Special Issue Article

Philipp Breidenbach and Sandra Schaffner*

Real estate data for Germany (RWI-GEO-RED)

https://doi.org/10.1515/ger-2019-0126

Abstract: The development and conditions of the housing market are an important part of economic research, receiving growing attention also in Germany. The RWI-GEO-RED dataset covers data for the German residential housing market from 2007 onwards on 1 km² level and closes the gap for small-scale, nationwide and up-to-date housing data for Germany. The data has a large potential in analyzing the real estate market itself or to evaluate policy interventions through treatments on small regional levels. Prominent examples are the closing of nuclear power plants and the German rent control. The dataset can be used on its own but also has a large potential for combination with different data. The dataset is provided by the FDZ Ruhr at RWI that hosts small-scale regional data for Germany. The data can be obtained for scientific research as scientific use file.

Keywords: Housing prices, Real estate data, Germany

JEL Classification: R31, R21, R10, O18, E30

1 Introduction

Regional and urban disparities become more apparent in many economic fields. This is also true for analyses regarding Germany. While nationwide aggregates of important economic indicators such as unemployment, GDP etc. show stable developments, disparities in the society gain increasing attention. This perception seems to result in the rise of a populist party following the same trend as many other countries. Since one explanation of the perceived disparities can be due to regional disparities, regional components become part of more economic, social and political disciplines.

^{*}Corresponding author: Sandra Schaffner, RWI – Leibniz-Institute for Economic Research, Hohenzollernstrasse 1-3, 45128 Essen, Germany, e-mail: schaffner@rwi-essen.de, ORCID: https://orcid.org/0000-0002-4941-4058

Philipp Breidenbach, RWI – Leibniz-Institute for Economic Research, Hohenzollernstrasse 1-3, 45128 Essen, Germany, e-mail: breidenbach@rwi-essen.de

However, traditional German data do not cover all German regions (e.g. SOEP) or the regional level that can be used is quite rough (e.g. Microcensus, SIAB). The research data center (FDZ) Ruhr at the RWI aims to close the gap by providing data for scientific research that fully covers Germany on a small regional level (1 km²), RWI-GEO. One essential component of the provided data are the German residential real estate data (RWI-GEO-RED). This dataset can be applied in studies covering a wide variety of subjects.

Housing prices have become a prominent subject in the German political discussion which results in policy interventions like the German rent control, Bestellerprinzip (client-pays-agent principle), Berlin rent cap or government grant scheme to support families building homes (Baukindergeld). The German housing market was characterized by small nominal price changes and even decreases in real prices in the early 2000s until the financial crisis in 2009 (e.g. Kholodilin et al., 2018). Afterwards prices started to rise and there is a societal and political discussion on price bubbles in Germany (e. g. Kajuth et al., 2013; Kholodilin et al., 2018). Although, price developments are very heterogeneous regarding German regions and neighborhoods (e.g. Klick and Schaffner, 2019a), evidence on the existence of price bubbles in Germany is mixed (e.g. Kholodilin et al., 2018; an de Meulen and Micheli, 2013).

Additionally, housing prices can serve as one indicator for the quality of a neighborhood. For the US, there exist a large literature on neighborhood effects: The neighborhood where children grow up has long-lasting impact on their outcomes as adult (Chetty et al., 2016; Elliott et al., 2006; Sampson, 2012; Sharkey, 2013; Wilson, 1987). More recent, similar analyses and findings are available for Europe as well (Hedman et al., 2015). These lasting neighborhood effects can be observed regarding economic as well as social and physical outcomes of residents. Neighborhood effects are a product of two factors: Selection and structural effects. The selection effect comprises the finding that families in disadvantaged neighborhoods are selected by worse socioeconomic characteristics like education, income, wealth etc. Further, it can be observed that there are fewer social and institutional resources in these neighborhoods.

Finally, climate change and the energy transition in Germany from nuclear and fossil-fuel energy to renewables are accompanied by many changes for consumers especially regarding heating, electricity, building, but also regarding infrastructure and disamenities of their surroundings. It is of economic and societal interest how consumers react to incentives and to observe willingness to pay and willingness to accept, respectively for aspects of the transition like improved heating, insulation but also overhead cable, windmills, solar panels etc. Some of these can be measured through housing prices (e.g. Bauer et al., 2017).

All these fields need precise information on housing, the location and condition of individual objects as well as information on prices. The FDZ Ruhr provides the unique dataset RWI-GEO-RED data on private housing obtained by ImmobilienScout24. The dataset entails information of real estate offers on prices as well as on various observable characteristics that determine the value of a property. It is provided on a monthly basis starting in January 2007. The most recent version of the regularly updated data is until February 2020. The data are organized into four separate datasets; houses for sale, houses for rent, apartments for sale and apartments for rent.

ImmobilienScout24 is the largest internet platform on real estate offers in Germany. It gives real estate owners and agents the opportunity to advertise their objects for a fee. The fee depends on the type of offeror, duration of advertisement and the object. Short-term advertisements for private offerors have been for free or fees of up to 10 €. Commercials can decide between single advertisements or (different types of) membership. Single offers depend on the region and price of the object. The prices have changed over time and are not published on the website. However, since single short-term advertisements are available (dependent on the region) for 0 to 40 €, we assume that prices for commercial offerors are relatively low in comparison to their expected revenues. The platform is open to private and commercial entities. Yet, the present dataset only includes residential real estate. ImmobilienScout24 has a self-reported market share of about 60 % of all real estate objects offered for sale or rent in Germany (Blog immobiliencommunity¹).

The data bases upon user-generated information. To advertise an object, the owner must fill out a questionnaire asking for different characteristics of the property. This means that the price at which the owner is willing to sell or rent out the object is to be understood as an offering price. The advertised price is not binding, i.e. the data do not comprise transaction prices. The price information is available for almost all advertisements. In addition to the price, advertisers are free to include further object-specific characteristics. This helps to present an object adequately, and ideally, it increases the chance of selling at a favorable price or renting out an apartment respectively.

This unique dataset has many advantages compared to other housing data available for the German real estate market. First, due to the high market share of ImmobilienScout24, the dataset consists of a substantial number of observations. This allows for analyzing small-scale housing markets, such as the investigation of the effect of policy intervention on local house prices (e.g. Bauer et al., 2015) and house price responses to location-specific shocks (e.g. Bauer et al., 2017).

¹ https://immobiliencommunity.de/2016/10/27/immobilienscout24-preise/

Second, the FDZ Ruhr gets the exact location of each advertisement for which it is known by ImmobilienScout24. Since this information is mandatory from mid-2016 onwards it is included in almost all advertisements in recent years. By contrast, data scraped from the internet only covers the information that is published on the website. The precise information on the location allows to refer the data to 1 km² grid level. Therefore, analyses on a very small regional level are possible. This allows a combination with other data on the same regional level such as further data provided by the FDZ Ruhr. Besides the 1 km² grid level, information on municipality and district are also available. The use of the exact geo-coordinate of the advertisements is also possible as an on-site access at RWI in Essen.

Third, further information only available for the provider of the data are accessible. The data give information on the days an advertisement was online, as well as the number of clicks and hits by users.

Finally, the data is updated regularly to the very recent end. This enables researchers to analyze bubbles and interventions up to date. By contrast administrative data for housing in Germany is available with several years between observation and publication.

2 Variables and summary statistics etc.

The dataset covers a variety of variables: Identifier of the object, the time period of the advertisement, variables that cover the object features, general information on the object, information on the area, energy and structure information, price and region of the object. Finally, there is dome meta-information on the advertisement as well as generated technical variables.

The dataset covers the period from January 2007 up to February 2020 whereas it is updated regularly. The dataset is divided into four subsets: apartments for rents, houses for sale, apartments for sale and houses for rent. Figure 1 presents the number of advertisements by year (beginning of the advertisement) and category. Since the share of renters is high in Germany and turnover is higher than for sales, apartments for rent is by far the largest subset – almost half of all advertisements are apartments for rents. By contrast, houses for rent are only a very small share. The number of advertisements is relatively stable over time with a dip between 2016 and 2018 especially for apartments for rent.²

² There are different reasons for the evolution, websites which do not charge the offeror (e.g. ebay kleinanzeigen) are on a rise for housing advertisements and word-to-mouth recommendation getting more effective especially in tight markets (especially for apartment rents).

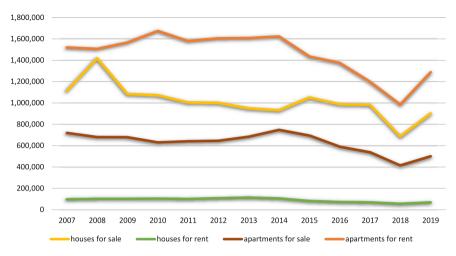


Figure 1: Number of advertisements by year (2007–2019).

Source: RWI-GEO-RED v2; own calculations. Note: Advertisements for 2020 are not illustrated in the figure, as it only contains two months yet.

Table 1 gives an overview on the variables and summary statistics of the sample for apartments for rent. Summary statistics of the other subsets are available upon request and through FDZ Ruhr webpage. Some of the variables are not available for each of the subsets. Further, there are some variables not available for the whole observation period. These variables have missing labels indicating that the variable is not yet or not any longer available. These variables are the number of ancillary rooms (nebenraeume), common charge for community association in EUR/month, heating costs, a variable indicating assisted living for the elderly (betreut), type of the offeror and heating medium. There are some variables with a high share of missings. This is true for some information like the inclusive rent (basic rent plus ancillary costs), heating costs but also year of construction for apartments for rent. By contrast, year of construction for houses (not shown in the table) has a very low share of missing. The data also allow to get some impressions on the popularity (or the demand) for each offered object by indicating the exact number of days an offer is online (laufzeittage) as well as number of clicks and number of hits (indicating how often an object fits to the search criteria of users).

The data also covers the start and end month of each advertisement. Table 2 summarizes the duration of the advertisements. More than half of them is avail-

 $^{{\}bf 3}\ http://www.rwi-essen.de/media/content/pages/forschung-und-beratung/pdf/summary_statistics.pdf$

Table 1: Summary Statistics Apartments for rent.

Variable	Variable description	V	Mean	SD	Min	Max
spell	Spell counter within object identifier	19 066 073	1.45	1.39	1	95
laufzeittage	Days of availability of ad	19 066 073	39.50	109.8	1	4531
hits	Number of hits of ad	19 066 046	6.626	1803	0	1035937
click_schnellkontakte	Number of clicks on contact button	19 065 324	13.50	44.26	0	17380
click_customer	Number of clicks on customer profile	11 019 875	1.753	7.59	0	066
click_weitersagen	Number of clicks on share button	19 066 072	0.65	48.32	0	135213
click_url	Number of clicks on customer URL	19 066 019	3.16	9.19	0	8626
mietekalt	Exclusive rent in EUR	19 066 073	543.1	366.1	0.01	2000
mietewarm	Inclusive rent in EUR	9 368 529	680.7	412.8	0	15000
nebenkosten	Utilities in EUR	17 724 325	140.6	73.29	0	1500
heizkosten	Heating costs	4 7 68 554	73.64	40.73	0	1500
baujahr	Year that object was built	11 619 636	1966	40.83	1000	2020
letzte_modernisierung	Year of last modernisation of object	3 993 629	2010	6.32	1800	2018
wohnflaeche	Living area	19 066 073	72.79	29.78	15	400
nutzflaeche	Usable floor space	4 665 675	50.19	42.29	0	1000
etage	Floor on which object is located	14 455 534	2.10	1.73	0	45
anzahletagen	Number of floors	10875356	3.48	2.04	0	45
zimmeranzahl	Number of rooms	19 066 073	2.62	0.98	0	10
nebenraeume	Number of ancillary rooms	870	0.94	0.51	0	2
schlafzimmer	Number of bedrooms	10,546,249	1.63	0.79	0	66
badezimmer	Number of bathrooms	13 037 909	1.08	0.29	0	5
parkplatzpreis	Price of parking space in EUR	4 866 356	40.89	30.15	0	1000
ev_kennwert	Energy consumption per year and square meter	5 259 317	122.3	53.35	0	009
ajahr	Beginning of ad, year	19 066 073	2012.72	3.67	2007	2020
amonat	Beginning of ad, month	19 066 073	6.25	3.43	1	12

Table 1: (continued)

Variable	Variable description	N	Mean	SD	Min	Мах
jahr	Ending of ad, year	19 066 073	2013	3.67	2007	2020
emonat	Ending of ad, month	19 066 073	6.27	3.43	1	12
aufzug	Elevator in object	19 066 073	0.04	0.19	0	1
balkon	Balcony at object	19 066 073	0.62	0.49	0	1
betreut	Assisted living for the elderly	4 461 893	0.056	0.23	0	1
denkmalobjekt	Protected historic building	1 049 801	0.00	0.04	0	1
einbaukueche	Kitchenette in object	19 066 073	0.35	0.48	0	1
ev_wwenthalten	Warm water consumption included in energy consumption	8 202 947	0.20	0.40	0	1
foerderung		12 482 780	0.04	0.18	0	1
gaestewc	Guest toilet in object	19 066 073	0.13	0.34	0	1
garten	(Shared) garden available	19 066 073	0.18	0.39	0	1
heizkosten_in_wm_enthalten	Heating costs covered by inclusive rent	15 400 858	99.0	0.47	0	1
keller	Cellar in object	19 066 073	0.51	0.50	0	1
parkplatz	Garage/parking space available	2 869 234	0.91	0.29	0	1
rollstuhlgerecht	Accessible, no steps	537 498	0.79	0.41	0	1
ausstattung	Facilities of object	19 066 073	1.05	1.29	0	4
energieeffizienzklasse	Energy Efficiency Rating	652 439	4.47	1.89	1	6
energieausweistyp	Type of Energy Performance Certificates (EPCs)	5 429 388	1.69	0.46	1	2
haustier_erlaubt	Pets allowed	5 120 595	0.70	98.0	0	2
heizungsart	Type of heating	14 550 954	10.85	3.88	1	13
kategorie_Haus	House type	12 654 256	3.39	2.21	1	11
objektzustand	Condition of object	13 017 678	5.21	2.01	1	10

Additional variables not displayed at table: grid, federal state, local labor market district, municipality, zip code, object-ID, Classification of object identifiers used more than once, plot area (for houses only), Brokerage at contract conclusion, Purchasing price in EUR (for sale only), Security deposit, Rented when sold (for sale only), Rental income per month in EUR (for sale only), heating medium, type of offeror.

Table 2: Duration of advertisements (in	months).
--	----------

	N	mean	median	SD
Houses for sale	9 528 050	3.96	2	5.82
Houses for rent	919 863	4.13	2	9.76
Apartments for sale	6 170 673	3.69	2	5.65
Apartments for rent	14 930 421	4.17	2	9.90

Several spells of one dwelling are summarized. Therefore, the number of observations is lower than the overall number.

Table 3: Price per m² (in Euro).

	N	mean	median	SD	Maximum
Houses for sale	12,539,103	1784.64	1590.87	1143.39	100000
Houses for rent	1,177,713	8.10	7.50	3.34	667
Apartments for sale	8,022,837	2316.21	1942.53	2423.55	1652203
Apartments for rent	19,066,073	7.47	6.60	3.34	250

Note: Extreme outliers (possibly due to typos etc. or very special non-representative objects) are excluded. The following exclusion rules are applied:

Houses for sale: number of rooms >15, price >5 Mill., price 0 or missing, living areas <50 m² or >600 m², plot size >2500 m², number of floors >5;

Houses for rent: price 0 or missing, living areas <15 m² or >600 m²;

Apartments for sale: number of rooms >8, price >2 Mill., price 0 or missing, living areas <27 m² or >240 m²:

Apartments for rent: rent >5000, rent 0 or missing, living areas <15 m² or >400 m².

able for only one or two months (median is 2). However, the mean is higher. This is mainly driven by some advertisements that are on the webpage for a very long time. Further, there are some advertisements that are updated by the offeror. Using the underlying original data from ImmobilienScout24 allows to track the updates and to generate a new spell for each update. Therefore, it is possible to observe price changes. Table 3 summarizes the price per m² for each of the subsamples. The mean and median price per m² is higher in houses for rent than in apartments for rent while the difference in the mean price per m² for houses and apartments for sale are quite small. However, median prices for apartments are higher.

Technically, offerors seem to be able to use the same identifier for different apartments. The FDZ Ruhr puts much effort on identifying such non-linked spells with identical identifiers and gives the researcher an additional indicator for cleaning the dataset (see Boelmann and Schaffner, 2019 for a thorough description). Based on this indicator it is possible to construct a dataset that shows only the last or first price during the offering period and further gives the opportunity to observe if an apartment or house is rented out several times. However, since the offeror of objects for sale changes, it is not possible to identify repeated sales.

The 1 km² grid information is anonymized in the scientific use file (SUF) to cover the General Data Protection Regulation of the European Union 2018 for scarcely populated grids. Given the precise geo-information in our data, objects in scarcely populated grids are likewise easy to be reidentified. Aiming to design a solid anonymization strategy, we highly benefit from observing the universe of all houses and households on the 1 km² grid level obtained from the RWI-GEO-GRID dataset (see Breidenbach and Eilers, 2018 for an overview).

Basically, we have two anonymization criteria. First, we censor the gridinformation for apartments or houses located in grids containing less than 5 houses (apartments)⁴ in total. Second, grid information is also censored in slightly more populated grids if a high share of houses/apartments is offered. We censor in grids with less than 10 houses (apartments) if more than half of them are offered within one year.⁵ Given the high probability of each respective object to be in the data, a re-identification would be to easy regarding the precise information in our data.

Given our censoring strategy, no advertisements are deleted due to censoring. Once an object has a risk of re-identification, we delete the information on the 1 km² grid cell. The object itself remains in the data with formation on the municipality and zip-code. In sum, the grid information of 0.67 % of houses for rent, 1.14 %, of houses for sale, 0.16 % of apartments for sale and 0.10 % of apartments for rent are anonymized.

Figure 2 gives an overview on the regional distribution of prices for houses for sale and apartments for rent. The indices are derived by hedonic price regressions. Therefore, the price indices capture differences in the characteristics of dwellings between German districts. There are large differences between regions in Germany. The derived price indices show the deviation from the German mean. If the index is 0, the region is at the German mean (on dwelling level). Housing indices of 100 mean that the price is doubled the German mean. Most districts are below the German mean (yellow or light orange). However, the densely populated areas are mainly above the mean. In both categories Munich and its surroundings are the most expensive area in Germany. The indices are published as an additional dataset, the RWI-GEO-REDX, that is available for public use, too.

⁴ We use the number of households in each grid to approximate the number of apartments (which we cannot observe directly).

⁵ Exceptions are houses that are still under construction.

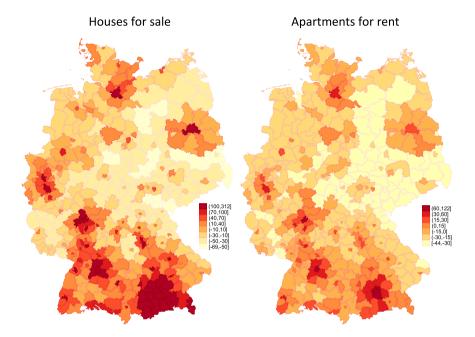


Figure 2: Hedonic Price indices – houses for sale and apartment for rent; 2019 districts.

Source: RWI-GEO-REDX v3 (Klick and Schaffner, 2019b) that is derived by RWI-GEO-RED.

The overall price trend for Germany is displayed in Figure 3. As for the regional price indices, the indices are based on hedonic price regressions. Prices for houses for sale decreased even in nominal terms in the first years observed which can be due to the financial crisis. In the end of 2014 prices reached the level of 2008 again. However, afterwards the positive slope is steeper than before. Prices for apartments for sale increased the most during the observed period. They increased by 50 % within eleven years which is far above the overall price development.

3 Existing evidence and potential of the data

The RWI-GEO-RED data draw their strength by three main factors i) the precise information on geo-locations, ii) the detailed information on the individual conditions of each object and iii) the comparable long history of data (since 2007) without time gaps in between which allow to obtain regionalized price indices. The following section gives an overview on different applications of the data in existing analyses making use of these different assets of the data.



Figure 3: Hedonic Price indices - yearly; reference year 2008.

Source: RWI-GEO-REDX v3 that is derived by RWI-GEO-RED.

Many of the conducted studies, analyzing the RWI-GEO-RED data make use of the precise geo-information available in the data. One prominent example is Bauer et al. (2017) who use the precise geo- and time-information to estimate house price effects in the surrounding of German nuclear power plants after the Fukushima event. They can show that objects located close to a nuclear power plant suffer from the event and the policy measures right after. House prices decreased in the aftermath, especially when the plants were closed due to the Fukushima event.

The data also allow to evaluate place-based policy such as the Emscher-Umbau, a large ecological reconstruction measure, analyzed in Bauer et al. (2015). Regarding the precise geo-location, the value of specific stationary amenities or disamenities can be evaluated. This is done by Frondel et al. (2019a), who study the effect of wind turbines close to advertised objects and Breidenbach et al. (2019a), who study the effect of airport noise and airport connectivity on rent prices. Both studies make use of the exact geo-coordinate available by on-site access at the RWI in Essen. Frondel et al. (2019a) are able to define the precise distance of an offered object to the next wind turbines, finding negative effects on house prices. Breidenbach et al. (2019b) can merge additional data on airport noise and OpenStreetMap drivetimes calculated by the exact position of each object. These exact drivetimes are also applied in Cohen and Schaffner (2019) who analyze house price effects of a newly built motorway (the A38 between Göttingen and Leipzig).

Regarding research in migration economics, Kürschner Rauck and Kvasnicka (2018) use the RWI-GEO-RED dataset to analyze the effect of the 2015's refugee inflow in Germany on local rent price. The described examples present a bandwidth of research in economics and linked fields that can benefit from the dataset and its exact information on geo-location.

The second strength of the data are the precise information on characteristics of each offered object. Yet, this information is predominantly applied in policy evaluations. Frondel et al. (2019b) evaluate the introduction of a mandatory quality disclosure regime in Germany and its effect on house prices. Breidenbach et al. (2019b) evaluate the rent control introduced in Germany in 2015. They make use of the indication of the objects' conditions which can be obtained for each offer in the data and which defines exemptions from the binding rent control.

As the data cover the full stock of residential objects offered on Immobilien-Scout 24 since 2007, they do not suffer from gaps in the observed time period. This allows to obtain profound price indices on different regional levels. The indices are published in the RWI-GEO-REDX datasets for the years 2008 to 2019 on the level of labor market regions, districts and municipalities (see Klick and Schaffner, 2019a). Additionally, work has been done on indices using small area estimations to overcome problems regarding small sample sizes on the regional level (Eilers, 2017).

4 Potential combination of the data

The 1 km² grid used in the data is defined by the INSPIRE regulation of the European Union, further the zip-codes, administrative municipalities and districts are part of the dataset. Based on this information it is possible to combine the data with all other datasets that cover information on these regional levels. Since the FDZ Ruhr focusses on small-scale regional data, there are also data in their portfolio to combine with.

First, the **RWI-GEO-GRID** dataset offers an opportunity to merge socioeconomic characteristics on the 1 km² grid. In detail this is information on the population by age group, gender and migration background. Further, there is information on household composition. These characteristics can be used to classify the neighborhood of an advertisement by the composition of the residents. Besides information on residents there is also information on the number of houses, households and commercial properties. Besides classification of the neighborhood this information can be used to analyze the share of dwellings on the market – to quantify the housing supply. For the residents there is also information on the sum of purchasing power and to which risk classes of credit fail risk the residents belong to.

These numbers are also valid indicators for the quality of neighborhoods but can also be used to analyze whether housing prices and purchasing power of the residents develop similarly or whether there seem to be some crowding out of the original residents. Finally, there is information on car brands, types and mode of driving of the cars. They can serve as an indicator for the eco-friendliness of the neighborhood. This information can be used as additional control or for heterogeneity if analyses on heating, insulation and energy performance certificates are made with the RWI-GEO-RED data.

The FDZ Ruhr already derived housing price indices on municipality and district level based on the RWI-GEO-RED data. These price indices (RWI-GEO-REDX) are for apartments (for rent and for sale) and houses separately. The price indices are based on hedonic price regressions and updated regularly (see DOI: http://doi.org/10.7807/immo:redx:V3). See Klick and Schaffner (2019a) for more details. These indices can be combined with the original data to rank the observed price in the regional price distribution.

Since there is information on municipality and district all administrative data can be merged to the RWI-GEO-RED data. This can be economic and social information provided by regional data base Germany (https://www.regionalstatistik. de/genesis/online/) of the Federal statistical office or by the Federal Institute for Research on Building, Urban Affairs and Spatial Development through the INKAR database (https://www.inkar.de/). Further, some municipalities provide shapefiles for urban districts. If the shapefiles are accessible, they can be combined with the data and used to combine with data on urban district level provided by Kostat (https://www.staedtestatistik.de/kostat.html?&K=0&F=1) or individual municipalities (e.g. https://opendata.leipzig.de/).

Up to now, the dataset is mainly used to analyze effects of (dis)amenities like noise, parks, windmills or infrastructure. Information on new built streets, airports etc. can be assessed by local authorities, Google maps, Open Street Map, Federal Agency for Cartography and Geodesy (BKG) etc. Since there are 1km² or geocodes it is possible to calculate the distance to points of interest or to combine it with shapefiles for noise etc. Therefore, it is possible to analyze how changes in (dis)amenities are valued in the housing market by renters or buyers. Driving times can be estimated by using Open Street Map or Google Maps. Further, FDZ Ruhr provides driving times between the 1 km² grid cells in Germany which can also be implemented.

All the above described combinations take the RWI-GEO-RED dataset as the main dataset for the respective analysis. However, it is also possible to take the

dataset for calculation of control variables like neighborhood quality, price development – changes in neighborhood, housing turnover etc. It is possible to combine the data with all micro datasets that cover small regional information. One prominent example is the German Socioeconomic Panel (SOEP), further data could be the sociological dataset Pairfam that gives information on families and pairs. Most of the data provided by the IAB at least has information on district level. Therefore, it is also possible to combine labor market data (e.g. SIAB) with the housing data RWI-GEO-RED.

5 Data access

The data are available to researchers for non-commercial use. There are two versions of the RWI-GEO-RED data. First, the Scientific Use Files (SUF) covering all information except the exact geo-coordinates. Further, some grid cells are anonymized due to the anonymization rules in section 2. Second, the full datasets are available in the Data Secure Room of the FDZ Ruhr in Essen (on-site access). The data can be obtained as a Stata® datasets (.dta) or .csv files. Data access to both versions requires a signed data use agreement. Both versions are restricted to non-commercial research and only researchers of scientific institutions are eligible to apply for data access. The SUF may be used at the workplace of the users. Data access is provided by the Research Data Centre Ruhr at the RWI – Leibniz-Institute for Economic Research (FDZ Ruhr am RWI). Data access can be applied for online at http://en.rwi-essen.de/forschung-und-beratung/fdzruhr/datenzugang. The application form includes a brief description and title of the project, potential co-operation, information on the applying department, expected duration of data usage as well as further participants in the project.

Data users shall cite the datasets properly with the respective DOIs. The DOIs of V1 of the Scientific Use Files are: http://doi.org/10.7807/immo:red:hk:suf:v2 for houses for http://doi.org/sales10.7807/immo:red:hm:suf:v2 for houses for rent: http://doi.org/10.7807/immo:red:wk:suf:v2 for apartments for sale http: //doi.org/10.7807/immo:red:wm:suf:v2 for apartments for rent.

6 Description of FDZ Ruhr and data overview

Following the idea to improve data access, the FDZ Ruhr is a research data center which provide data for scientific research. Its service catalogue is based on criteria of the German Data Forum (RatSWD) from which it is accredited since 2010. The

data of the FDZ Ruhr can be distinguished into two strands: RWI-MICRO and RWI-GEO, RWI-MICRO comprises data sets that were collected within research of the RWI. They cover a wide range of subjects. The most famous data are the German Residential Energy Consumption Survey (RWI-GRECS) and the Socio-Ecological Panel (Green-SÖP). These two household panels cover information on energy consumption and attitudes towards climate change measures in Germany.

The FDZ Ruhr am RWI puts much effort on providing small scale information for researcher. Therefore, RWI-GEO is the major strand of dataset which is constantly enriched. RWI-GEO comprises all those datasets of the FDZ Ruhr covering information on a small-scale level. As far as possible, "small-scale" is defined by 1×1 km grid cells. These grid cells allow very granulated analyses and they are standardized by the EU-INSPIRE guideline which assures a constant spatial definition of grids over time (in contrast to administrative border). All data available on the grid level can be merged to the RWI-GEO-RED datasets. The datasets are the Socio-Economic Data on grid level (RWI-GEO-GRID) and the Real Estate Indices (RWI-GEO-REDX) that are described in section 4. Further, there is a Small-Scale Population Projection (RWI-GEO-GRID-POP-FORECAST) on 1 km² level, Gas Prices (RWI GEO-Gas) for each gas station in Germany and driving times that are based on calculations with open street map. As shown in sections 3 and 4 this particularly accounts for socio-economic neighborhood characteristics as well as information on amenities (and disamenities) in the proximity.

References

- an de Meulen, P., and M. Micheli. 2013. "Droht Eine Immobilienpreisblase in Deutschland?" Wirtschaftsdienst 93 (8): 539-544.
- Bauer, T. K., S. T. Braun, and M. Kvasnicka. 2017. "Nuclear Power Plant Closures and Local Housing Values: Evidence from Fukushima and the German Housing Market." Journal of Urban Economics 99:94-106.
- Bauer, T. K., R. Budde, M. Micheli, and U. Neumann. 2015. "Immobilienmarkteffekte des Emscherumbaus?" Raumforschung und Raumordnung 73 (4): 269-283.
- Boelmann, B., and S. Schaffner. 2019. "FDZ Data Description: Real-Estate Data for Germany (RWI-GEO-RED V1) - Advertisements on the Internet Platform ImmobilienScout24 2007-03/2019." RWI Projektberichte.
- Breidenbach, P., and L. Eilers. 2018. "RWI-GEO-GRID: Socio-economic Data on Grid Level." Jahrbücher für Nationalökonomie und Statistik 238 (6): 609-616.
- Breidenbach, P., J. P. Cohen and S. Schaffner. 2019a. "Continuation of Air Services at Berlin-Tegel and its Effects on Rental Prices." Ruhr Economic Papers #822. RWI. https://doi.org/10.4419/86788953.

- Breidenbach, P., L. Eilers, and J. Fries. 2019b. "Rent Control and Rental Prices: High Expectations, High Effectiveness?" Ruhr Economic Papers #804. RWI. https://doi.org/10.4419/86788932.
- Chetty, R., N. Hendren, and L. F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." American Economic Review 106 (4): 855-902.
- Cohen, J., and S. Schaffner, 2019, "A New Highway in Germany and the Impacts on Real Estate Prices." Ruhr Economic Papers #821.
- Eilers, L. 2017. "Is My Rental Price Overestimated? A Small Area Index for Germany." Ruhr Economic Papers #734. RWI. https://doi.org/10.4419/86788854.
- Elliott, D. S., S. Menard, B. Rankin, A. Elliott, W. J. Wilson, and D. Huizinga. 2006. Good Kids from Bad Neighborhoods: Successful Development in Social Context. Cambridge University Press.
- Frondel, M., G. Kussel, S. Sommer, and C. Vance. 2019a. "Local Cost for Global Benefit: The Case of Wind Turbines." Ruhr Economic Papers #791. https://doi.org/10.4419/86788919.
- Frondel, M., A. Gerster, and C. Vance. 2019b. "The Power of Mandatory Quality Disclosure: Evidence from the German Housing Market." Journal of the Association of Environmental and Resource Economists 7 (1): 181-208. https://doi.org/10.1086/705786.
- Hedman, L., D. Manley, M. van Ham, and J. Östh. 2015. "Cumulative Exposure to Disadvantage and the Intergenerational Transmission of Neighbourhood Effects." Journal of Economic Geography 15:195-215.
- Kajuth, F., T. Knetsch, and N. Pinkwart. 2013. Assessing House Prices in Germany: Evidence from an Estimated Stock-flow Model Using Regional Data.
- Kholodilin, K. A., C. Michelsen, and D. Ulbricht. 2018. "Speculative Price Bubbles in Urban Housing Markets." Empirical Economics 55(4): 1957-1983.
- Klick, L., and S. Schaffner. 2019a. "FDZ Data Description: Regional Real Estate Price Indices for Germany (RWI-GEO-REDX) - Version 2: 2008-02/2019." RWI Projektberichte, Essen.
- Klick, Larissa, and Sandra Schaffner. RWI; ImmobilienScout24. 2019b. "RWI-GEO-REDX: Regional Real Estate Price Index for Germany, 2008-02/2020. Version: 3." RWI – Leibniz Institute for Economic Research. Dataset. https://doi.org/10.7807/immo:redx:v4.
- Kürschner Rauck, K., and M. Kvasnicka. 2018. "The 2015 European Refugee Crisis and Residential Housing Rents in Germany (No. 12047)." IZA Discussion Papers.
- Sampson, R. J. 2012. Great American City: Chicago and the Enduring Neighborhood Effect. University of Chicago Press.
- Sharkey, P. 2013. Stuck in Place: Urban Neighborhoods and the End of Progress Toward Racial Equality. University of Chicago Press.
- Wilson, W. 1987. The Truly Disadvantaged. The Inner City, the Underclass, and Public Policy. Chicago|London.