

Thiel, Patrick

Research Report

FDZ data description: Regional real estate price indices for Germany (RWI-GEO-REDX) - Version 13: 2008-11/2023

RWI Datenbeschreibung

Provided in Cooperation with:

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Thiel, Patrick (2024) : FDZ data description: Regional real estate price indices for Germany (RWI-GEO-REDX) - Version 13: 2008-11/2023, RWI Datenbeschreibung, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen

This Version is available at:

<https://hdl.handle.net/10419/301663>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

**FDZ Data description: Regional Real
Estate Price Indices for Germany
(RWI-GEO-REDX) - Version 13: 2008-11/2023**

June 2024

Patrick Thiel



Impressum

Herausgeber:

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstraße 1-3 | 45128 Essen, Germany

Postanschrift:

Postfach 10 30 54 | 45030 Essen, Germany

Fon: +49 201-81 49-0 | E-Mail: rwi@rwi-essen.de
www.rwi-essen.de

Vorstand

Prof. Dr. Dr. h. c. Christoph M. Schmidt (Präsident)

Prof. Dr. Thomas K. Bauer (Vizepräsident)

Dr. Stefan Rumpf (Administrativer Vorstand)

© RWI 2024

Der Nachdruck, auch auszugsweise, ist nur mit Genehmigung des RWI gestattet.

RWI Datenbeschreibung

Schriftleitung: Prof. Dr. Dr. h. c. Christoph M. Schmidt

Gestaltung: Magdalena Franke, Claudia Lohkamp

FDZ Data description: Regional Real Estate Price Indices for Germany
(RWI-GEO-REDX) - Version 13: 2008-11/2023

June 2024

Patrick Thiel

Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

FDZ Data description: Regional Real Estate Price Indices for Germany (RWI-GEO-REDX) – Version 13: 2008-11/2023

June 2024

Patrick Thiel

Contents

List of Figures	5
List of Tables	5
Abstract	6
1 Introduction	7
2 Data	9
3 Methodology	10
3.1 Estimations of Regional Indices	10
3.2 Aggregation to Higher Regional Levels	11
3.3 Combination of Individual Housing Type Indices	11
4 Descriptive Evidence	13
5 Changes Between Versions	15
6 Data Access	16
7 References	17
8 Appendix	18
8.1 Information on data restriction in the RWI-GEO-REDX Price Indices	18
8.2 List of Included Independent Variables	19

List of Figures

1	Time Development of the German Price Indices for House Sales, Apartment Rentals, Apartment Sales and Combined Index	Change in percentage points to base year	
	2008	13
2	Regional Pattern of the Price Index for Houses in 2023	Level Values and Change	
	Values Relative to 2008 at the County Level	14

List of Tables

A1	Explanatory Variables in Analysis of House, Apartment and Rent Prices	19
-----------	--	-------	----

Abstract

The FDZ Ruhr at RWI offers price indices for apartments and houses (rentals and sales) in Germany, commencing in 2008. The price indices are derived from the RWI-GEO-RED real estate dataset, which integrates real estate listings from ImmoScout24. Three types of hedonic regressions are performed to allow for comparisons across time (annual or quarterly), across regions (1 x 1 km, municipality, and district), and in combination across time and regions. Furthermore, starting with V13, we also offer a combined index, whereby all three data sources (apartment rents and sales, and house sales) are combined into a single index. The data are publicly available (PUF) and can be requested from the FDZ Ruhr website. A weakly anonymized version (SUF) is also available for researchers upon request.

This data report constitutes an updated version of previous reports and refers to RWI-GEO-REDX v13, which encompasses data up to November 2023.

1. Introduction

One of the most pressing issues currently being debated in the German public and political spheres is the evolution of housing prices. The issue of rising rents, the shortage of housing in urban areas, and the phenomenon of rural-urban migration are among the topics currently under discussion. Despite the topicality of the subject, there is a paucity of data on the recent development of house prices in Germany. To address this gap in the literature, the RWI-GEO-REDX dataset quantifies regional differences in house sales, apartment rents, and apartment sales at the 1 x 1 km grid level and aggregates at higher regional delineations.

A variety of techniques are employed to derive real estate indices. Firstly, median sales prices are the most straightforward to construct and are employed by the U.S. Census Bureau. However, these indices do not adjust for the quality of the properties on the market (Ghysels et al., 2013).

Case and Shiller (1989) propose a repeated-sales price index, which is a constant-quality index that uses only information on houses that have been sold at least twice during the sample period. This index requires very little data. However, these properties may not be representative of the market (Clapp & Giaccotto, 1992). Furthermore, the index relies on the crucial assumption that quality does not change due to renovation or modernization. The U.S. Federal Housing Finance Agency employs a repeated-sales methodology for the HPI index.

Third, hedonic price indices are based on hedonic price regressions, which express the price as a function of the characteristics of the dwelling or house. The characteristics thus serve to describe the quality of the property. For all three types of indices, it is crucial to establish a correlation between the sale of a property and local economic conditions. This is because more expensive homes tend to be put on the market during periods of economic expansion. As noted in Gatzlaff and Haurin (1998), it is important to consider the selectivity of using only sold houses. In a further contribution to this field, Englund et al. (1998) combined the hedonic index with the repeated sales index. Finally, there are stock market-based indices that are obtained by trading real estate investment trust shares.

While a simple study of regional average rents and prices is not sufficient to make reliable statements on housing prices, the RWI-GEO-REDX is based on hedonic price regressions, which provide a more robust analysis. The hedonic price regressions incorporate qualitative differences in housing supply, such as varying amenities. Hedonic price regressions are a commonly employed methodology for real estate pricing in Germany. For instance, the hedonic EPX of Eurospace AG (2019), a commercial online distributor of real estate financing, and the real estate price index of Verband Deutscher Pfandbriefbanken e V. (vdp) (2019), which processes transaction data from certain financial institutions, utilize this approach. The real estate price index by bulwiengesa AG (2018) focuses on cities and larger towns, but does not include consistent information on data structure and methodology and is available for commercial use only. The AK-OGA, a collaboration of all German federal advisory committees on real estate matters, publishes housing price indices for various housing types on an annual basis and at the federal and subdistrict levels. The price indices are derived in part by means of median sales observation and hedonic price regressions. Furthermore, the IMX Offer Index is a hedonic price index derived from the offers placed by the online real estate agency ImmobilienScout24 for the sale and rental of houses or apartments.

The initial two indices present house price changes at the national level, with no further information provided for smaller regions. This study addresses this gap by combining a comprehensive, up-to-date dataset with a hedonic price regression. The resulting regional price indices relative to the German mean provide insight into regional differences, the region-specific time trend, and the national development over time. The analysis is based on a dataset of online real estate advertisements, the RWI-GEO-RED. The comprehensive dataset was obtained from the real estate

portal ImmobilienScout24. The data is updated on a regular basis, allowing for the analysis of the most recent developments. Consequently, it is feasible to update the price indices on a regular basis.

A hedonic price regression is employed to derive price indices, which permit the capture of a range of characteristics of sales and rental prices that extend beyond the median. It is possible that the repeat sales approach may not be representative of the German real estate market, as Voigtländer (2012) notes.

In contrast to other published price indices, which are constructed using financial data, the following analysis deals with seller-listed listings. This implies that only apartments and houses that are available for rent or sale and listed on the listing website are included in the dataset. Consequently, the analysis is based on advertised prices rather than transaction prices.

Furthermore, indices for home sales are incorporated into the dataset, thereby providing additional insight into regional price information for home sales and rentals. Furthermore, a combined index is presented, which aggregates data from all three housing types into a single index. The indices can be retrieved at the following levels: grid (1 x 1 km), municipality, and district.

2. Data

We use the RWI-GEO-RED data (RWI - Leibniz Institute for Economic Research, [2023a](#), [2023b](#), [2023c](#)) of the FDZ Ruhr at the RWI to generate the price indices. The data are derived from real estate listings published on Germany's largest real estate listing website, ImmobilienScout24. Here, property owners and real estate agents may advertise their houses and apartments for a fee. All information is provided by the owner or agent selling or renting the property. While some information is required to place an advertisement online, the majority of information is provided voluntarily. The provision of additional information about the property is likely to facilitate an appropriate presentation and increase the probability of a sale or rental.

The dataset encompasses data on the prices of property listings and various characteristics that determine the value of a property. The data utilized in this data product are monthly in nature. The present dataset encompasses the period from January 2007 to December 2023. Due to the limited number of observations in 2007, the subsequent data analysis is restricted to the years 2008 to November 2023. The restriction to November 2023 is intended to avoid a look-ahead bias in the time trend.

The RWI-GEO-RED provides information on the building at the level of the housing unit. Furthermore, the data encompasses information regarding the dimensions of the residence (e.g., living space, plot area, number of rooms), its amenities (e.g., balcony, garden, bathrooms, level of facilities), financial aspects (e.g., price and additional costs), and energy consumption. Unfortunately, some variables are characterized by a considerable number of missing values, which must be taken into account in the subsequent analysis.

The variables selected for analysis were chosen based on two considerations. Firstly, it is essential to ensure coherence within the data set in order to guarantee comparability between properties. Secondly, the set of characteristics employed in the analysis must be sufficiently comprehensive to encompass the diverse characteristics of apartments and houses. In order to approximate the market price of the property as closely as possible, advertisements are included only in the most recent month of publication, that is, the month in which they are removed from the listing website. Previously updated versions of the listed apartment or house are not included in the analysis. This strategy is designed to provide the closest approximation possible to the actual sale price by comparing the published, self-reported listing price with the actual sale price. Further details on the excluded homes from the original dataset can be found in the Section [8](#).

During the aggregation process (see Section [3.2](#)), we calculate price indices for districts and municipalities based on the regional definitions of 2019 (Federal Agency for Cartography and Geodesy, [2019](#)). In 2019, Germany was comprised of 401 counties, each exhibiting a unique combination of area and population size. Furthermore, there were about 11,000 municipalities.

3. Methodology

3.1. Estimations of Regional Indices

It is important to consider the specific characteristics of the property in question when calculating regional price indices, as well as the regional and temporal differences that may affect the results. The regression is comparable to common hedonic price regressions (e.g., Sirmans et al., 2005) and to such as those applied for Germany in Bauer et al. (2011).

The listing price per square meter is employed as the dependent variable for the housing unit for sale, while the exclusive rent per square meter ("Kaltmiete") is utilized for rentals. Consequently, the endogenous variable is not dependent on the size of the apartment or house. Other characteristics that influence the rental or listing price are incorporated into the hedonic price regression as exogenous variables. In order to control for unobserved heterogeneity, all independent variables are demeaned by the national mean in the estimation process.

All models are estimated at the grid level, and the results are aggregated to larger regional delineations if necessary. Additionally, we provide a combined index that integrates the individual housing types at the grid level. Furthermore, the combined index is aggregated to higher regional levels if needed. The aggregation and combination processes are described in greater detail below.

The objective of the indices is to illustrate the temporal evolution of housing prices in Germany, the regional disparities therein, and the temporal evolution of these regional differences. In order to estimate these three types of indices, different models are required. First, we employ the following model to estimate price indices for the entirety of Germany over time (year or year-quarter):

$$\ln(y_{igt}) = \beta X_{igt} + u_g + \delta_t + \epsilon_{igt}, \quad (1)$$

where the dependent variable y is the sale or rental price per square meter of dwelling unit i in grid g in year t or year-quarter t . The characteristics of the property are included in vector X . The characteristics in the hedonic regressions vary between rental, owner-occupied, and home ownership; a list of the characteristics used as exogenous variables is given in Table A1 in Section 8.

The initial model incorporates a time-independent price index, u_g , for grid g (fixed effects). Furthermore, year or year-quarter fixed effects, denoted by δ_t , are included. The time evolution in Germany is described by the evolution of fixed effects (δ_t) for each year $t = 2008, \dots, 2023$, respectively each year-quarter $t = 2008\text{-Q1}, 2008\text{-Q2}, \dots, 2023\text{-Q2}$. Given that all regions and years are considered collectively, it is assumed that the characteristics are valued in a uniform manner across all years and regions. The error term ϵ_{igt} is assumed to have a standard normal distribution.

The second regression employs a yearly cross-sectional approach, wherein a regional price index u_g for grid g and a specific year or year-quarter t_0 are utilized:

$$\ln(y_{igt_0}) = \gamma X_{igt_0} + u_{gt_0} + \epsilon_{igt_0} \quad (2)$$

In this approach, it is assumed that the characteristics are valued in the same way throughout Germany at the respective time t_0 and in the respective grid. The index is calculated annually from 2008 to 2023. The price index u indicates the price differences between the grids at time t . The indices derived from this regression describe the regional price discrepancy to the German average for all properties offered at a specific time. More specifically: What is the regional premium for a square meter of living space for a property of the same quality and features relative to all listed properties in the respective year?

The initial approach offers insights into the overall development over time, whereas the second approach provides insights into year-specific regional differences. In addition to absolute price differences, regional differences over time are also of interest for research and policy discussion. Consequently, a third approach is employed to assess the respective temporal evolution of grid g via the year-grid fixed effects, denoted by η .

$$\ln(y_{igt}) = \theta X_{igt} + \eta_{gt} + \epsilon_{igt} \quad (3)$$

The specific time evolution of region g between year t_0 and t_1 can be derived by

$$\eta_{gt_1} - \eta_{gt_0} \quad (4)$$

The primary assumptions underlying this analysis are analogous to those presented in Equation (2), namely that the characteristics are perceived to be of equal value in each region and across time. Furthermore, the temporal evolution may vary between regions. This approach is employed for all three types of regions on an annual basis, with the objective of comparing the development observed in each region relative to the base year of 2008 or the year-quarter 2008-Q1.

All regression estimates are de-logged for the output and expressed in percent.

3.2. Aggregation to Higher Regional Levels

In addition to the grid level price indices, we also provide price indices at higher regional aggregations (districts and municipalities) for the regressions presented in Equation (2) and Equation (3). To obtain aggregated estimates, we take the output at grid level and weight each estimate by:

$$W_{grt} = \frac{NOBS_{gt}}{NOBS_{rt}} \quad (5)$$

The year-specific weight is determined by calculating the ratio of the number of observations (NOBS) in grid g to the NOBS in region r (district or municipality). This results in a weighted index per grid cell. Subsequently, the individual grid cells are aggregated for a given region r .

$$Weighted\ Pindex_{rt} = \sum_r^R W_{grt} \times Pindex_{gt} \quad (6)$$

For the change at the regional level (Equation (3)), we calculate the deviation between the base period and each subsequent time period after aggregation in order to obtain the aggregated change in the price/ rent index for region r .

This procedure is specific to the type of housing in question, and thus is performed for each housing type (apartment sales and rents, and house sales) separately.

3.3. Combination of Individual Housing Type Indices

In order to create a single index that incorporates the various housing types, a type-specific weight is calculated.

$$W_{ht} = \frac{NOBS_{ht}}{NOBS_t} \quad (7)$$

The weight is defined as the number of observations (NOBS) for housing type h at time t divided by the total NOBS at time t . Given this weight, the weighted price index at time t is calculated.

$$Pindex_t = \sum_{h,t}^{H,T} W_{ht} \times Pindex_{ht} \quad (8)$$

The aggregation of the combined index to higher regional levels follows the strategy outlined in Section [3.2](#).

4. Descriptive Evidence

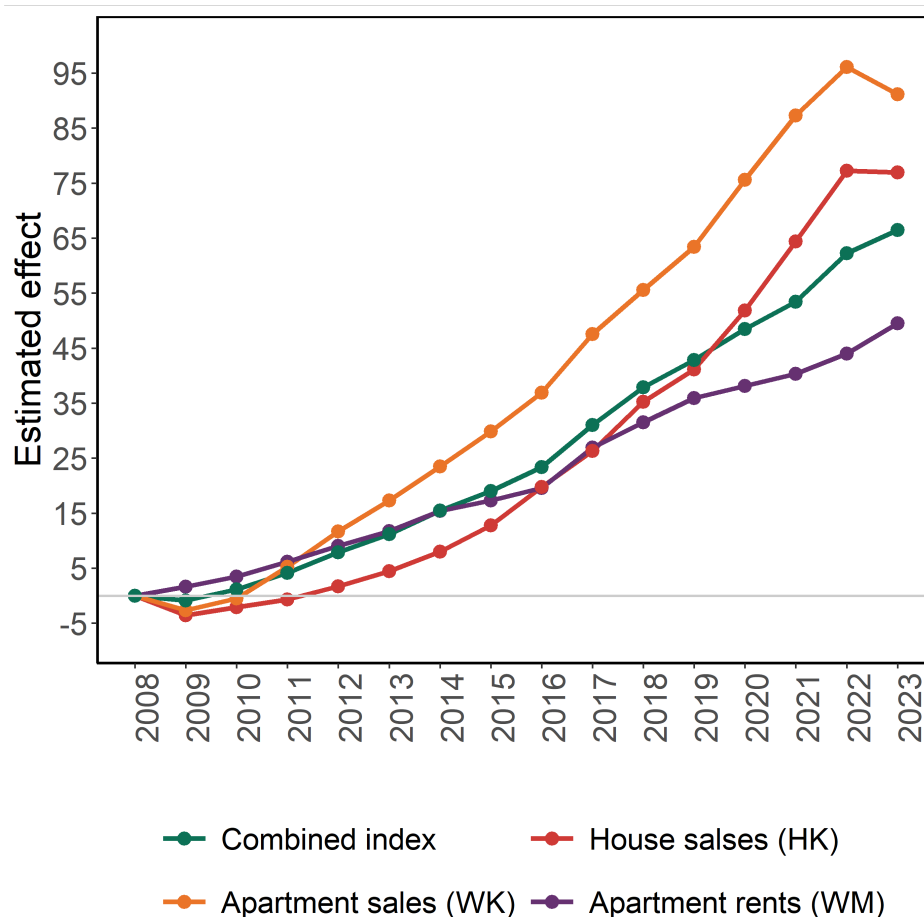
Figure 1 shows the time evolution of the German price indices derived from Equation (1). House and apartment prices declined or stagnated at the beginning of the period. While apartment rents rose steadily, apartment sales began to rise in 2011 and house sales followed suit in 2014. All three types then experienced strong price increases. The ever-increasing trend for sales (both apartments and houses) was only interrupted in 2022. Rising inflation and interest rates, the war in Ukraine, and a lack of investment in the construction sector led to a shortage of housing, making it difficult for people to finance new homes. Apartment rents, on the other hand, remain on an upward trend.

The figure also shows the combined index, which is a combination of the individual indices as described in Section 3. It follows the evolution of the other indices and ranks between them in terms of the estimated effect, which is constructed.

Figure 1

Time Development of the German Price Indices for House Sales, Apartment Rentals, Apartment Sales and Combined Index

Change in percentage points to base year 2008



Notes: The time effects are measured in changes in δ_t in percentage points.

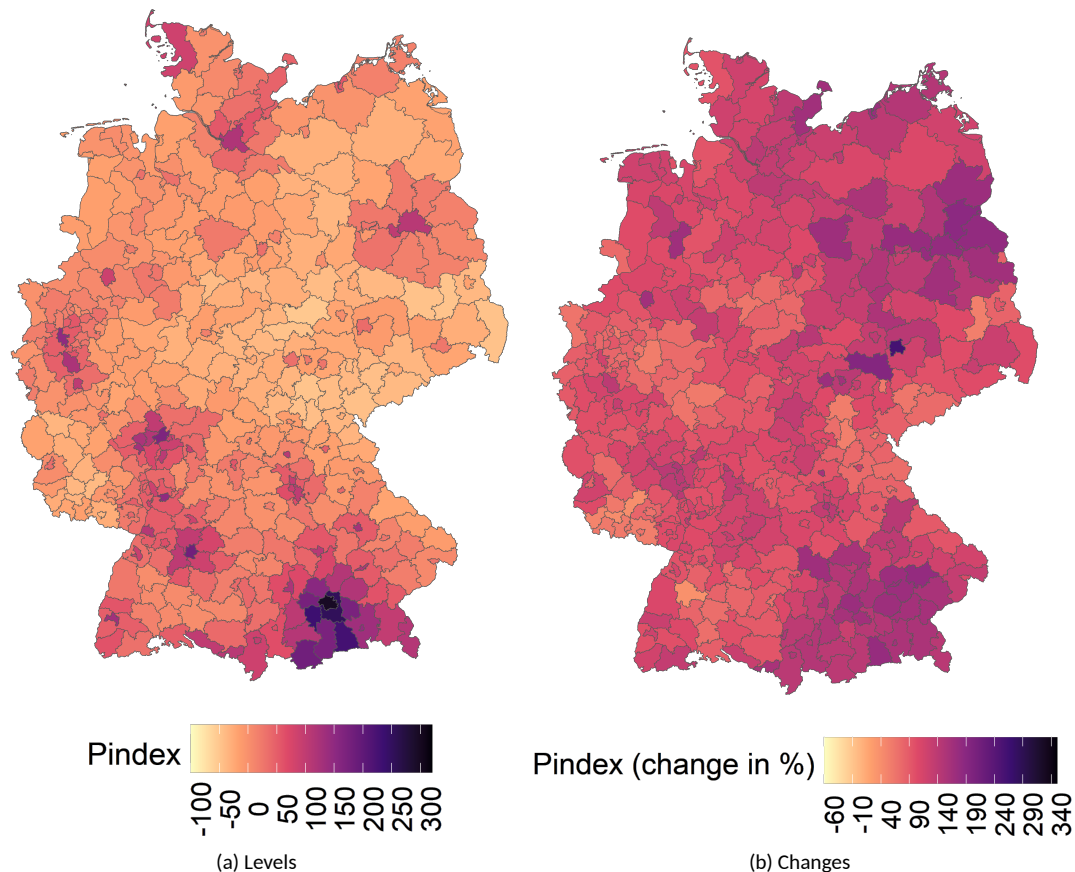
Source: Authors' graph based on RWI-GEO-REDX (RWI - Leibniz Institute for Economic Research, 2024).

To illustrate the possibilities offered by the RWI-GEO-REDX dataset, a few examples of descriptive analyses are given below. The same can be done for the indices for rental housing and apartment sales. We limit our results here to home sales. These have a wide coverage, even in more rural areas, while rental and apartment ownership tend to be concentrated in urban areas.

Figure 2

Regional Pattern of the Price Index for Houses in 2023

Level Values and Change Values Relative to 2008 at the County Level



Notes: The figure shows the geographical distribution of the price index for house sales in 2023 represented as levels (Panel A) and changes relative to 2008 (Panel B) at the county level.

Source: Authors' graph based on RWI-GEO-REDX (RWI - Leibniz Institute for Economic Research, [2024](#)).

Figure [2a](#), which shows the regional price index for home sales (levels and changes) at the county level in 2023. The areas with the highest prices are located in or around major cities, which makes sense as these locations are typically hot spots for working and living. It also reflects the tight housing market with limited housing supply in urban areas.

Figure [2b](#), which reflects the change in house prices between 2008 and 2023, shows that house prices have risen sharply in most regions over time. Areas with particularly strong increases are again cities such as Berlin and Munich.

5. Changes Between Versions

This section provides a concise overview of the modifications incorporated between the current version (V13) and the previous version (V12).

A fundamental shift was made in the estimation strategy, with a transition to an approach based on grids. As detailed in Section 3, the estimation strategy is applied at the grid level for all three regression equations. The estimated effects are then aggregated to higher regional delineations by weighting the effects based on corresponding observations. This approach allows for considerable flexibility in the production of aggregate values at various regional levels, while maintaining the detailed modelling that is necessary.

Furthermore, we introduced the combined regional index with V13. This index combines the individual housing types (apartments for rent and sale, and houses for sale) at the grid level, thereby providing a comprehensive snapshot of the local housing market that is independent of the housing type. Furthermore, the aggregation of the combined index to other regional entities is provided.

The variable "assisted living" was removed from the list of independent variables for the estimation of houses for sale. ImmoScout24 ceased the utilization of the variable in 2016, resulting in a negligible proportion of observations exhibiting a value for this variable.

6. Data Access

The data can be obtained as a Public Use File (PUF) or a Scientific Use File (SUF) from the FDZ Ruhr at RWI. The FDZ Ruhr is the research data center of the RWI - Leibniz Institute for Economic Research. In order to ensure that the indices are not driven by small sample sizes, the PUF dataset includes only those indices that are based on at least 50 observations per year and region. The indices based on less than 50 observations per year and region are also available on request as an SUF for scientific research purposes only. It is up to the researcher to decide whether this lower threshold is reliable in the considered research design. Since the RWI-GEO-REDX subsumes aggregated information, it does not contain any information that is restricted for data security reasons. The indices presented here are available as Excel (.xlsx) files.

Data access does not require a data use agreement, but users must register to access the data. Interested users should visit our website <https://www.rwi-essen.de/en/research-advice/further/research-data-center-ruhr-fdz/data-access>.

It is not permitted for users to undertake any action that might result in the re-identification of individual homes or apartments.

Users are requested to cite the source correctly and to inform the FDZ Ruhr about publications using the data. When using the dataset RWI-GEO-REDX, please cite the data as RWI (2023): RWI-GEO-REDX: Regional Real Estate Price Index for Germany, 2008-11/2023. Version: 1. RWI - Leibniz Institute for Economic Research. Dataset. <http://doi.org/10.7807/immo:redx:v13>. Further, we recommend citing this data description.

Users may also want to cite the data generating process: <https://doi.org/10.5281/zenodo.12166545>.

7. References

- Bauer, T. K., Micheli, M., Schmidt, T., Kiefer, M., Wilke, L.-H., Feuerschütte, S., et al. (2011). *Ein hedonischer Immobilienpreisindex auf Basis von Internetdaten 2007-2011* (tech. rep.). RWI Projektberichte.
- bulwiengesa AG. (2018). Bulwiengesa-Immobilienindex 1975-2017. https://www.bulwiengesa.de/sites/default/files/immobilienindex_2018.pdf. [Accessed: November 2019].
- Case, K. E., & Shiller, R. J. (1989). The Efficiency Of The Market For Single-Family Homes. *The American Economic Review*, 79(1), 125.
- Clapp, J. M., & Giaccotto, C. (1992). Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods. *The Journal of Real Estate Finance and Economics*, 5, 357–374.
- Der Obere Gutachterausschuss für Grundstückswerte im Land Nordrhein-Westfalen. (2017). Grundstücksmarktbericht 2017 Nordrhein-Westfalen. https://www.boris.nrw.de/borischdate/n/gmb/2017/GMB_000_2017_pflichtig.pdf. [Accessed: November 2019].
- Englund, P., Quigley, J. M., & Redfearn, C. L. (1998). Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices. *Journal of Urban Economics*, 44(2), 171–196.
- Eurospace AG. (2019). Index EPX Hedonic. <https://www.europace.de/index-epx-hedonic/>. [Accessed: November 2019].
- Federal Agency for Cartography and Geodesy. (2019). Verwaltungsgebiete (2019) 1:25000 - VG250 und VG250-EW. <https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/verwaltungsgebiete.html>.
- Gatzlaff, D. H., & Haurin, D. R. (1998). Sample Selection and Biases in Local House Value Indices. *Journal of Urban Economics*, 43(2), 199–222.
- Ghysels, E., Plazzi, A., Valkanov, R., & Torous, W. (2013). Forecasting Real Estate Prices. *Handbook of Economic Forecasting*, 2, 509–580.
- Gutachterausschuss für Grundstückswerte in Sachsen-Anhalt. (2017). Grundstücksmarktbericht Sachsen-Anhalt 2017.
- RWI - Leibniz Institute for Economic Research. (2023a). RWI-GEO-RED: RWI Real Estate Data - Apartments for Sale. <https://doi.org/10.7807/immo:red:wk:v9> [Data set].
- RWI - Leibniz Institute for Economic Research. (2023b). RWI-GEO-RED: RWI Real Estate Data - Apartments for Rent. <https://doi.org/10.7807/immo:red:wm:v9> [Data set].
- RWI - Leibniz Institute for Economic Research. (2023c). RWI-GEO-RED: RWI Real Estate Data - Houses for Sale. <https://doi.org/10.7807/immo:red:hk:v9> [Data set].
- RWI - Leibniz Institute for Economic Research. (2024). RWI-GEO-REDX: Regional Real Estate Price Index for Germany, 2008-11/2023. <https://doi.org/10.7807/immo:redx:v12> [Data set].
- Sirmans, S., Macpherson, D., & Zietz, E. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, 13(1), 1–44.
- Verband Deutscher Pfandbriefbanken e V. (vdp). (2019). INDEX: Der Immobilienpreisindex des vdp. https://www.pfandbrief.de/site/dam/jcr:120e19a8-cfd8-499d-8611-82141b5b0043/vdp_Index-Broschuere_Q3_2019_DE.pdf. [Accessed: November 2019].
- Voigtländer, M. (2012). Real Estate Data Sources in Germany. In T. Just & W. Maennig (Eds.), *Understanding German Real Estate Markets* (pp. 3–17). Springer Berlin Heidelberg.

8. Appendix

8.1. Information on data restriction in the RWI-GEO-REDX Price Indices

In order to estimate the RWI-GEO-REDX price indices, it is necessary to exclude luxury apartments from the original RWI-GEO-RED dataset. Consequently, rental apartments with rents exclusive of utilities in excess of €5,000 per month are excluded. The living area is limited to a range of 15 to 400 square meters, with a maximum of seven rooms per rental apartment. The data set with these restrictions encompasses a considerable range of rent prices and living space, although it excludes the smallest or most exclusive apartments from the portfolio. It seems plausible to suggest that the luxury apartment market is not fully integrated into the average rental market. This makes comparisons over regions across these markets challenging. Furthermore, a minimum living space of 15 square meters was set to include only self-sufficient apartments and to exclude typographical errors.

Similarly, restrictions are placed on the sale of houses. The living area of these houses ranges from 50 to 600 square meters, while the price of the house itself can vary up to 5 million euros. The number of rooms is limited to 15. The objective is to ensure self-sufficiency while preventing the construction of houses that may be defective or overly luxurious. The analysis of house sales focuses on single-family homes, with apartment buildings being excluded. Furthermore, holiday homes are excluded if explicitly declared by the seller, as are houses with more than five floors.

Constraints are similarly imposed on offers for apartment sales. For the purposes of this estimation, apartments with a price in excess of 2 million euros, comprising more than eight rooms and an advertised living area falling below the 1st percentile and above the 99th percentile are excluded.

In the context of online advertising, the user is afforded the autonomy to determine the specific details of the real estate in question that are to be made public. The use of self-declared information in this context often results in the presence of missing values in numerous variables, which must be handled with particular care. For binary variables, a missing value is treated as a zero, indicating that the offer does not meet the specified feature. This appears to be a reasonable assumption, given that the owner or agent is likely to publish the benefits of the real estate in order to attract searchers with certain preferences. Moreover, in some years, numerous characteristics are collected via checkboxes, which renders the distinction between "no" and "no answer" meaningless. Such examples include particularly advantageous characteristics of the property, such as a balcony or guest toilet. In the analysis, missing values are treated as a distinct category for categorical variables. In the metric variable under consideration, the number of rooms, missing values are indicated as "zero rooms."

8.2. List of Included Independent Variables

Table A1

Explanatory Variables in Analysis of House, Apartment and Rent Prices

Variable	Description	House sales (HS)	Apartment rents (AR)	Apartment sales (AS)	Restrictions
Numerical variable					
Number of rooms	Number of rooms in the unit	x	x	x	Restricted to: 15 rooms (HS), 7 rooms (AR), and 8 rooms (AS)
Categorical variable					
Number of total floors	1 := Missing				
	2 := 1-3 Floors				
	3 := 4-5 Floors			x	
	4 := 6-10 Floors				
	5 := More than 10 floors				
Floor number of housing unit	0 := Missing				
	1 := Ground floor (UG)				
	2 := First floor (EG)				
	3 := 2nd to 3rd floor			x	
	4 := 4th to 5th floor				
	5 := 6th to 10th floor				
Endowment	6 := Above 10th floor				
	0 := Missing				
	1 := Simple				
	2 := Normal	x	x	x	
	3 := Sophisticated				
Year of construction	4 := Deluxe				
	1 := Missing				
	2 := Before 1900				
	3 := 1900-1945				
	4 := 1946-1959				
	5 := 1960-1969	x	x	x	
	6 := 1970-1979				
	7 := 1980-1989				
	8 := 1990-1999				
	9 := 2000-2009				
	10 := After 2009				
Plot area (in sqm.)	0 := Missing				
	1 := (0-200]				
	2 := (200-400]				
	3 := (400-600]	x			
	4 := (600-800]				
	5 := (800-1,200]				
	6 := (1,200-2,500]				Restricted to 2,500 sqm. due to possible agricultural use
Binary variable					
First occupancy	1 if new owner or renter move in as first occupancy	x	x	x	
Detached house	1 if house is detached	x			
Semi-detached house	1 if house is semi-detached	x			

Continued on next page

Table A1 – Continued from previous page

Variable	Description	House sales (HS)	Apartment rents (AR)	Apartment sales (AS)	Restrictions
Terraced house	1 if house is a terrace house	x			
Exclusive house	1 if the property is declared as a mansion or castle	x			
Multi-family house	1 if house hosts more than one family	x			
Other house type	1 if house is categorized differently	x			
Balcony	1 if property has a balcony		x	x	
Garden	1 if apartment has access to a private garden		x	x	
Guest toilet	1 if housing unit includes a guest toilet	x	x	x	
Built-in kitchen	1 if housing unit comes with a built-in kitchen		x	x	
Granny flat	1 if the property contains a separate “granny flat” or secondary suite	x			
Basement	1 if cellar room is available		x	x	
Common charge	1 if common charge is declared in offer			x	
Lift	1 if property contains a passenger lift			x	

Notes: In the report on the property market of the federal state of North Rhine-Westphalia (Der Obere Gutachterausschuss für Grundstückswerte im Land Nordrhein-Westfalen, 2017) the referees do not count sales of undeveloped rural plot area under 2,500 sqm. in their statistics of farmland sales. This is also the case in the report on the property market for the scarcely populated state of Saxony-Anhalt (Gutachterausschuss für Grundstückswerte in Sachsen-Anhalt, 2017). This gives a notion that plot areas above the margin of 2,500 sqm. can be of interest for agricultural and not only residential purposes. To focus on house sales for living purposes without further commercial use, only plot areas smaller than 2,500 sqm. are included in the following.

Source: Authors' table based on RWI-GEO-REDX (RWI - Leibniz Institute for Economic Research, 2024).



Leibniz-Institut für
Wirtschaftsforschung

Das RWI wird vom Bund und vom Land
Nordrhein-Westfalen gefördert.

