table III								
1	-0.002	-0.002	+0.000	+0.197	-0.003	-0.003	-0.000	+0.566
2	-0.004	-0.003	+0.000	+0.150	-0.004	-0.003	-0.001	+0.349
3	-0.005	-0.004	+0.000	+0.131	-0.005	-0.004	-0.001	+0.270
4	-0.005	-0.004	+0.000	+0.113	-0.004	-0.003	-0.001	+0.158
5	-0.005	-0.004	+0.000	+0.111	-0.004	-0.003	-0.002	+0.147
6	-0.005	-0.004	+0.000	+0.115	-0.004	-0.003	-0.002	+0.166
7	-0.004	-0.004	+0.001	+0.120	-0.004	-0.004	-0.002	+0.197
8	-0.004	-0.003	+0.001	+0.120	-0.004	-0.004	-0.003	+0.183

Note: The table shows the root mean squared error (RMSE), the mean absolute error (MAE), the mean error (ME) and the correlation (Corr.) comparing price changes forecasted by the adjusted listing index using an expanding window between January 2017 through July 2023 (**Index D**), with the corresponding changes in the adjusted listing index estimated using the full sample (**Index C**). The mean error is defined as the mean of the change in the index estimated using the full sample minus the forecasted price change. The time-until-sale, probability of sale, and sales-to-listing price ratio used to estimate the adjusted listing indexes are estimated using the full sample.

Table V: Summary of of the decomposition of the forecast performance at a horizon of 6 months
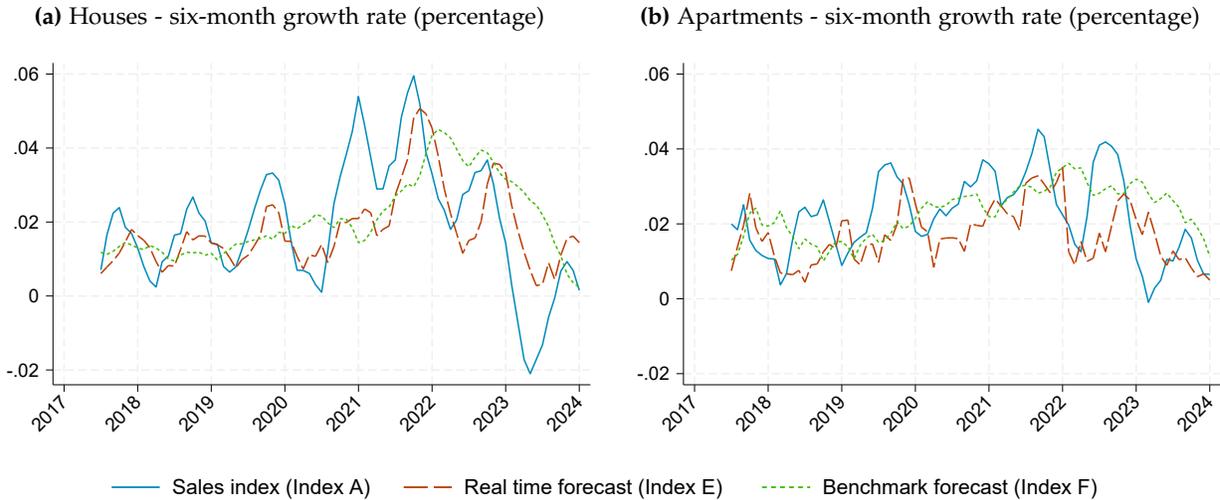
	Houses				Apartments			
	RMSE	MAE	ME	Corr.	RMSE	MAE	ME	Corr.
Index C	0.007	0.005	0.000	0.918	0.006	0.004	0.002	0.861
Index D	0.008	0.007	-0.001	0.870	0.008	0.006	0.004	0.776
Index E	0.012	0.010	0.001	0.679	0.012	0.010	0.007	0.483

Note: The table shows the root mean squared error (RMSE), the mean absolute error (MAE), the mean error (ME) and the correlation (Corr.) comparing six-month growth rates in prices measured using the sales price index (**Index A**) with (1) six-month growth rates in prices measured using the adjusted listing index estimated on the full sample (**Index C**), (2) the six-month growth rates forecasted using the adjusted listing index, estimated using an expanding window except for the estimation of the sales-to-listing price ratio, the time-until-sale and probability of sale (**Index D**), and (3) the real-time six-month forecasts using the adjusted listing index, estimated using an expanding window (**Index E**).

4.2.3. Graphical analysis of the forecast performance

In order to better understand the forecast we analyze it visually on Figure 9. The figure compares the six-month growth rate measured using the sales price index estimated using the full sample of transactions (**Index A**) with the real-time forecasted price change using the adjusted listing index (**Index E**) or the benchmark model (**Index F**) at a six-month forecast horizon. Note that for the forecasts, each point on the graph is estimated using a different sample, only containing data up to six months prior to the corresponding month on the graph.

Figure 9. Comparison of the sales index, real-time forecast, and benchmark forecast

(a) Houses - six-month growth rate (percentage)

(b) Apartments - six-month growth rate (percentage)

Sales index (Index A) — Real time forecast (Index E) — Benchmark forecast (Index F)

Note: The figure shows the six-month growth rate measured using the sales price index estimated using the full sample of transactions (**Index A**), the corresponding real-time forecast using the adjusted listing index (**Index E**), and the forecast using the benchmark model (**Index F**).

Much of the variation predicted by the adjusted listing index, but not captured by the benchmark model, is attributable to seasonality. Indeed, Figure 7 shows that while sales prices tend to be highest in the autumn, initial listing prices appear to peak during the summer. By shifting listing prices forward in time, most of this variation can be predicted. However, one could easily construct a benchmark model that also accounts for such seasonality.

While it is not difficult to forecast housing prices in a market with stable growth rates, it is more interesting to detect turning points in the sales price index early, and to predict the magnitude of price changes at those turning points. The sales index has two such turning points during the period analyzed in this study (both of which are more pronounced in the index for houses than in the index for apartments). The first is a sharp increase in price growth rates following the start of the COVID-19 pandemic in 2020. The second turning point is a drop in growth rates caused by increasing mortgage interest rates in the second half of 2022. Both of these turning points are primarily caused by external factors, and thus impossible to predict using just information on listings. However, once the market is hit by such an external shock, listing prices will likely react faster than sales prices, given the contract period of approximately three months on average.

Let's consider the perspective in September 2022. Just two months earlier, the ECB had announced an increase in its interest rates for the first time in over 10 years. While available transaction data would not yet have shown any decrease in prices, the forecast using the adjusted listing index would have informed us that (quality adjusted) sales prices for houses would only increase by 1.76% in the coming six months. This is significantly lower than the average six-month growth rates in the years prior. The prediction, however, overestimates the real growth rate in sales prices between September 2022 and April 2023, which amounts to -0.7% (if estimated using **Index A**, or 0.02% if estimated using **Index C**). This is because, while we would already have observed declining listing prices, we would not yet have observed the extent to which the time-until-sale, probability of sale, and (importantly) the sales-to-listing price ratio had responded to the market shock.

Figure 10 compares six-month price growth, measured using (1) the adjusted listing index estimated on the full sample (**Index C**), (2) the index estimated with an expanding window (**Index**

E), and (3) the index estimated with an expanding window but using the full sample to estimate time-until-sale, probability of sale, and the sales-to-listing price ratio (**Index D**). **Index C** and **Index D** are closely aligned, even at the turning point at the end of 2022. This suggests that our forecasting method's accuracy in predicting the magnitude of price changes at turning points depends on how well we can extrapolate or forecast the evolution of the time-until-sale, the probability of sale, and the sales-to-listing price ratio. Future research could explore ways to improve the method of extrapolation for these indicators, either by using information already contained in the listing data (e.g. recent listing price developments or online search activity) or by incorporating external, timely data (e.g. mortgage interest rates).

Figure 10. Comparison of the sales index, semi-real-time forecast, and real-time forecast
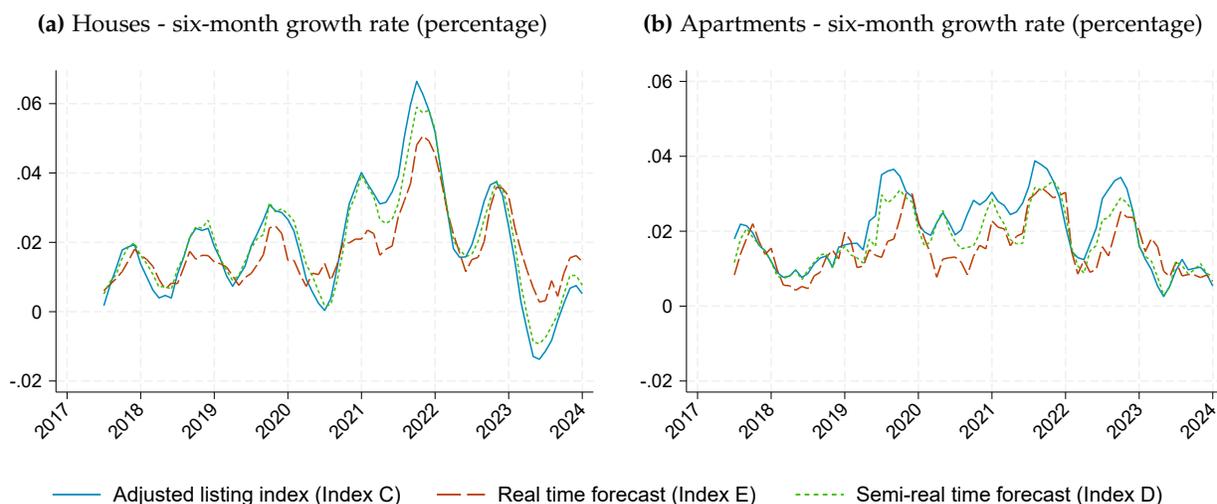
(a) Houses - six-month growth rate (percentage)   **(b)** Apartments - six-month growth rate (percentage)

Note: The figure shows the six-month growth rate measured using the adjusted listing price index estimated using the full sample of transactions (**Index C**), and the corresponding real time and semi-real-time forecast using the adjusted listing index (**Index E** and **Index D**)

5. Conclusion

In this paper we constructed a more timely house price index for Belgium by integrating listing and transaction data. The key idea is that asking prices for properties that are currently being advertised online can serve as a predictor of sales prices of transactions that will occur in a couple of months. To obtain a more timely index, we start from an index based on listing prices and make three adjustments. We move the index forward in time by estimating a time-until-sale for each listing, we weigh observations in the index by estimating the probability that a property advertised online will be sold, and we multiply the asking price by the estimated relative difference between the asking and transaction price.

Our novel method results in a listing index that not only closely aligns with the traditional transaction index but also anticipates price changes in the coming months. The assessment of the index's forecasting accuracy indicates an average six-month prediction error of 1 percentage point for houses and 0.9 percentage points for apartments, outperforming a benchmark forecasting model. Visual inspection of the forecast reveals that listing prices can predict changes in sales prices shortly after an external shock hits the market. The accuracy of these predictions, however, depends on the ability to anticipate changes in the time-until-sale, the probability that a property will get sold, and the sales-to-listing price ratio. These indicators are also likely to change together with market conditions, but they are only observed with the same delay as transaction prices. In this way, our method shifts the challenge from predicting price changes to predicting evolutions in secondary market indicators.

While our study uses simple extrapolation methods to predict movements in these secondary market indicators, future research could improve these predictions by using information already contained in the listing data (e.g. recent listing price developments or online search activity) or by incorporating external, timely indicators (e.g. mortgage interest rates or consumer confidence indicators). Furthermore, machine learning techniques could improve the accuracy of both the estimation and the extrapolation of the time-until-sale, probability of sale, and sales-to-listing price ratio.

Future research could investigate the extent to which the predictive power of initial listing prices generalizes across different countries. We anticipate that the forecasting value of listing data is greater in countries where there is a longer delay between signing the sales agreement and the official registration of the transaction. For example, in countries like Belgium and France, this delay averages around three months, whereas in the United States, the process typically takes about six weeks (Anenberg and Laufer 2017). This longer interval means that listing prices in Belgium can offer more forward-looking information, enabling more accurate forecasting of official sales price indexes.

Furthermore, the usefulness of listings for forecasting improves when the sales-to-listing price ratio remains relatively stable over time. As such, our method might also prove useful for forecasting price developments in the rental market, as rental prices are rarely negotiated. Finally, the value of house price forecasting in general is greater in countries that experience more volatile housing market patterns, as these are more difficult to anticipate than the relatively stable house price growth observed in Belgium.

In this study, we analyze the ability of a listings-based index to anticipate changes in a traditional index in which sales prices are recorded at the time the deed is signed. However, Anenberg and Laufer (2017) suggest that an index in which prices are recorded at the time the buyer and seller negotiate the price is better correlated with market conditions and general economic developments. While we do not observe this time directly in our listing or transaction dataset, the date at which a listing is removed from a real estate website provides a close approximation. As we estimate the listing and contract period separately, our methodology could also be used to approximate and forecast an index in which sales prices are recorded at the time of delisting.

REFERENCES

Anenberg, Elliot, and Steven Laufer, 2017, A more timely house price index, *Review of Economics and Statistics* 99, 722–734.

Anglin, Paul M., Ronald Rutherford, and Thomas M. Springer, 2003, The trade-off between the selling price of residential properties and time-on-the-market: The impact of price setting, *The Journal of Real Estate Finance and Economics* 26, 95–111.

Bańbura, Marta, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin, 2013, Now-casting and the real-time data flow, in *Handbook of economic forecasting*, volume 2, 195–237 (Elsevier).

Eurostat, 2017, *Technical Manual on Owner-Occupied Housing and House Price Indices*, Eurostat.

Giannone, Domenico, Lucrezia Reichlin, and David Small, 2008, Nowcasting: The real-time informational content of macroeconomic data, *Journal of monetary economics* 55, 665–676.

Gluszak, Michal, and Radoslaw Trojanek, 2024, Supply, demand and asymmetric adjustment of house prices in poland, *Real Estate Management and Valuation* 32, 31–45.

Kolbe, Jens, Rainer Schulz, Martin Wersing, and Axel Werwatz, 2021, Real estate listings and their usefulness for hedonic regressions, *Empirical economics* 1–31.

Lyons, Ronan C, 2019, Can list prices accurately capture housing price trends? insights from extreme markets conditions, *Finance Research Letters* 30, 228–232.

Reusens, P, F Vastmans, J Vandenbergh, T van Kempen, and S Damen, 2025, Impact of the flemish energy renovation obligation on house prices, *Economic Review* 1–44.

Reusens, Peter, Frank Vastmans, and Sven Damen, 2023, A new framework to disentangle the impact of changes in dwelling characteristics on house price indices, *Economic Modelling* 123, 106252.

Wang, Xiaodan, Keyang Li, and Jing Wu, 2020, House price index based on online listing information: The case of china, *Journal of Housing Economics* 50, 101715.

Appendix A. Additional Figures & Tables

Table A.I: Forecast performance of the benchmark model

Horizon (months)	Houses				Apartments			
	RMSE	MAE	ME	Corr.	RMSE	MAE	ME	Corr.
1	0.004	0.003	0.000	0.300	0.004	0.003	0.000	0.087
2	0.007	0.006	0.000	0.297	0.006	0.005	0.000	0.104
3	0.010	0.008	-0.001	0.293	0.008	0.006	0.000	0.128
4	0.012	0.010	-0.001	0.306	0.009	0.007	0.000	0.198
5	0.014	0.011	-0.001	0.315	0.010	0.008	-0.001	0.218
6	0.016	0.013	-0.001	0.320	0.011	0.009	-0.001	0.217
7	0.018	0.014	-0.002	0.322	0.013	0.010	-0.001	0.228
8	0.020	0.016	-0.002	0.323	0.014	0.011	-0.001	0.268
Difference with main results reported in Table II								
1	+0.001	+0.001	-0.001	-0.446	+0.000	-0.000	-0.001	-0.280
2	+0.002	+0.002	-0.001	-0.493	+0.000	+0.000	-0.002	-0.408
3	+0.003	+0.003	-0.002	-0.494	+0.001	+0.000	-0.004	-0.428
4	+0.004	+0.003	-0.002	-0.470	+0.000	-0.000	-0.005	-0.358
5	+0.004	+0.003	-0.003	-0.428	+0.000	+0.000	-0.006	-0.343
6	+0.004	+0.002	-0.003	-0.359	-0.000	-0.000	-0.007	-0.265
7	+0.003	+0.002	-0.003	-0.287	-0.001	-0.001	-0.009	-0.199
8	+0.003	+0.002	-0.002	-0.217	-0.002	-0.002	-0.010	-0.116

Note: The table shows the root mean squared error (RMSE), the mean absolute error (MAE), the mean error (ME) and the correlation (Corr.) comparing price changes forecasted by the benchmark model using an expanding window between January 2017 through July 2023, with the transaction index estimated using the full sample (**Index A**). The mean error is defined as the mean of the change in the index estimated using the full sample (**Index A**) minus the forecasted price change. The benchmark price change forecast is the average monthly price change of the transaction index of the last observed year multiplied by the length of the forecast horizon.

Table A.II: Real-time forecast performance of the adjusted listing index using different methods for extrapolating the time-until-sale, probability of sale, and sales-to listing-price ratio

Horizon (months)	Houses				Apartments			
	RMSE	MAE	ME	Corr.	RMSE	MAE	ME	Corr.
A.) TOM, POS & S/L price ratio stay constant (main result Table II)								
1	0.003	0.002	0.000	0.745	0.004	0.003	0.001	0.367
2	0.005	0.004	0.001	0.790	0.005	0.004	0.002	0.512
3	0.007	0.005	0.002	0.787	0.007	0.006	0.004	0.555
4	0.008	0.007	0.002	0.777	0.009	0.007	0.004	0.556
5	0.010	0.008	0.002	0.743	0.010	0.008	0.005	0.561
6	0.012	0.010	0.001	0.679	0.012	0.010	0.007	0.483
7	0.015	0.012	0.001	0.610	0.014	0.011	0.008	0.427
8	0.017	0.014	0.000	0.540	0.015	0.012	0.009	0.384
B.) Average TOM, POS & S/L price ratio of last observed year								
1	0.003	0.003	0.001	0.694	0.004	0.003	0.001	0.405
2	0.006	0.005	0.001	0.725	0.005	0.004	0.003	0.576
3	0.008	0.007	0.002	0.703	0.007	0.006	0.004	0.609
4	0.009	0.008	0.002	0.718	0.008	0.007	0.005	0.646
5	0.011	0.009	0.002	0.687	0.010	0.008	0.006	0.634
6	0.013	0.011	0.002	0.626	0.012	0.009	0.007	0.551
7	0.016	0.012	0.002	0.549	0.013	0.011	0.008	0.492
8	0.018	0.014	0.001	0.457	0.015	0.012	0.010	0.436
C.) Trend in TOM, POS & S/L price ratio of last observed year continues								
1	0.003	0.002	0.000	0.741	0.004	0.003	0.001	0.353
2	0.005	0.004	0.001	0.784	0.005	0.004	0.002	0.499
3	0.007	0.005	0.001	0.778	0.007	0.006	0.003	0.528
4	0.008	0.006	0.001	0.765	0.008	0.007	0.004	0.517
5	0.010	0.008	0.000	0.725	0.010	0.008	0.004	0.510
6	0.013	0.010	0.000	0.660	0.011	0.009	0.005	0.433
7	0.015	0.012	-0.001	0.602	0.013	0.011	0.006	0.398
8	0.018	0.014	-0.003	0.553	0.015	0.012	0.007	0.373

Note: The table shows the root mean squared error (RMSE), the mean absolute error (MAE), the mean error (ME) and the correlation (Corr.) comparing price changes forecasted by the adjusted listing index using an expanding window between January 2017 through July 2023 (**Index E**), with the corresponding change in the transaction index estimated using the full sample (**Index A**). The mean error is defined as the mean of the change in the index estimated using the full sample (**Index A**) minus the forecasted price change (**Index E**). Panel A, B and C concern the the different extrapolation methods used to estimate the time-until-sale, probability of sale, and sales-to-listing price ratio, which are used to estimate the adjusted listing indexes.