Exploring spatial disparities in residential house prices on a county level in Germany

Abstract. In the segmented *German* residential real estate market, house price levels differ substantially among counties. This thesis explores a variety of drivers that may explain spatial disparities of house prices between German counties. Supply and demand for regional house price levels include different aspects as economic, socio-cultural, infrastructural, real estate, and regulatory drivers. An OLS model is built to investigate the associations with house price levels on a county-level. A distinction is made across space between 16 federal states and 401 *German* counties. Further, I characterize differences across regions where I pay special attention to a variety of drivers, including the urban-rural sprawl. I find that not all drivers have equal dominance levels for regional house price levels. Findings not only reveal that personal income is positively and average age of the inhabitants is negatively associated with the regional house prices, but also that urban and western counties are positively associated with house prices.

Keywords. German house prices, residential price variations, urban-rural divide, spatial disparities

Christoph Klare

Master Thesis Real Estate Studies

Final Version

COLOFON

Title	A cross sectional analysis of German regional house prices and associations
Version	V9
Author	Christoph Klare
Student number	S4115155
Supervisor	Prof. dr. ir. A.J. (Arno) van der Vlist
Assessor	Dr. M. (Mark) van Duijn
E-mail	c.klare@student.rug.nl
Date	Nov 18, 2020
Word count	15635 Words

Disclaimer: “Master theses are preliminary materials to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the author and do not indicate concurrence by the supervisor or research staff.”

Contents

1. Introduction ... 1
1.1 Motivation ... 1
1.2 Literature review ... 2
1.3 Research problem statement ... 4
2. Associations influencing house prices ... 7
2.1 Demand drivers ... 7
2.2 Supply drivers ... 10
2.3 Other external drivers ... 10
2.4 Stock-Flow model ... 11
2.5 Hypotheses ... 13
3. Data & Method ... 14
3.1 Data sources ... 14
3.2 Regional housing markets in Germany ... 15
3.3 Descriptive analysis ... 18
3.4 Real Estate Ratios ... 25
3.5 Model development ... 29
4. Results ... 32
4.1 House price differences across federal states ... 32
4.2 House price differences across counties and urban areas ... 34
4.3 Regional house price drivers ... 36
4.4 Chow Test ... 40
5. Conclusion ... 44
References ... 46
Appendix A (Definitions) ... 50
Appendix B (Descriptive Analysis and Ratios) ... 52
Appendix C (OLS Results and Tests) ... 62
Appendix D (Assumption testing) ... 77
Appendix E (Stata Syntax) ... 83

1. Introduction

1.1 Motivation

House price differences are an often discussed issue in German society. House prices differ substantially within Germany. For example, official German statistics reveal that in 2017 the average house prices in metropolitan areas are substantially higher than the average house price of rural areas in Germany (Destatis, 2020). For housing seekers, the price level is an important determinant for the decision to buy. The question arises as to what drivers are associated with these house price differences.

This paper explores the main associations that drive house prices between counties (Definition in Table appendix A, Table A.1). Interested parties should receive a deeper understanding, why one area is more expensive than the other one. By revealing what associations determine regional prices, for instance the average dwelling size in square meters, housing seekers can balance their preferences and have a deeper understanding of why one county is more expensive than the neighbouring counties. As an illustration, the German newspaper FAZ (for Definition, see appendix A, table A1) reports that municipalities within Germany can have up to 13 times higher square meter prices than other regions (Papon, 2019). Consequently, due to the fact that in recent years a substantial party is interested in buying their own property, a high share of the population is possibly interested in the main associations that drive house prices.

The implications of regional house price differences are far-reaching as it is associated with migration as a driver as well. The increasing trend of migrating to new counties is among others connected to demographic drivers as for example job prospects and better infrastructure for young families are demanded (Bleck & Wagner, 2006). This leads to migration from rural to urban areas as well as from eastern to western regions (Diekmann, 2019). The question may arise, how regional migration balances affect the house price levels and whether other drivers stimulate house prices differently on a regional level. When people choose a new place to live, the demand for homeownership also increases more unequally within Germany, affecting local house prices. In 2018 the average homeownership rate in Germany was at 51.5% with a geographical disparity (Trading Economics, 2020; Zumbro, 2014).

One of the stylized facts of housing markets is that regional house prices may vary considerably. Including national and international investors on the demand side to buy residential real estate, the real estate market has become more unequally distributed and untransparent. Investors typically prefer urban regions because of personal risk and profit structures (Fabricius, 2015). The imperfect German real estate market reasoned the new subsidy program "GRW" (see appendix A, Table A*1)* which started January 1st, 2020 in order to create an equilibrium of economic power including house prices throughout Germany. Therefore, it also supports broadband expansion in rural areas for private persons and other subsidies and infrastructural measures to attract investors (BMWI, 2020). Thus, the question arises what drivers are associated with regional house prices?

1.2 Literature review

This research connects to the literature on regional house price differences. The following subchapter lists the contributions provided by several scientists that examined different house price levels, separated into paragraphs for demand, supply and institutional factors. For further considerations the house prices will represent the perceived housing values as a proxy, yielding a common basis for comparisons for the following literature review.

Similar to the research questions of this paper, Blanco et al. (2015) address the regional house price differences in Spain. More specifically, the authors attempt to find price levels in the different regions of Spain, meaning that in this case groups of regions are on the same house price level. Based on macro-economic factors the authors stress that the house prices differ, as the independent variables affect each group of regions differently, resulting in the categorization of 4 house price groups. For demand factors, Blanco et al. 2015 use economic variables as household income and a rent-to-price ratio, while demographic and social input variables are immigration, population growth and population composition. Apart from the 3 main factors for the demand side, the authors introduce construction costs, housing stock, land availability and climate as supply factors. For this research, the population and housing stock, as well as vacant homes, are considered very convincing determinants for Spain's regional house price differences.

Regional house prices are also linked to other broader factors like the regional labour market. Deschermeier et al. (2016) specify these labour patterns for the German market and clarifies that rural areas, with a shortage of skilled workers, are likely to face decreasing demand for housing which leads to a negative price adaptation when holding the stock of housing constant. Additionally, skilled workers are likely to immigrate from these weak regions to regions with better job expectations (Deschermeier et al., 2016). These associations of agglomeration have been examined by researchers for their effects and have been found to contribute to the attractiveness of a region (Jacobs, 1996). It would therefore not be surprising if the population concentration in urban areas is also accompanied by stronger economic growth. These results go in line with the research by Chen et al. (2007) who conducts the relationship between house prices and income in Taiwan, stressing that an equilibrium relationship between regional house prices and income levels exists. However, findings like those mentioned before seem missing for a current observation of Germany.

Peripheral regions are also currently facing spatial disparities as demand for living and average house price levels are higher in urban areas. However, Röhl (2018) finds that slagging rural house prices need concepts to cope with the major challenges of demographic change, which will affect the economy and living conditions here more strongly and more than in the urban regions. Takáts (2012) confirms these findings in his work about aging and house prices, indicating that in the upcoming years an increasing average age of an economy leads to decreasing house prices. For people willing to move to a place with better life expectancies, Bleck & Wagner (2006) conduct an analysis about the migration behaviour within a country. The authors state that with a migration to a different county, often the type of residential unit changes. This can be implied by migration from a rural area and a single-family home to an urban area and into an apartment. Meaningful work by Rosenthal & Strange (2004) highlight that regional economies attract the workforce to migrate rather than to commute and thus indicating that labour markets positively associate the demand for housing in a region. These agglomerations offer higher individual wages and local amenities, leading to a higher willingness to pay for quality-adjusted house prices (Rosenthal & Strange, 2004).

Local policy decisions and regulations affect regional house prices across space and over time. Vermeulen & Van Ommeren (2009), indicate that land-use planning restrictions strongly affect the economic activities of a region and thereby the wealth of a region. The authors conclude house prices are responsive to the net internal migration, which is correlated with employment growth. The outcome of land use regulation is feasible in construction permissions, although the attractiveness of a region also determines the demand for new dwellings. Therefore, Einig (2003) researches how current building land prices and construction permissions are the consequences of policy measures and change house prices. In conclusion, high building land prices and a low amount of building permissions can thus indicate a high regional demand with a limited supply of new constructions.

The given literature on demand, supply and institutional factors can explain regional house prices to an adequate level, but not all of it. Regions differ in the perceived attractiveness, amenities and other social factors that cannot easily be quantified. Hence, the ‘propensity to own' a home is one indicator that differs among the regions and thus associates with the regional house price level. Mulder (2006) stresses that home-ownership and family formation is closely connected, while people are more likely to settle where the cost of rearing children is low. Generally, every county has a certain perceived attractiveness and thus a constant term for explaining house price levels remains.

1.3 Research problem statement

In general, recent literature provides no insight into drivers of regional house prices in Germany. Studies indicate it for other nations and cross-national comparisons (Takáts, 2012), however these theories are mostly not applicable to the German market. For instance, Blanco et al. (2015) include in their house price analysis for Spain, climate variables for Spain, which are very meaningful for a country with different climate zones, but also because house price groups are determined as the dependent variable in the regressions. Furthermore, the existing literature rather focuses on growth rates and time series perspectives and less on a direct comparison between regions for a given year. For instance, Rosenthal & Strange (2004) compare several concepts on agglomeration economies over the last decades and proved a positive association between agglomeration economies and the corresponding regional house price level. With no insight into the regional house price disparities in Germany, a research gap for the German housing segmentation remains, initializing my contribution to the academic research as an aggregated, quantitative and contemplating view on the existing research. With my cross-sectional approach, more associations can be researched, as the existing literature focuses on the differences between a few associations. While adding my analysis with current and relevant data, an aggregated view on the German regional market is the focused research aim which thus can be defined as the following:

How can regional house price levels be understood on a county-level in 2017?

The scientific relevance of this thesis is provided by ascertaining the driving determinants for regional house price disparities. To achieve the research aim, the aim will be decomposed into 3 research questions:

1. *How are regional house prices determined?*
   For the first research question, most of the aggregated academic literature will be regarded. To measure the disparities, a literature review will set the basis of the research aim, with definitions on urban-rural disparities, associations of urbanization and house price levels. Bleck & Wagner (2006) Rosenthal & Strange (2004) and

Vermeulen & Van Ommeren (2009), will be the driving sources. Furthermore, sale prices of residential objects on a county level are explored on a square meter price in Euro. Expressed as real estate market ratios, regional house price levels offer a national comparison between counties. The ratios are important for understanding market connections and differences across counties and consequently aiming at visualising regional differences. The determination of a rent-to-price ratio will also guide housing seekers through the, for now, non-transparent market. As a matter of fact, some administrative counties, as Hannover and Aachen, will be reallocated into different urban-rural groups. This is necessary because the population size, determining the categorization of a county is sometimes more widely spread. Also, the associations of highway and public transport development programs will be discussed, based on the findings of research question 2.

2. *What is the variation in regional house prices in Germany??*

   The second research question builds upon the first one. By referring to associations, described by Kempermann et al. (2019), Capozza et al. (2002), Meen (2002) and by using a log-linear model, house prices can be dismantled into their associations. Regional fixed effect measures will be derived in a table and interpreted to create hedonic price indices for each regional distinction.

3. *Which regional drivers are associated with regional house price disparities?*

   The last subquestion is dedicated to the categorization of current house price groups based on common associations. Regional drivers are explored with the underlying house price levels found in research question 2. Based on the existing literature, tables and maps present the aggregated associations of research questions 2 and 3. Interaction variables between spatial effects are included to reflect the German real estate market.

After some academic insights are given in chapter 1.2 and the derived research questions, this thesis is further scientifically motivated. It aims to isolate strong associations that lead to regional disparities in house prices. By analysing available macro data and combining existing literature on urbanization, migration behaviour, economics and regulations, possible drivers for regional house price levels are considered. With a focus on spatial and economic drivers, this master thesis tries to contribute to further research as real estate ratios, fixed effects and dominance tests enable future research to compare municipalities. New insights of this thesis can be used for example for further investigations or cross-country analyses or to establish regional house price rankings. The following conceptual model delivers an overview of the concept that will be followed during the regression and based on the findings of previous studies. The simplified baseline assumption is that regional house prices are responsive to external drivers. Firstly, external drivers can be expressed as demand and supply drivers. For now, unknown associations, like institutional, unobservable associations and fixed locational effects, are marked with a dashed line and indicate factors associated with regional house prices. Institutional and other factors as social and historical associations may change regional house prices directly and indirectly, and cannot be related to demand and supply associations directly.

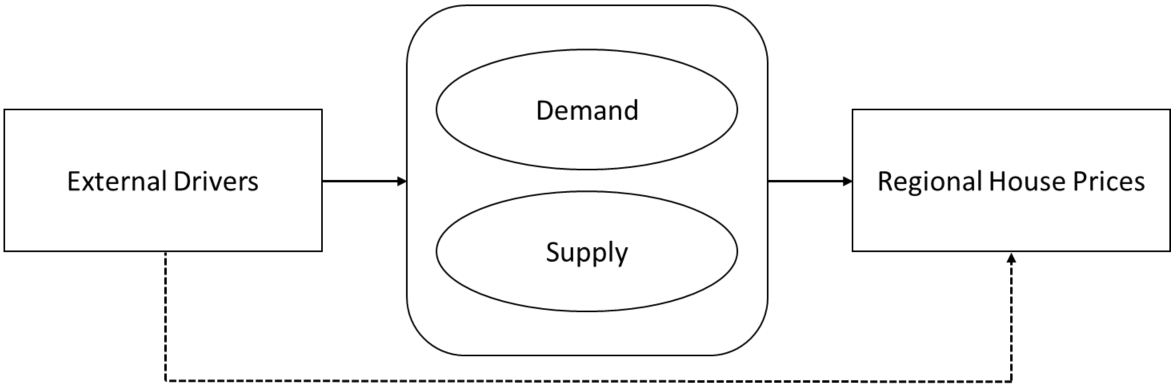


Figure 1: Conceptual model for the determination of regional house prices

This paper is organised into 5 chapters. Chapter 2 firstly examines existing theories to receive an understanding of the market dynamics. Chapter 3 stresses the German regional housing markets. The third chapter presents the data and method for the quantitative part of this thesis and sets hypotheses. The next chapter highlights the findings of the regression and visualizes new insights. Chapter 5 concludes and offers a critical review as well as an outlook for further research.

2. Associations influencing house prices

The following subchapters theorize on regional house price drivers, for two reasons: first, to answer the question of which drivers are associated with regional house price levels in order to formulate a model in Chapter 3. Secondly, to postulate hypotheses.

2.1 Demand drivers

Certainly, one important association for regional house price levels, in all areas, are economic activities. Jacobs (1969) already stresses that urban areas have greater economies compared to economies in less densely populated regions. Thus, in urban areas, economical actors benefit from complementary knowledge exchange and general agglomeration associations. These high economic activities in urban areas demand workforce. Hoogstra et al. (2017, p. 365) state that the inflow to regions with good job prospects is high. In a meta-analysis these authors show that “jobs follow[ing] people”, meaning that a more divergent settlement of companies takes place, driving towards higher regional disparities of economic activities.

Individual economic wealth can be expressed in personal income and certainly associate the spending power and house values. Therefore, research positioned house prices and income as a cointegrated relationship, showing a long-term relationship (Capozza et al., 2002; Meen, 2002). Gallin (2006) however, state that on smaller geographical levels in the USA, the association between income and house prices, expressed as ratios, only show a small variance when compared nationally. However, it does not prove a long term equilibrium if time series data is considered. Other authors found similar associations, e.g. Chen et al. (2007) proved for Taiwan that income accounts for 25% of the house price in Taiwan. Generally, the OECD (2016) states that the demand for owner-occupied housing increases with higher household income.

Introducing a new indicator, the driver income will be adjusted to real income which associates with regional house price levels. Real income includes all income components of a year that relate to a surveyed household as a whole, as well as all individual gross incomes of the people currently surveyed in the household (market income from the sum of capital and earned income including private transfers and private pensions). Besides, income from statutory pensions and pensions as well as social transfers (including social assistance, housing benefit, child benefit, support from the employment office) are taken into account, and finally, with the aid of a simulation of tax and social security contributions, annual net income is calculated - including one-off special payments (13th and 14th monthly salaries, Christmas bonus, vacation bonus, etc.) that are taken into account (Grabka & Goebel, 2018). The authors stress that higher real income leads to fewer poverty risks and responsively a higher willingness to pay for houses is given – absolutely (house price) and relatively (price per square meter). Income as a ratio can further be taken into account to distinguish between regional house price levels. To stress regional disparities Philiponnet & Turrini (2017) build a ratio of income levels on a European country level with the corresponding house prices – similarly for rent price levels. The result is an applicable comparison and a benchmark for European countries, indicating that some regions have substantial differences in affordability as the house prices differ from the benchmark level of the corresponding income level.

With a deeper focus on the economy, it can be stated that smaller economic activities lead to low regional employment levels. This results in less spending power for housing. Mainly this is the case in rural regions, while high employment levels in urban regions result in high house prices. Buch et al. (2014) indicate that labour migration takes place from cities with high unemployment towards cities with low unemployment rates and high wages. Buch et al. (2014) regress labour market indicators on house prices and find that they play a crucial role in migration behaviour. Results show that 44%, 11% and - 32% explaining the net migration rate of the workforce with employment growth, wage level and unemployment growth, respectively. The highlighted results are confirmed by earlier work of Renkow (2003) who analyses commuting behaviour and within county labour market adjustments. Furthermore, an important finding of his results is the fact that neighbouring employment rates and population levels can have a significant positive association on the labour force level of a specific region. Bleck & Wagner (2006) report similar outputs for explaining the associations of migration and the trend of suburbanization. Accordingly, this phenomenon can be explained by the increasing demands for space on the site of the population as well as for companies.

Additionally, migration is accompanied by the demand for a new type of dwelling. An interesting observation is the fact that the regional house prices are higher in counties where people migrate to own a single family or duplex home (Bleck & Wagner, 2006). With surveys, Bleck & Wagner (2006) conduct that migration is motivated based on the old and new house prices and the form of housing. Conversely viewed, table 1 shows the highlighted housing situation of migrants from urban to rural areas. The findings indicate an increase in the propensity to own, which is in line with the family formation aim presented by Mulder (2006).

Table 1: Dwelling type before and after migration

	Pre Migration	Post Migration
Apartment	78.2%	46.2%
House	20.7%	52.2%
Renting	84.3%	59.4%
Owning	12.0%	38.4%

Note: In some studies, there are further distinctions, which have either been disregarded or summarized to the above. As a result, individual categories do not always add up to 100%. Significance is throughout given at a 1% confidence level. Source: own presentation based on Bleck & Wagner (2006)

Therefore, Bleck & Wagner (2006) collected and summarized other questionnaire data sets and presented that migrants are demanding more space in rural areas, by preferably owning it. 25% of those polled justify their choice of migrating to a rural area and owning a house or an apartment with lower house prices. However, explicit housing values were not regarded in this research.

Next, demographic drivers are associated with regional house price levels, for instance the regional average population age structure and their demand for (family and single) space. Thus, Berndgen-Kaiser et al. (2014) state that next to migration patterns the geographical fertility and mortality rates drive the demand and supply for houses, and responsively the house prices. The authors' analysis categorizes house price levels based on the risk of regional vacancy, taking into account that higher average ages of the population are present in more rural areas. Berndgen-Kaiser et al. (2014) stress that approximately 48.49%, or 0.86 million out of 1,77 million observed, of rural dwellings and only 6,6%, or 0,073 million out of 1,1 million observed dwellings are considered as attached with high vacancy risks. Thus, Berndgen-Kaiser et al. (2014) state that the age structures and other demographical levels are positively correlated with regional house prices. The consequences for high risk assets are higher vacancy rates and as a consequence, lower house prices. Especially, the average population age, as a result of out-migration, of a county might be associated with the house price level (Berndgen-Kaiser et al., 2014). The risks of value decreases are lower in urban areas. This leads to higher regional house price levels. This seems coherent as shown by Berndgen-Kaiser et al. (2014). When considering regional working-age households, and their purchasing power, the age structure becomes an important determinant influencing house prices. Takáts (2012) proved for OECD countries that the demography of the economy substantially determines real house price levels. Furthermore, Mankiw and Weil (1989) and DiPasquale and Wheaton (1994) indicate that the share of working-age residents is a main association for the demand in house prices and services, as they require more space than young and old people.

Another way to measure regional consumption power for housing is to compare purchasing powers. While most literature covers country comparisons with purchasing power standards, smaller-scale regional research, e.g. by Cadil et al. (2014), find that for the Czech Republic regional disparities with spatially adjusted purchasing powers lead to regional house price differences. With a low personal purchasing power, lower investments in an own dwelling are possible (Reichert, 1990).

On the infrastructure side, the housing supply is associated with the accessibility of a region. For that reason, Efthymiou & Antoniou (2013) test regional transportation infrastructure in Greece and evidence positive as well as negative associations with regional house prices. While trams, metros and bus stations have a positive association with house prices, airports and ports intrude the region with noises, leading to a negative association on house prices.

2.2 Supply drivers

Next to demand drivers, the supply of dwellings shapes, regional house prices. The supply can be expressed with regional residential building permits and land prices. Shiller (2007) observes a positive association of regional house prices and the building land price with different regional magnitudes. While mainly in urban areas building land is scarce, home-owner and investors decide to construct high rise buildings, leading to lower building land prices per square meter of living space in apartment buildings. Einig (2003) find that a high number of building permits, relative to the number of inhabitants in a county, are used to encounter supply imbalances.

2.3 Other external drivers

Positive and negative associations with house prices can also be found in regional policy measures. Next to creating and approving more building land, regional public aids as economic and infrastructural subsidies, are raised to increase the regional attractiveness that has positive associations on the regional house price level (Buhr, 1981). infrastructural subsidies as studied by Efthymiou and Antoniou (2013) are mainly present in regions that aim to attract people to immigrate or where the purchasing power and current house prices are low. High investments in rural transportation systems and low land taxes are measures that can positively increase the regional house price level (Poterba, 1983). Similarly, investments in urban transportation systems can positively affect the regional house price, while the motivation behind the expansion can also be even more diverse. Another policy implication is given by tax subsidies for owner-occupied housing in order to stimulate the regional house price level.

As a consequence of low regional affordability on housing, policy tools, are used to narrow down regional house price differences towards the national house price level. In a market with excessive demand, however, the price people are willing to pay, often mismatches the offered supplied prices. In a market clearing process, the excessive demand adjusts to the regional house price level (Kulikauskas, 2015). If regional demand exceeds the given supply, emigration shifts to regions with higher housing stocks are the consequence. Assuming an increase in demanded quantity for housing, e.g. by migration or higher regional income, the market adaptation will lead to higher regional house prices. Alternatively, if the housing stock cannot be extended, policy tools, as an increase in land taxes, can decrease the excessive demand and vice versa.

The demand, supply and other drivers are gathered in the following table 2, presenting the reference to the paper and the direction of the associations.

Table 2: Demand, supply and other drivers that associate with regional house price levels

Driver	Association to regional house price	Reference
Immigration (= population)	+	Jacobs (1969); Buch et al. (2014); Bleck and Wagner (2006)
Employment level	+	Buch et al. (2014)
Economic activity	+	Renkow (2003); Bleck & Wagner (2006)
Urbanization	+	Bleck & Wagner (2006)
Average population age	-	Berndgen-Kaiser et al. (2014); Mankiw and Weil (1989); DiPasquale and Wheaton (1994)
Personal income	+	Capozza et al. (2002); Meen (2002), Gallin (2006); Chen et al. (2007); OECD (2016)
Purchasing power	+	Cadil et al. (2014)
Regional rent price level	+	Philiponnet & Turrini (2017)
Building land price	+	Shiller (2007)
Transportation infrastructure	+/-	Efthymiou and Antoniou (2013)
Infrastructure aids	+/0	Efthymiou and Antoniou (2013)
Tax subsidies	+	Poterba (1983)

Note: All drivers refer to a regional characteristic, associating with the regarded regional house price level per square meter.

2.4 Stock-Flow model

For a comprehensive view, the different social, economic, geographic and infrastructural influences have to be tied together, following a stock-flow approach, commonly used in literature. The advantage of this model for this thesis is that supply, demand and drivers in particular can be related and visualized in equations with each other. This way, interactions can be observed and the origin of variables comprehended. The following model is guided by the work of DiPasquale and Wheaton (1994), which has been used and adjusted extensively by further research. Although some parts, e.g. house price developments over time, are not important for the analysis part of this research, the full baseline model is presented for one county to receive a complete understanding of the market associations.

$$D_t = H_t(\alpha_0 - \alpha_1 U_t) \tag{1}$$

$$U_t = P_t(M_t - I_t) \tag{2}$$

The two equations describe the demand function (1) with the user cost function (2), where for time t=1,.., T, H is the population variable, U presents the user costs of owner-occupier, M is mortgage costs, I are capital gains and P is the house price. Besides, α is a given response parameter. Alpha can be α_0, which indicates the intercept (the share of owner occupiers in case of 'no' user costs) and α_1, the response of D to the user cost.

$$D_t = S_t \tag{3}$$

Following the equilibrium condition in equation (3) where S is the Stock of dwellings, equations (1) and (2) can be rearranged. Equation (4) shows that house prices, in the short run, are driven by S_t, H_t, M_t and I_t.

$$P_t = \frac{(\alpha_0 - S_t/H_t)}{\alpha_1(M_t - I_t)} \tag{4}$$

In the long run, new construction of dwellings C and the demolition of dwellings δ (given as share of annual demolition) change the stock of dwellings endogenously. With the condition that new construction only occurs when supply is lower than demand, the equilibrium stock ES is given in equation (7).

$$C_t = \tau(ES_t - S_t) \geq 0 \tag{7}$$
$$ES_t = -\beta_0 + \beta_1 P_t$$

τ is a parameter that indicates how fast new construction respond to a disequilibrium *(ES-S)* and β is a parameter that can be β_0 for the intercept and β_1 as the response parameter of ES to house prices.

After substituting (7) into (5) a steady-state market establishes, with the stock being driven by the house price of P_{t-1}.

After the equilibrium shock and after passing the steady-state, a new equilibrium is obtained. The new house price is given in equation (8) and the new supply is given in equation (9).

$$P^* = \frac{\alpha_0 H_t(\tau+\delta)+\tau\beta_0}{H_t(\tau+\delta)\alpha_1(M_t-I_t)+\tau\beta_1} \tag{8}$$

$$S^* = \frac{\tau}{(\tau+\delta)}\left[-\beta_0+\beta_1\left[\frac{\alpha_0 H(\delta+\tau)+\tau\beta_0}{\alpha_0 H(\delta+\tau)\alpha_1(M_t-I_{t)}+\tau\beta_1}\right]\right] \tag{9}$$

2.5 Hypotheses

The theory allows to derive the following hypotheses. With regard to the research questions, the two hypotheses are used to further extend the aim of this work.

$$H_1\text{: } Metropolitan\ counties\ are\ associated\ with\ higher\ regional\ house\ price, c.p.$$

$$H_2\text{: } The\ regional\ average\ anual\ personal\ income\ is\ positively\ associated\ with\ the\ regional\ house\ price\ level, c.p.$$

3. **Data & Method**

3.1 Data sources

The data comes from 4 different cross-sectional sources, with two of them being open-source data. Firstly, for the dependent variable different house prices are provided by Postbank, which runs regional analyses based on annual market observations. Current regional house prices per square meter are available on new and existing dwellings as a median calculation and as an arithmetic calculation for existing dwellings only. The second source is the public German Regional Database (Regionaldatenbank Deutschland) which provides detailed results of the official statistics of the federal and state governments. The tables offered are based on the regional statistical data catalogue and the Regio-Stat special program. Based on the existing literature variables are sorted to their sources and units in table A2 in appendix A. In the following analyses, the variables will be considered. The third origin is a platform called "INKAR", the abbreviation for "Indicators and maps for spatial and urban development" (Indikatoren und Karten zur Raum- und Stadtentwicklung). INKAR is legally subordinated to the Federal Institute for Building, Urban and Spatial Research (BBSR) and thus to the Federal Office for Building and Regional Planning (BBR), which in turn is directly subordinate to the Federal Ministry of the Interior, Building and Home Affairs (BMI).

Furthermore, the source for most of the socio-cultural data is provided by the "Bertelsmann Stiftung", which is a platform giving insight to many demographical indicators, on a high-quality level. The Bertelsmann Stiftung forwarded a dataset with roughly 100 variables for this thesis, covering the years 2006 – 2018. Whilst the intersection with the other dataset, especially with the house price values is limited to the year 2017, the research question is answered for 2017 exclusively – as a cross sectional comparison of regional house price levels. With the last source, the German Postbank, providing data on regional house prices and their corresponding inflation, this work becomes notable in terms of a comprehensive and high-quality data set.

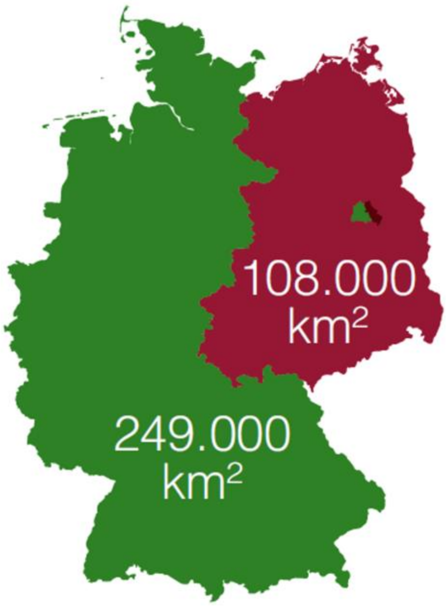
As the different sources are split up into different regions, from a very small municipality level, towards a federal state level, several assumptions have to be made. Firstly, and most importantly, the datasets had to be merged, while the demographical indicators like population age and population density, by "Wegweiser Kommune" covered 3362 municipalities and the house price variables by Postbank only covered 401 rural counties and urban areas. Due to a lack of qualitative data on house prices in time and across space, that is available without paying high processing fees, the concession is to reduce the regarded regions to 401 counties so that the regarded number of controlled regions becomes n=40. The regions are categorized with assigned official regional codes, and averages of the 3362 municipalities are determined to the corresponding counties. The second constraint is the operationalization of the data, which are the regional names and their explanations on the structure. Thus, string variables are changed to binary and categorical variables and missing values are added manually whenever possible, for instance by adding population numbers based on official census data. Inconsistency and logical errors in the data are also corrected – especially correcting commas and names after they have been checked with other data sources. This way a descriptive analysis and selection of the variables becomes feasible.

3.2 Regional housing markets in Germany

The context of the German market is examined in this subchapter. The aim is to explore regional housing markets in Germany. Recent work focuses on an east-west, north-south and urban-rural segmentation.

East-west disparity

Economic power and amenities of living differ substantially between Germanies' eastern and western federal states. Diekmann (2019) stresses the perceived attractiveness of living and state that this is one driver for regional house price differences (=disparities). Surfaces and densities are presented by the work of Kempermann et al. (2019) in figure 2 to show the differences between the old and new federal states after the German reunion in 1990.



	Surface		Population		Urban area	
	km²	%	Absolute number	%	Absolute number	%
East	107.99	30.2	16,538,250	20	20	18.5
West	249.592	69.8	66,153,000	80	86	81.5
Total	357.582	100	82.69125	100	106	100

Figure 2: Isolation of former and new federal states. Source: own presentation based on Kempermann et al. (2019)

With 16,184 million people living in the east and 66,608 million in the west of Germany in 2017, the population density per km² is 73.07% higher in the west of Germany than in the east. Based on historical categorization, the eastern federal states, also called new federal states, contain Berlin, Brandenburg, Mecklenburg Western Pomerania, Saxony, Saxony-Anhalt and Thuringia. Due to the chosen county-level comparison in the given data set, there is no separation of East and West Berlin and Berlin is considered as an eastern county. For the western federal states Schleswig Holstein, Hamburg, Bremen, Lower Saxony, Northrine-Westphalia, Hessen, Rhineland Palatinate, Saarland, Rhineland Palatinate and Bavaria are grouped so that all 16 states are covered.

Comparing counties between east and west Germany differ in regional house prices, Kempermann et al. (2019) stress that controlling for only for inhabitant numbers, medium size cities in the east have 27% lower square meter prices than in the west. Based on the fact that fewer metropolitan areas are located in the east, the assumption of lower house prices in the east becomes more comprehensible as Rosenthal & Strange (2004) elaborate on economic activities. Socio-cultural comparisons between the former east and west counties of Germany are considered by Hiller & Lerbs (2016). In their work, the historical circumstances of east Germany lead to a migration from east to west, resulting in heterogenous demographic developments, including fertility, life expectancy and age structures.

The average age structure of inhabitants also differs across Germany. That accumulation of people and the created population density are, what Rosenthal & Strange (2004) mention as key determinants for creating economic activities and external effects. These agglomerations create wealth and are reflected in increasing house prices. In form of a literature review, Rosenthal & Strange (2004) state that instead of commuting to their work people are willing to pay higher prices for rents and owning an apartment or house. In 2016 the average age in German urban areas was 42.8, whereas in rural areas the average age was 44.9 (Henger & Oberst, 2019). Another study, by Hiller & Lerbs (2016) investigate the 87 biggest cities and found that a 1 percentage point increase in age leads to a 0.8 and 0.5 percentage point for the real apartment and single house, respectively.

North-south disparity

For the next regional disparity, the discussion of the German context is helpful. For a north-south comparison in Germany, local job prospects explain the context for regional house prices. A regional economic analysis by Wolf (2016) shows these differences, that are reflected in related economic indicators. In the past 20 years, the number of employed persons is on a high level - 19.7% and 19.0% which is nowhere else as strongly as in Bavaria and Baden-Württemberg, respectively. Within northern Germany, only Hamburg (+18.1%) and Lower Saxony (+17.7%) were able to create an above-average number of jobs compared to the rest of Germany. Net immigration in the south is also significantly higher than in northern Germany. The lower unemployment rate of 3.5% in the south can also be seen as an indicator for economic strength and consequently for higher square meter prices of residential properties (Wolf, 2016). In existing comparisons, solely 11 federal states are considered to be divided into North and South (Gradmann, 1931):

- South: Rhineland-Palatinate, Baden-Wuerttemberg, Bavaria and Saarland
- North: Schleswig-Holstein, Hamburg, Lower Saxony, Bremen, Brandenburg, Mecklenburg-Western Pomerania and Saxony-Anhalt

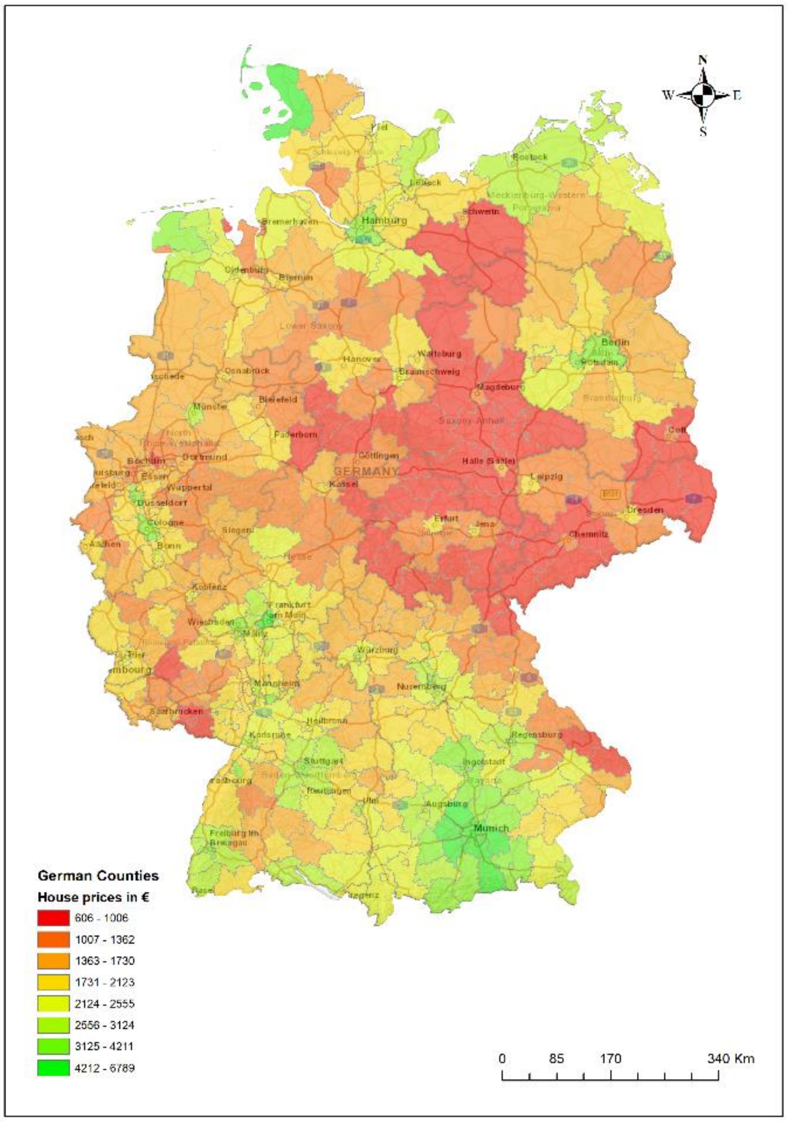
Urban-rural disparity

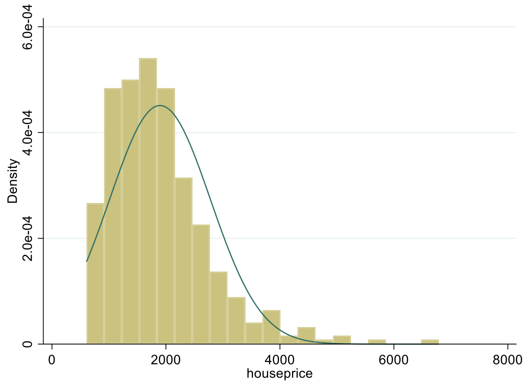
Finally, important regional segregation on the German house market is observed throughout the nation and thus is also a substantial association for the other two house price disparities. The urban-rural disparity occurred by many of the associations illustrated in Chapter 2. In particular, Henger & Oberst (2019) examine the demography on the 401 counties and find that the median age in the urban counties remained on a constant level between 2006 - 2016. For the 293 rural areas, the median age increased significantly from 39.5 to 44.8 during the study period. This is also due to the recent refugee policy, which describes the naturalization of the 108 urban areas as more reliable. The exceptionally high total migration to other regions is due to the distribution of initial reception facilities for refugees. Overall, Germany accepted 643,000 foreign residents annually in the years 2012 to 2016, of which 42 percent ultimately moved to the large cities (Henger & Oberst, 2019). Results by Pomogajko & Voigtländer (2012) prove on a 1% significance level with a sufficient coefficient of determination that the expectation of an increase in demand for space by 1 percent in 2011 leads to an increase in current prices of around 18 percent. The buying power index and the population also indicate positive correlations towards house prices, which was already shown in Chapter 2 by referring to various authors for different countries. Generally, Grabka & Goebel (2018) create deciles for real income groups across Germany and found an income gap of 40% of the highest and lowest deciles in 2015. This results in different willingness's to pay for absolute house prices. As the above shown economic activity is locally situated, the different income groups are heterogeneously distributed across urban and rural counties. A common measure to indicate income disparities is the Gini-Coefficient that gathers incomes, here real incomes, in a respected market which provides a value between 0 and 1, where a high Gini-Coefficient presents a high-income disparity (Grabka & Goebel, 2018). For 2016 the Gini-Coefficient was 0,319, inferencing that many geographically segmented income groups are not willing to pay high house prices in urban areas (Grabka & Goebel, 2018; Kempermann et al., 2019).

3.3 Descriptive analysis

Dependent variable

After examining the German context, the next step is to consider the quality of the data and have a first narrative-exploratory analysis of the data. Therefore, first of all, attention is given to the dependent variable, namely the regional house price per square meter.





houseprice

	Percentiles	Smallest		
1%	692,32	606,42		
5%	837,21	623,54		
10%	958,47	675,52	Obs	401
25%	1265,17	675,77	Sum of Wgt.	401
50%	1758,36		Mean	1896,425
		Largest	Std. Dev.	884,461
75%	2270,62	5205,3		
90%	3017,79	5239,92	Variance	782271,2
95%	3572,1	5646,81	Skewness	1,506525
99%	4836,15	6789,44	Kurtosis	6,704369

Figure 3: Median house prices with a square meter price level over 401 counties. The selection of house prices in the form of median values offered a better normal distribution than the arithmetic value, as both values are logged as aggregates of each region, leaving the arithmetic value as too sensitive for outliers.

Throughout the German nation, variation in house price levels can be observed. Figure 3 illustrates an overview of the German house price situation, with an average price of 1,896.43 Euro per square meter and a distribution skewed to the right. House price levels seem to have spatially concentrated, especially for lower house price levels.

Further regional house price differences, between the determined disparity regions from chapter 2, are given in table A6 in appendix B. It showed that the differences in urban-rural comparisons are the highest, followed by east-west and north-south respectively. In 2017, the eleven biggest metropolitan regions covered 56,79 million inhabitants, including regions as the Rhein-Ruhr Valley, Berlin-Brandenburg, München, Frankfurt-Rhein-Main-Area, Stuttgart, Hamburg and so on (Statista, 2019). Roughly, 77% of the total population, which was 82,792,000 million in 2017, lived in urban areas (World Bank, 2020).

Given the right-skewed distribution, the regional house price per square meter is transferred using the natural logarithm. Figure 4 shows an almost normal distribution of regionals house prices per square meter as the natural logarithm.

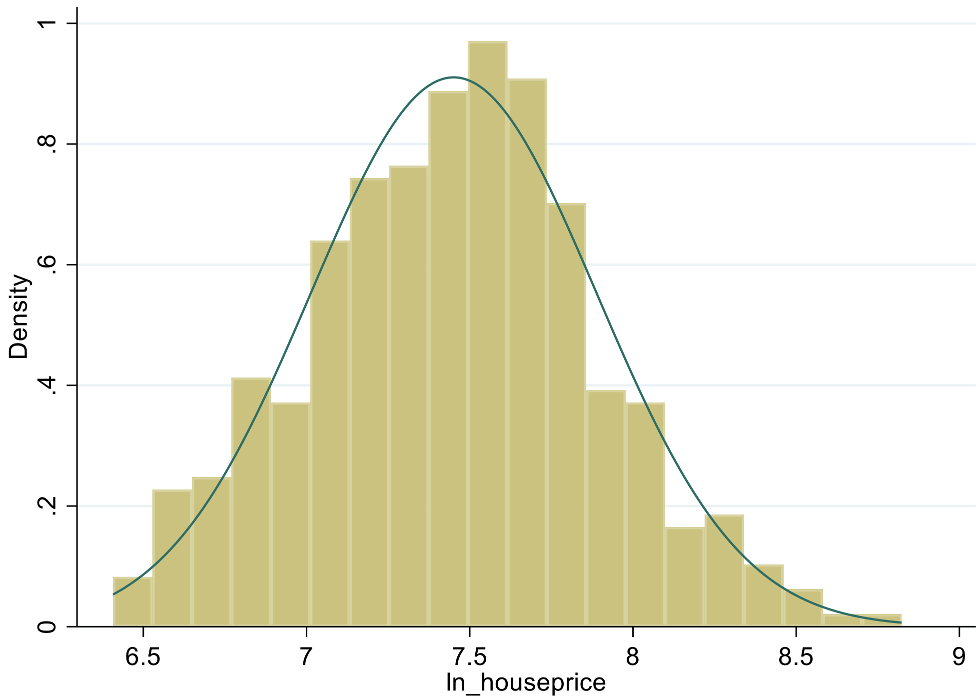


Figure 4: Histogram of ln (house price). The figure shows the LN house price values of each county per square meter, summed by frequency.

The house prices refer to the existing stock of dwellings in a region. Regional convergences of house price regions are visible, especially in the east with low square meter prices, in the west with average square meter prices and in the south, mainly around Munich, with high prices.

<u>Independent variables</u>

For the regional distinction, several dummy variables are created. Firstly, a differentiation between urban and rural regions to account for urban-rural disparities in the regression. The variable Urban introduces the distinction between urban (=1) and the reference group rural (=0) into a dichotomous and applicable variable, to consider the associations separately as either urban or rural counties as a dummy variable. The urban and rural definitions are given by the commonly used definition criteria on the core population and urban fringe definitions

(Borcherdt, 1977). Regarding Germany, these categorizations are predefined however, as mentioned in the introduction at research question 3 we categorized the counties of Aachen and Hannover as urban counties as they fit more adequately to urban areas. Other researchers proceed likewise, i.e. Henger & Oberst (2019), leading to 293 rural and 108 urban counties in Germany.
For the east-west and the north-south disparity, the federal states of the counties are grouped into dummy variables, which can be comprehended in table A3 in appendix A. Further usage in the analysis model is explained in chapter 3.5. A summary statistic on the spatial level of interest will be added in chapter 3.4.

The selection of explanatory variables of interest is derived from previous studies that are expounded in chapter 2. The regional rent price average is considered an indicator of house prices. In a county that is attractive for residents and high housing values, the average rent trends in the same direction.

Also, the real estate structure might be revealing for interpretation. The distinction is twofold. On the one hand single-family and duplex homes and on the other hand apartment complexes. While in cities with high square meter prices the single-family and duplex home rate is comparatively lower, which is mostly the case in urban areas.

Also, for the identification of regulatory associations, the building land price is a variable that measures regional building land values in Euro per square meter. They are derived from the building land purchases of each region in 2017. Intuitively, regulations on originating new building land, form a market place and determine, among other drivers, the price (Vermeulen & Van Ommeren, 2009). The building land value is determined, by a demand and supply approach, where the demand exceeds the inelastic supply, and regulated by policy makers.

Next, the variable gross domestic product and the gross value added, GDP and GVA respectively, are compared (definitions: see appendix A, Table A1). Apart from the individual income per person, the GDP can be considered as a parameter for economic wealth in a region, and thus might associate with the spending power on housing. However, after analysing the GDP it became clear that the operability for regressions is not as suitable as the GVA, which is more normally distributed and measures the value of goods and services that are produced regionally.

On a personal level, the average household income indicates the spending power and thus the willingness to pay for housing. It is measured in available income in € per inhabitant for a regarded county. It can be understood as the amount that is available to private households for consumption purposes or for saving.

The economic variable “Beds for Tourist” addresses regional tourism and states how man beds are available per 1.000 inhabitants. Accommodation companies that host more than eight guests at the same time and only temporarily are accounted - campsites are excluded.

For socio-cultural variables the populations' average age, as promoted by DiPasquale & Wheaton (1994), describes the average age of a defined group, here the population as a whole. It is stated as the arithmetic means of the age of all people in one county at the end of 2017.

The independent variable migration describes the balance of as the sum of total out and inflows into a county, per 1,000 inhabitants.

For the set of infrastructural variables, a wide range of regional characteristics can be considered. Population density is a variable that describes how many inhabitants per km² of settlement and traffic area live in a region. The population density of a county can be used as an indicator (Rosenthal & Strange, 2004). Obviously, urban counties have a higher population density than rural counties with the density being a more suitable indicator than the population measured in absolute numbers. Figure 5 graphically scatters the positive relationship between the population density and house prices.

Another infrastructural variable is the average broad band access as the share of households in a county with a broadband connection of at least 50 Mbit / s.

Furthermore, the distance to the next long-distance railway station is expressed as the variable Train distance. It is the area-weighted average value of car travel times to the IC or ICE stop. The selected train stations are all of the IC, EC and ICE system stops of the German public train service, even those in which only individual trains operate. The accessibility calculations of motorized private transport are based on route searches in a road network model. The lowest value is observed in the north-eastern rural counties, while 48 counties, all of them are metropolitan regions, received an average distance of 0 Minutes.

The variable floor space ratio describes the average living area in square meters per household in a county.

Building permissions, as introduced by Einig (2003) are the variable to measure a county's growth and as well to cover the institutional strength of creating more living space. It is measured in building permissions for new apartments per 1,000 residents in a county.

Infrastructure and economic aids aim at the improvement of the regional economy and infrastructure in long terms respectively. The variables are measured as financial subsidies in Euro per inhabitant in a county.

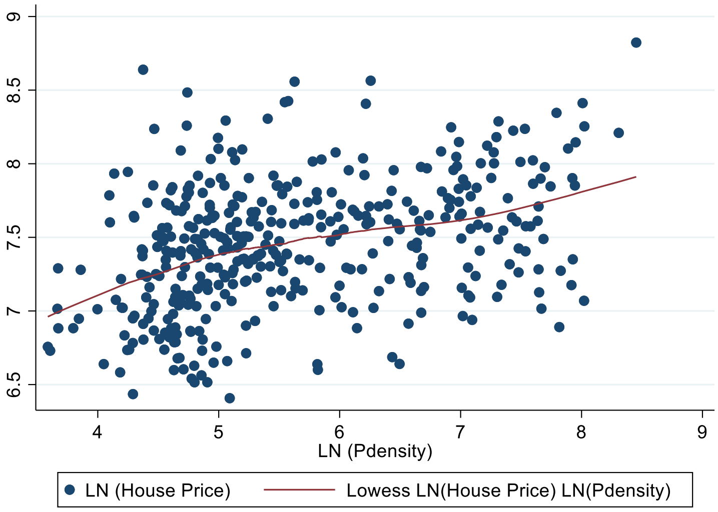
Creating a table on the possible determinants including the mean, standard deviation as well as minimal and maximal values, table 3 gives an overview of the data, while table A4 of appendix B summarizes the regressor variables including their LN for the regression model.

Table 3: Descriptive Statistics of Variables

Variable	Mean	Std.Dev.	Min	Max
House price	1896.43	884.461	606.42	6789.44
Gross value added	62.438	10.887	46.73	147.49
Annual income	22470.76	2589.134	16382.04	38909.04
Beds for Tourist	41.867	49.249	3.7	405.7
Average age	44.54	1.967	39.81	50.21
Population density	533.753	702.7	36.13	4686.17
Migration balance	-10.361	29.725	-149.37	62.66
Broadband access	76.663	15.45	27.42	99.6
Train distance	21.925	15.379	0	79
Rent average	7.256	1.802	4.63	17.36
Share single and duplex homes	83.143	10.558	50.1	96.1
Floor space ratio	46.449	4.601	35	67.5
Building land price	175.26	221.133	11.5	2428.7
Building permissions	3.468	1.94195	.4	15
Economic aid	52.595	132.028	0	1034.7
Infrastructure aid	77.533	67.731	.3	430.9
Rent-to-price	5.087	1.329	1.534	9.93
Price-to-Income	8.31	3.354	3.00	23.3
Purchase Power-to-Price	6.249	2.863	1.74	21.64
Observations	401			

Note: Explorative statistics of regressor variables, before transforming with log ln. All variables in the table are national averages, build by the arithmetic average of the 401 counties. The dependent regression variable (without log ln) is included for comprehensiveness reasons – as well as the real estate ratios that will be introduced in chapter 3.4.

Figure 5: Scatterplot of ln (house price) per square meter and ln (population density)



Note: house prices and population density are expressed on a county level as the median average.

With a graphical overview of the data in figure A2-A4 (appendix B)**,** the visualization of the typical measures of central tendency is completed. High differences between the respected possible associations for house prices can be observed. For instance, the fundamentals like the Gross value added figure can be more than three times higher in a county, in Wolfsburg where the VW car production plant is located, than in the county Südwestpfalz, a rural eastern county. Similarly, the variable tourist reaching from 3.7 to 405.7, also indicate big differences in the attractiveness between counties. The more significant variable turned out to be the regional gross value added figure, as the normality of distribution delivers more robust output than GDP. Interestingly, the number of observed single family homes and duplex homes varies between roughly 50% in metropolitan areas and 96% in very rural regions. Apart from the given characteristics a detailed table with all other explanations on the variables and data origins is given in appendix A**,** Table A2. Exploring the regional house price per square meter ratios, high deviations are observed that will be further affiliated and explained in chapter 3.4.

Finally, the mean values of the considered variables for all counties, to answer the research sub question 3, are listed in table A5 of appendix B according to their spatial location. The mentioned normalisation of variables is a necessary step for the operability in an ordinary least square regression model (Brooks & Tsolacos, 2010) and is graphically presented in Table A5 to A10 of appendix C.

3.4 Real Estate Ratios

We now consider real estate ratios to further explore regional house price differences in Germany. This part of the descriptive analysis gives information about different house price levels and operationalizes the numerical relationships to compare counties with each other. With existing literature as the foundation, the ratios will visualize different house price levels. Among others, explanatory power is found for regional house prices in migration behaviour, age and economic wealth. These economic associations will now be integrated into 3 indexes (Kulikauskas, 2015; Philiponnet & Turrini, 2017), with the full list of results for all regions being available in table A8 of appendix B, as well as an overview on the accuracy of the data:

<u>1. Rent-to-price ratio</u>

The first ratio is commonly accepted for private persons and institutional investors, as well as in research. While a rent-to-price ratio is of prevalent interest for societal matters, we keep the original values given by the data set instead of using logarithmic values. Thus, the following equation for the regional rent to price ratio is

$$RtP_i = \frac{R_i}{P_i} * 100 \qquad (1)$$

where RtP stands for the rent-to-price ratio and R presents the regional average net annual rent per square meter in Euro, which is obtained by multiplying the average rent of the dataset by 12. Similarly, P stands for the regional median house price value measured in Euro per square meter. Deleting the Euro per square meter on the variables R and P, the rent-to-price ratio is defined by a dimensionless value that typically ranges from 0 to 1, based on the nature of the positive numbers of R and P. The multiplication with 100, delivers a percentage value. Socially relevance is reflected in finding explanations, for the increasing demand for owner-occupier in supply-limited regions. For instance, the lower the RtP ratio the better for buyers, c.p. This is because a low RtP ratio indicates either low regional house prices or high rents per square meter, which both can indicate a “buy” investment decision. The following thresholds are used by Eilinghoff (2019) for the German market which can be a decision guideline for potential buyers:

Table 4: Investment thresholds based on rent-to-price ratios

Ratio		Factor		Investment
from	to	from	to	decision
100.00	5.00	1	20	buy
4.99	4.00	20.01	25	rent, buy
3.99		25.01		rent

Note: The table gives an average orientation of whether it is worth to invest in an own dwelling rather than to rent a dwelling. It compares the regional rent price level and house price level per square meter. The investment decision row gives a suggestion how to decide in a region, as home-owner costs (e.g. yearly land taxes) are not considered in a rent-to-price ratio.

The factor expression in table 4 is the conversely formulated rent-to-price ratio, namely the price-to-rent ratio with a switched numerator and denominator in equation (1). Noticeably, for a regional house price with a factor value of 4% a buy-decision can still be the better choice, as the capital gains over time are not taken into account. Thus, owner-occupiers face risks as indicated by Berndgen-Kaiser et al. (2014) but are also confronted with costs, e.g. closing costs, mortgage costs, maintenance, property taxes, and insurances, resulting in the above-given thresholds.

Calculating the ratio on the given data set, the segmented market on the German house price becomes obvious. With an RtP ratio ranging from 1.5% to 9.9% and a mean of 5.09%, generally the decision to buy, and thus the demand side is strong, however, with regional disparities. Table A8 in appendix B summarizes all the ratios including the rent-to-price index sorted alphabetically, so that interested parties can look up regions. For instance, if a family is considering to move and ceteris paribus considers the RtP indicator as decisive, leaving out factors as commuting and amenities, the RtP can give a first impression on the distractive situation. The following table 5 lists the 5 highest and lowest RtP values.

Table 5: Extract of rent-to-price values

Rank	Name of county	House Price	RtP	Sqm Living Area
1	Nordfriesland	5646	1.53	58.1
2	Aurich	3553	2.16	54.8
3	Regensburg	3778	2.55	44.0
4	Landshut	3454	2.66	45.4
5	Rostock, County	2787	2.70	46.0
…	…	…	…	…
397	Dessau-Roßlau, City	763	8.92	46.6
398	Vogtlandkreis	606	9.16	47.5
399	Salzlandkreis	675	9.16	47.6
400	Nordhausen	692	9.57	45.1
401	Kyffhäuserkreis	623	9.93	49.0

Note: The table shows Germanies highest and lowest rent-to-price ratios and the corresponding square meter house prices and average available living space per inhabitant – on a 401 county-level perspective.

Sorting the table highlights that low RtP values are found in very demanded regions with limited space. For instance, the county of Nordfriesland includes the attractive island travel destinations of Sylt, while high RtP values are achieved in more rural regions in the east of Germany. The same accounts for the county Aurich, which includes among others the popular islands Juist and Norderney. With reference to the composition of the index, consciously a square meter comparison for RtP was selected, for a better comparison of the counties instead of choosing the RtP based on the average living area in each region. This method yields higher adjustability for private demand on the dwelling size, admitting that the ratio itself does not take into account the widespread all locational factors. However, with the next ratios, a clearer regional distinction becomes possible for housing seekers.

<u>Income-to-price ratio</u>

The second index, the price-to-income ratio, a common measure for regional comparisons, indicates to what extend households are able to afford a dwelling for their own residency. In contrast to Philiponnet & Turrinia (2017), who asses house price levels on a European country scale over time rather than between countries, the index here will compare the 401 counties in Germany in 2017.

$$ItP_i = \frac{I_i * 12}{P_i} * 100 \tag{2}$$

Equation (2) formulates the layout of the index ItP being the income-to-price ratio, resulting from dividing the annual average personal income I by regional house price value on a square meter price in the same region i. The results of the ratio are on the one hand presented based on the urban-rural, east-west, north-south distinction in table A7, appendix B**.** For the ItP ratio, it is further graphically illustrated in figure 6.

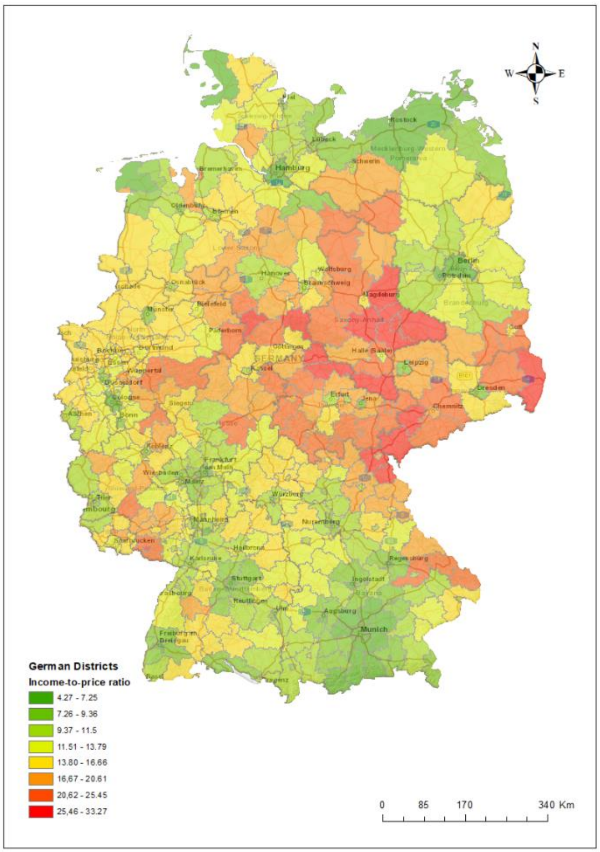


Figure 6: Map of Germany with income to (house)price ratio on a county level.

Regional disparities are visible with high values in the metropolitan areas and the coastal regions, while high ItP values are mainly situated in the centre and east of Germany – typically rural regions with low house price values. While low house prices do not necessarily lead to a high ItP ratio, in combination with a sufficiently high income distribution the house price value does so.

Definite differences on a broad level are obvious while also smaller ItP ratio convergences of counties are measurable – mainly in Bavaria and as well as around Munich and Frankfurt am Main. Especially if compared to other academic work in the field (Zhang et al., 2016; Goodman, 1988) the ItP ratio is more suitable as it is used more frequently than the purchase-power ratio which is introduced in the following.

Purchase-power-to-price ratio

The comparison approach of Cadil et al. (2014) of different purchasing power derivation and indexes, including time and timeless dimensions, delivers the basis for this new and rarely used ratio. Adding the economic index of the income-to-price ratio, the purchasing power-to-price ratio accounts for spatial characteristics across regions. Purchasing power is the measure of the value of money and offers comparisons between counties. The purchasing power of money indicates the number of goods that can be bought with a monetary unit or a certain amount of money. However, the prices of goods are constantly changing. So heating oil is more expensive in winter, while other goods such as bicycles or computers become cheaper in the same period. The purchasing power of money can therefore only be measured in relation to certain goods. For this purpose, a shopping cart is put together, which contains the typical goods that are bought by a household and which is used in the calculation of the consumer price index. If the price index of living expenses has increased (decreased), the purchasing power of money concerning goods in the shopping basket has decreased (increased). The purchasing power of money has therefore increased (decreased) when more (less) goods can be bought for a unit of money than at an earlier point in time. The monetary value and price level are therefore reversed and instead of comparing it across time, it is compared across space (Bpb_b, n.d.). The dataset for this work contains an already calculated and averaged purchasing power index for all 401 German regions with a reference value of 100 for Germany as a country.

$$PPtP_i = \frac{PP_i}{P_i} * 100 \tag{3}$$

With equation (3) the purchase-power-to-price ratio, defined as $PPtP$ in Region i described as the division of median house price measured in Euro per square meter by the purchase power of the same region. Multiplying it by 100 results in a percentage value.

Finally, all ratios are summed in table 6 while table A7 in appendix B gives insight into the disparity regions separately. A high standard deviation of 5.42 for the ItP ratio indicates a very diverse distribution. More specifically, the highest RtP, ItP and are located in eastern-urban and eastern counties, respectively.

Table 6: Overview of Real-Estate-Ratio Distributions

Ratio	Mean	Std. Dev.	Min	Max
Rent-to-price	5.087	1.329	1.534	9.93
Income-to-price	13.95	5.42	4.27	33.27
Purchase-power-to-price	6.061	2.241	1.732	13.80
Observations	401			

3.5 Model development

The empirical model follows the concept of Fingleton (2006), who asses a similar approach to examine the distribution of residential property prices on a cross sectional level for London for 2001. At this point, it is necessary to develop a model as the before given real estate ratios do not consider the variables of interest on a satisfying level. The model is expressed in the following equations, in order to identify whether regional house price differences exist on different spatial levels and how dominant the variables are.

$$\text{LN}\,(P_y) = \alpha_1 + \beta \sum_1^{16} FedState_y + \varepsilon_y \tag{5}$$

The analysis starts with a basic model that expresses location fixed effects in two ways. Firstly, equation (5) determines ceteris paribus how the location factors associate the regional ln house price per square meter. The reduced form of the model starts with the natural logarithm of the house prices P as the dependent variable and is determined by a constant term α_1, while β is the coefficients for the fixed federal effects (FedState) for each of the sixteen federal states in Germany, indicated with the subscript y. ε is the stochastic error disturbance of associations that cannot be explained by the rest of the equation. The purpose of equation (5) is motivated by the analysis to identify contextual differences in house price levels on a federal state level.

$$\begin{aligned} \mathrm{LN}\,(P_y) = {} & \alpha_1 + \beta \sum_1^{16} FedState_y + \alpha_2 Gva_y + \alpha_3 \ln\,(Income)_y + \alpha_4 Tourist\ beds_y \\ & + \alpha_5 \ln\,(Age)_y + \alpha_6 \ln\,(Popoulation\ density)_y + \alpha_7 Migration_y \\ & + \alpha_8 \ln\,(Broadband\ access)_y + \alpha_9 Train\ distance_i + \alpha_{10} Rent_y \\ & + \alpha_{11} Single\ family\ home_y + \alpha_{12} \ln\,(Floor\ ratio)_y \\ & + \alpha_{13} \ln\,(Building\ land\ price)_y + \varepsilon_y \end{aligned} \tag{6}$$

In equation (6) the model is extended to a federal state level, where FedState represents the dummy of a county located in federal state y. The α coefficients are structural attributes reflecting federal characteristics. The 11 input variables for federal state y are on the one hand economic figures. Gva indicating the gross value added per inhabitant and $\ln\,(Income)$ presenting the average annual income per inhabitant in Euro as a natural logarithm ln. As in Chapter 3.3. the distribution of the variables was taken into account, resulting in the transformation of the raw data into log ln values, normally distributed regressors delivering significant outputs. Furthermore, it includes the measure $\mathrm{Tourist}$ which represents the beds in a county available for tourism. On the other hand, for socio-cultural associations, the model contains the average age of the inhabitants in the respective region Age expressed as ln, the ln of the population density of a region $\mathrm{LN}(Population\ density)$, measured as the ln of the inhabitants per km² of settlement and traffic area and the migration balance $ln(Migration)$, which is the ln of the sum of out and inflows into a region, per 1,000 inhabitants. The infrastructural variables include the average broad band access $ln\ (Broadband\ access)$ as the share of households with a broadband connection of at least 50 Mbit / s and the variable $Train\ distance$ which represents car travel time to the nearest long-distance train station in minutes. Real Estate variables are gathered, with $Rent$ being the net average regional rent per m², $\ln(Single\ family\ home)$ describing the share of total dwellings being single-family and duplex homes and $Sqmliv$ as the average m² of dwelling space per inhabitant. $\ln\,(Building\ land\ price)$ is the variables explaining local housing regulations as a price mechanism of how much building land is made available by local government and is measured in a square meter price.

$$\text{LN}(P_i) = \alpha_1 + a_2\, location_i + \alpha_3 Gva_i + \alpha_4 \ln(Income)_i + \alpha_5 Tourist\ beds_i + \alpha_6 \ln(Age)_i + \alpha_7 \ln(Popoulation\ density)_i + \alpha_8 Migration_i + \alpha_9 \ln(Broadband\ access)_i + \alpha_{10} Train\ distance_i + \alpha_{11} Rent_i + \alpha_{12} Single\ family\ home_i + \alpha_{13} \ln(Floor\ ratio)_i + \alpha_{14} \ln(Building\ land\ price)_i + \varepsilon_i \quad (7)$$

While the first two models lead this works methodology, equation (7) is the combination of the priors models to address the research sub questions 2 and 3. The cross sectional comparison is downsized to a county level, in order to identify the variables that are associated with regional house price disparities. Based on the given supply and demand determinants, economic, socio-cultural and infrastructural associations, the variable location reflects the disparity regions urban-rural, east-west, north-south separately as well as in an aggregated form later on.

4. Results

This chapter presents the results and discusses several tests that prove the statistical soundness of the approach. First, the analysis focuses on house price differences on a federal state level in order to understand the context of locational drivers. The idea is to narrow down associations and then finally isolate driving variables. For this, house prices are regressed on a set of federal state fixed effects. The results will show how regional house prices vary across federal states. Then, as a next step, it is examined whether regional house prices can be associated with regional, county-level characteristics like urbanization level and geographical location of the counties. Finally, it is examined whether regional house prices can be associated with regional demand and supply drivers.

4.1 House price differences across federal states

The first step aims at controlling for differences, whether on a federal state level fixed effects for house prices can be observed. In this model, fixed effects on a federal state level indicate federal state specific drivers on the regarded house price level. Assuming significant results a positive coefficient for a federal state imply generally a higher house price level in the state, without controlling for other variables.

Table 7: OLS regression on federal state level

VARIABLES	(1) Federal States
Hamburg	0.709**
	(0.359)
Lower Saxony	-0.316***
	(0.104)
Bremen	-0.363
	(0.262)
North Rhine-Westphalia	-0.301***
	(0.102)
Hesse	-0.0644
	(0.113)
Rhineland-Palatinate	-0.251**
	(0.107)
Baden-Württemberg	0.0840
	(0.104)
Bavaria	0.0762
	(0.0966)
Saarland	-0.358**
	(0.168)
Berlin	0.573
	(0.359)
Brandenburg	-0.308**
	(0.122)
Mecklenburg-Western Pomerania	-0.0419
	(0.152)
Saxony	-0.679***
	(0.132)
Saxony-Anhalt	-0.870***
	(0.129)
Thuringia	-0.700***
	(0.115)
Constant	7.637***
	(0.0898)
Observations	401
R-squared	0.393

Note: The coefficients of this table are the locational effects of federal states with the ln of the house price per square meter as the dependent variable. The federal state of Schleswig-Holstein is the reference category for location. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The results of table 7 deliver the first impression on the locational effects of 16 federal states. States like Hesse and Mecklenburg-Western Pomerania are not statistically significant on a 90% significance level. However, positive and negative coefficients with high and low magnitude indicate that based on the locational fixed effects by itself, house price differences are observed. The highest positive coefficient, accounts for Hamburg with 0.709, while the highest negative coefficient represents Saxony-Anhalt with -0.870. These findings indicate that based on fixed effects, the federal state house price levels differ. More nuance patterns appear if one examines house price differences on a county level.

4.2 House price differences across counties and urban areas

As a second step, the analysis narrows down the regional perspective on a county level and controls for house price differences. The motivation at this point is to better characterize the drivers across regions. Table A9 of appendix C, therefore, proves that regional differences for the house prices, expressed as the log ln values per square meter, exist with a high adjusted R^2 of 0.993. Hence, the conclusion is that regional house price differences exist and a grouping of similar counties becomes feasible, based on county-characteristics.

The analysis of the German housing market continues with the differentiation of possible disparities regions, to summarize and synthesize differences across counties. The concept is to compare the locational associations separately from the other causal variables and control whether they differ in direction, magnitude and significance. Therefore, table 8 presents the delimited findings of the spatial associations on house prices.

Table 8: OLS regression on county level

VARIABLES	(1) Urban-Rural (UR)	(2) UR, East- West (EW)	(3) UR, EW, North-South (NS)
Urban	0.197***	0.143***	0.163**
	(0.0484)	(0.0486)	(0.0820)
East		-0.508***	-0.399***
		(0.0573)	(0.0623)
South			0.259***
			(0.0575)
North			0.124**
			(0.0602)
Urban × East		0.244**	0.299**
		(0.114)	(0.121)
Urban × South			0.0558
			(0.104)
Urban × North			-0.141
			(0.114)
Constant	7.398***	7.498***	7.323***
	(0.0251)	(0.0255)	(0.0477)
Observations	401	401	401
R-squared	0.040	0.211	0.278

Note: The table determines ceteris paribus how the location factor affects the regional ln house price per square meter. Interaction variables are included. The location factor categorizes a county into a defined cardinal direction and as well, whether the region is urban or rural. Urban and rural is measured on a county level, while East, West, South and North refers is measured on a federal state level. Regression (3) includes urban-rural east-west and south-north distinctions. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Observing the coefficients of the fixed effects, an interesting finding is that the urban sign in equation (2) is given with 0.143 which can be interpreted as regional house prices being higher in urban areas. However, the interaction of Urban and East, indicating that the variables depend on the value of the other is given with 0.244. As the interaction coefficient is higher than both single coefficients (Urban= 0.143 and East= -0.508) one could argue that for example urban counties are strongly associated with high house price levels in the eastern federal states. Therefore, the H_1 hypotheses can be rejected, that the location by itself has no association with the house price. Notably, the coefficient of the interaction variable Urban & East is 0.244 which partly goes in line with earlier findings and theory. Although Kempermann et al. (2019) report lower house prices in the east, it can be based on the fact that the eastern federal states only have a few urban counties, compared to the west. Then reversely, urban counties in the east can have a way higher house price level than rural counties.

These findings reveal that based on the geographical location of a county, house price level differences exist across Germany. After the findings of table 8 that controlled solely the view on the spatial associations, the input is extended. Therefore, the main findings based on equation (5) are given in table 9. Remembering that no counties are excluded, a complete perspective on the German house price market is possible.

4.3 Regional house price drivers

Next, now significance in house prices across regions is observed and differences across types of regions are analysed. Therefore, association with demand and supply drivers are considered.

For a broad view on the associations of demand and supply, a federal state level comparison is given in table A10 of appendix C. It distinguishes between all associations and highlights directions and magnitudes of the input variables. Starting with the control only for fixed effects, then including the introduced regressors and then finally considering 3 institutional factors. Institutional drivers that might be correlated with house prices are measured by building permits for new apartments per 1,000 inhabitants. Thus, building permits indicate a positive association with the ln of house prices. An increase of 1 unit of building permissions leads to a 1.8% increase of the regional ln square meter house prices. The associations of the GRW subsidy programs, as mentioned in chapter 1.1 only result in small positive coefficients. Infrastructure and economic subsidies aim at the improvement of the regional economy and infrastructure in long terms respectively and do not provide significant statements – therefore they are excluded from further considerations.

Table 9: Estimates of OLS regression model

VARIABLES	(1) All Counties	(2) UR	(3) UR, EW	(4) UR, EW, NS
Gross value added	-0.00368***	-0.00341***	-0.00298***	-0.00290***
	(0.00104)	(0.00103)	(0.00102)	(0.00101)
ln(Income)	0.267**	0.307**	0.527***	0.541***
	(0.120)	(0.120)	(0.129)	(0.131)
Tourist beds	0.00175***	0.00176***	0.00201***	0.00201***
	(0.000203)	(0.000201)	(0.000206)	(0.000207)
ln(Age)	-1.445***	-1.418***	-2.228***	-2.318***
	(0.308)	(0.306)	(0.357)	(0.355)
ln(Population density)	-0.0553**	-0.0744***	-0.0507**	-0.0253
	(0.0225)	(0.0233)	(0.0235)	(0.0246)
Migration	0.00102**	0.000618	0.000352	0.000290
	(0.000404)	(0.000426)	(0.000422)	(0.000418)
ln(Broadband access)	0.0993*	0.0813	0.154**	0.145**
	(0.0601)	(0.0599)	(0.0612)	(0.0610)
Train distance	-0.00259***	-0.00237***	-0.00210***	-0.00204***
	(0.000738)	(0.000735)	(0.000723)	(0.000726)
Rent	0.122***	0.124***	0.109***	0.102***
	(0.00997)	(0.00990)	(0.0104)	(0.0105)
ln(Single family home)	0.307**	0.407***	0.589***	0.530***
	(0.131)	(0.135)	(0.139)	(0.140)
ln(Floor space)	0.104	0.0893	0.0919	0.0652
	(0.0875)	(0.0869)	(0.0851)	(0.0846)
ln(Building land price)	0.182***	0.184***	0.188***	0.184***
	(0.0267)	(0.0265)	(0.0260)	(0.0258)
Urban		0.115***	0.114***	0.0679
		(0.0412)	(0.0404)	(0.0424)
East			0.160***	0.184***
			(0.0383)	(0.0391)
South				0.0728***
				(0.0244)
North				0.0670**
				(0.0260)
Constant	6.866***	6.085***	5.718***	6.187***
	(1.681)	(1.690)	(1.657)	(1.651)
Critical F value (99% significance level)		2.23	2.23	2.23
Chow F statistic		4.13	5.02	11.52
Observations	401	401	401	401
R-squared	0.853	0.855	0.862	0.865
	(1)	(2)	(3)	(4)

Note: The dependent variable is the log ln of house price per square meter on a county level. Similarly, the independent variables annual income (Inc), average age (Age), population density (Pdensity), broadband access (Bb), single family and duplex homes (Sfhd), average living area (sqmliv) and building land price are expressed as ln values and present elasticities for the ln house price. Urban, East, South and North are spatial (dummy) variables for the location of the county. A constant term has been included in all regression. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The answer to '*How can regional house price levels be understood on county-level comparison'* starts with table 9**.** The first regression (1) obtains all German counties, regardless of the location in the country. With a high R^2 of 0.853 the coefficient of determination is sufficiently high and stays above 0.85 throughout the other regression models. Thus, 85 percent of the variance of the dependent variable, ln house price, is explained. As introduced in chapter 3.1 the dummy variable urban refers to urbanized areas of a county level while East, North and South categorize a county to a group of federal states.

The regression model in column (1) shows that the average age of a county, has a high association with the price level for housing. Hence, my model predicts the elasticity from age on the house price. Keeping all other variables constant, a 1 per cent increase of the average age leads to a decrease of the house price per square meter by 1.445 per cent. Furthermore, the age coefficient is significantly different from zero on a 1 per cent level, which brings confidence that this measure is decisive. Thus, the conclusion can be made that the above listed demand and supply drivers determine the house prices on a satisfying level. On the demand side regional house prices are positively correlated with tourism, measured with tourist beds. On the supply side house prices are positively associated with the building land price. Only the average square meter dwelling size per inhabitant is not significant on a 10 per cent level, however, the other input variable, especially the ln of the average annual income shows a high positive and significant association on the regional house price level. For all models, the test statistics exceed the critical f value, therefore the pooled model is rejected for all models. Further explanations are given in chapter 4.4.

Along with the research aim, subquestion 2 receives comprehensive answers through the model built in chapter 3.5., which can now be confirmed on accuracy to present the German residential real estate market. The real estate, economic, socio-cultural, infrastructure measures that have been derived by existing literature, seem to be applicable on the German market. Especially high GVA figures, migration and broadband access positively affect house price values.

Regressions (2), (3) and (4) are devoted to answer sub question 3. Firstly, regression (2) separates urban and rural counties and investigate the same associations as in regression (1), while most of the input associations remained similar. The dummy variable $Urban$ therefore expresses that a county that is declared as urban, have on average a (exp(0.115)-1)100 = 12.18% higher house price value than the reference group rural has. However, the population density is increasing the negative association compared to the baseline regression (1). Now, a one per cent increase in population density, leads to a 7.44% decrease in the corresponding house price. This implicates a mismatch with the existing literature as Rosenthal & Strange (2004) and (Jacobs, 1996) stress higher concentrations of the population with more economic wealth. The third regression, (3), combines the second and third regression and delivers satisfying and conclusive output. With the highest R^2 (0.862) value and by far the lowest constant term (5.718) the model stresses the association on house prices for urban-rural and east-west aspects. With a focus on the coefficients, regions are measured differently throughout the tested models, indicating especially the average age, annual income, migration rates and the share of single family and duplex homes are changing in magnitude the most. While the association and thereby the demand for a close train station remains the same across the compared models, the before mentioned regressors are more volatile, though with statistical significance. Lastly, with a significance level of $p<0.01$, regression (4) shows on both territorial areas, North and South, positive coefficients. Although the coefficients are significant, we cannot observe a clear disparity of associations for both regions as both dummies are positive in direction.

Surprisingly, in regression (2), and (3), the dummy variable $East$ is positive. In (3), with a p-value below 0.01 and a coefficient of 0.161, regional house prices are - holding the other regressors constant - on average 16.1% higher than in the western federal states. However, a substantial increase of the age variable appertains to this effect. For a county in the east-west comparison, a one percent increase in age leads to a 2.228 per cent decrease of the house price. To my knowledge, this has not been identified by existing literature so far. A possible explanation of this observation might be the increase of the age coefficient and the corresponding high age in the eastern regions – see table A5. An interesting conclusion is derived, that when all right-hand side variables, as in table 9, hold constant, the house values are lower in the western counties than in the east.

Finally, dominances of the independent variables are controlled to understand what are the most important drivers for regional house prices.

Table 10: Dominance statistics of the main regression models

ln (house price)	(1) All counties	(2) UR	(3) UR, EW	(4) UR, EW, NS
Gross value added	0.038	0.038	0.035	0.0325
Ln(Income)	0.089	0.090	0.088	0.0796
Tourist beds	0.042	0.043	0.044	0.0452
ln(Age)	0.110	0.109	0.107	0.1015
ln(Population density)	0.032	0.032	0.029	0.0277
Migration	0.007	0.006	0.005	0.0054
ln(Broadband access)	0.033	0.032	0.030	0.0278
Train distance	0.037	0.036	0.035	0.0351
Rent	0.252	0.293	0.241	0.2289
ln(Single family home)	0.013	0.014	0.013	0.0130
ln(Floor space)	0.001	0.001	0.001	0.0009
ln(Building land price)	0.198	0.231	0.193	0.1828
Urban		0.011	0.009	0.0090
East			0.030	0.0224
North				0.0287
South				0.0224

Note: The table gives the results of the dominance statistics (not standardized), to stress how important a variable is for reducing the prediction error of the dependent variable.

Table 10 shows that the building land prices and rents next to income and age are very strong associations for regional house prices. Depending on the model specification, the least important drivers for reasoning the regional ln house price are migration, the average available floor space per person (Floor space) and the share of single family and duplex homes in a region.

4.4 Chow Test

This subchapter confronts the analysis with a test whether the true coefficients are equal across regressions and data sets. Explanations on heterogeneity and other assumption test are presented in appendix D.

Table 11: Chow test based on the regression models

	Pooled	Urban	Rural	East	West	North	South
Residuals	11.32140	2.652372	7.352085	1.93849	7.82357	3.50982	3.88979
Observations	401	108	293	77	324	103	182
F-Value	F(12, 388) =186.97	F(12, 95) =50.87	F(12, 280) =148.13	F(12, 64) =33.01	F(12, 311) =140.16	F(12, 90) =32.20	F(12, 169) =8.27
Critical F Value (99% significance level)	2.23	2.23		2.23		2.23	
Chow F statistic		4.13		5.02		11.52	

Exemplary, inserting the RSS for the urban-rural subsamples with 293 rural regions and 108 urban regions and the pooled regression leads to the following F statistic.

$$F = \frac{(11.32140 - 2.652372 - 7.352085)/12}{(2.652372 + 7.352085)/(293 + 108 - 2 * 12)} = 4.13$$

Comparing the F-value of the Chow test with the F-Value statistics table, the critical value is 2.23 on a 1% significance level, which is lower than the test result of 4.13. Thus, there is no parameter stability throughout both sample groups and therefore we can reject the null hypothesis that the insignificant improvement of fit in separating regressions into urban and rural regions. This approves the goodness of fit of the urban-rural divide. The test statistics of the subgroup division of east-west and north-south exceeds the critical F value as well, therefore the pooled model is rejected for all regressions, see table 11.

Table 12: Regression parameter estimates for Urban-Rural comparison

ln (house price)	Pooled model All Counties	Group 1 Urban	Group 2 Rural
Gross value added	-0.00368***	-0.00386***	-0.00308**
	(0.00104)	(0.00141)	(0.00155)
ln(Income)	0.267**	0.594***	-0.00303
	(0.120)	(0.198)	(0.156)
Tourist beds	0.00175***	0.00481***	0.00177***
	(0.000203)	(0.00147)	(0.000202)
ln(Age)	-1.445***	-0.942	-1.247***
	(0.308)	(0.722)	(0.377)
ln(Population density)	-0.0553**	-0.0394	-0.0456*
	(0.0225)	(0.0547)	(0.0262)
Migration	0.00102**	0.00214*	0.000749
	(0.000404)	(0.00123)	(0.000486)
ln(Broadband access)	0.0993*	0.307	0.0369
	(0.0601)	(0.243)	(0.0616)
Train distance	-0.00259***	-0.00283**	-0.00236***
	(0.000738)	(0.00133)	(0.000869)
Rent	0.122***	0.0890***	0.144***
	(0.00997)	(0.0174)	(0.0129)
ln(Single family home)	0.307**	0.111	1.108***
	(0.131)	(0.207)	(0.241)
ln(Floor space)	0.104	0.0961	0.0282
	(0.0875)	(0.167)	(0.0982)
ln(Building land price)	0.182***	0.170***	0.194***
	(0.0267)	(0.0446)	(0.0336)
Constant	6.866***	1.776	5.475**
	(1.681)	(3.189)	(2.651)
Observations	401	108	293
R-squared	0.853	0.865	0.864

Note: The table presents the separate model results, based on a grouping of the data set. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 13: Regression parameter estimates for East-West comparison

ln (house price)	Pooled model All Counties	Group 1 East	Group 2 West
Gross value added	-0.00368***	-0.00308**	-0.00319***
	(0.00104)	(0.00155)	(0.00100)
ln(Income)	0.267**	-0.00303	0.554***
	(0.120)	(0.156)	(0.130)
Tourist beds	0.00175***	0.00177***	0.00180***
	(0.000203)	(0.000202)	(0.000236)
ln(Age)	-1.445***	-1.247***	-2.078***
	(0.308)	(0.377)	(0.400)
ln(Population density)	-0.0553**	-0.0456*	-0.00256
	(0.0225)	(0.0262)	(0.0239)
Migration	0.00102**	0.000749	0.000925**
	(0.000404)	(0.000486)	(0.000404)
ln(Broadband access)	0.0993*	0.0369	0.0436
	(0.0601)	(0.0616)	(0.0767)
Train distance	-0.00259***	-0.00236***	-0.00147*
	(0.000738)	(0.000869)	(0.000822)
Rent	0.122***	0.144***	0.100***
	(0.00997)	(0.0129)	(0.0107)
ln(Single family home)	0.307**	1.108***	0.620***
	(0.131)	(0.241)	(0.139)
ln(Floor space)	0.104	0.0282	0.184**
	(0.0875)	(0.0982)	(0.0888)
ln(Building land price)	0.182***	0.194***	0.200***
	(0.0267)	(0.0336)	(0.0271)
Constant	6.866***	5.475**	4.644***
	(1.681)	(2.651)	(1.726)
Observations	401	293	324
R-squared	0.853	0.864	0.844

Note: The table presents the separate model results, based on a grouping of the data set. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 14: Regression parameter estimates for North-South comparison

ln (house price)	Pooled model All Counties	Group 1 North	Group 2 South
Gross value added	-0.00364***	-0.00314	-0.00268*
	(0.00127)	(0.00222)	(0.00137)
ln(Income)	0.224	-0.238	0.641***
	(0.160)	(0.394)	(0.171)
Tourist beds	0.00174***	0.00262***	0.000381
	(0.000237)	(0.000310)	(0.000393)
ln(Age)	-1.675***	-2.532***	-2.053***
	(0.403)	(0.700)	(0.548)
ln(Population density)	-0.0534*	-0.0784	-0.0774**
	(0.0282)	(0.0493)	(0.0329)
Migration	0.000973*	-0.00151	0.00149***
	(0.000518)	(0.00116)	(0.000501)
ln(Broadband access)	0.0952	-0.0126	0.0959
	(0.0774)	(0.114)	(0.0981)
Train distance	-0.00229**	-0.00400**	-0.00187*
	(0.000988)	(0.00200)	(0.000992)
Rent	0.114***	0.246***	0.0768***
	(0.0133)	(0.0307)	(0.0132)
ln(Single family home)	0.272	-0.191	0.196
	(0.189)	(0.332)	(0.228)
ln(Floor space)	0.118	-0.182	0.247**
	(0.118)	(0.211)	(0.119)
ln(Building land price)	0.186***	0.0393	0.223***
	(0.0332)	(0.0762)	(0.0311)
Constant	8.326***	19.67***	5.622**
	(2.346)	(4.886)	(2.281)
Observations	285	103	182
R-squared	0.817	0.811	0.862

Note: The table presents the separate model results, based on a grouping of the data set. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

In table 12 to 14 parameter estimates are reported to show differences to the pooled model. Group 1 and 2 respectively present the regional housing markets in Germany. The grouping is continued from table 14 and is based on chapter 3.2. As for all models the null hypothesis can be rejected, the coefficients of groups 1 and 2 in table 12 to 14 have higher explanatory power to describe the correlation of variables with the ln house price per square meter.

5. Conclusion

The focus of this study was the analysis of the German real estate market and its house price level. Existing literature has been used to identify drivers and associations that are correlated with regional house price levels. Subsequently, data for the drivers and house prices per square meter have been merged into one data set. In order to analyse the 401 German counties, the data came from Postbank Wohnatlas, Wegweiser-Kommune, INKAAR and Regionaldatenbank Deutschland. The data set has been analysed in a descriptive-narrative form as well as with a multivariate linear regression. The analyses concerned the German real estate market for 2017 and included 14 variables. The scale has been narrowed down to 401 counties, starting from a federal state level over an urban-rural, east-west and north-south comparison, in order to answer the research question: *How can regional house price differences be understood on county-level comparison in 2017?*

The main findings show that the location of a county determines the house prices in at least two ways. Firstly, by the geographical location itself and secondly by changing real estate, economic, socio-cultural and infrastructural drivers. Urban counties and counties in the western federal states show higher absolute house price levels, compared to rural and eastern counties. However, in the regression model of this thesis, the dummy coefficient for eastern counties is positive and significant. This might be reasoned due to the composition of the variable set. Average personal age and average personal income seem to be strongly correlated with regional house price levels, while the average personal age is higher in eastern counties. To my knowledge, a model with a positive correlation between eastern regions and house prices has not been reported in existing literature yet. Based on the given model it is confirmed that regional patterns exist. A disparity between urban and rural counties and between eastern and western counties is detected, while there is no evidence for a disparity between northern and southern counties. In general, the north has only a few but very high-priced residential counties. Merely these counties are Hamburg Bremen and the coastal regions have high priced housing values as outliers, while the south offers concentrations of high house price levels. However, the dominance statistic showed solely the medium importance of the location dummy variables.

With the regression results, the associations of drivers have become more transparent. From a social perspective, housing seekers receive an understanding of whether counties are more expensive than others, and what drivers can be associated with this. Possible explanations about the drivers in the German residential real estate market are delivered. The regional rent-to-price, price-to-income and purchase-power-to-price ratio, provide the foundation of a profound location decision for interested owner-occupiers. With these findings, policy makers and scientists receive new insights about the geographical differences that can explain different house price levels. The distance to train stations, broadband access and the share of single family and duplex homes positively affects the counties' regional house prices, which can be taken into account for scientific research on the accessibility of regions and spatial lags. Similarly, this accounts for the research.

Further research can consider the findings of this work. As the thesis partly used freely available data sources, the quality of the observation might be imperfect. Regulatory causal variables and recreation related variables could add value to the existing work of this thesis. Determinants for regulations are only stressed on a federal state level and in the existing literature, as including building permissions and land taxes delivered vague results. Including local economic and infrastructure subsidy programs do not deliver sound estimates. This might be reasoned in a homogenous distribution of financial aids – further research can bring new insights. Furthermore, increasing the number of observations, by comparing Germany on a municipal level would lead to more detailed insights. The municipality comparison would increase the number of observations from 401 counties to 3362. The focus on counties in this works goes back to the mentioned data limitations. Another interesting perspective to extend the research in this field would be to change the regression model. While this paper focuses on a separate distinction between counties, an aggregated view of counties with a spatial autocorrelation regression model would indicate the convergences of regions. Finally, concerning endogeneity, improvements for this model would benefit to decrease the correlation of explanatory variables, like rent and building land prices, with the error term.

References

Berndgen-Kaiser, A., Bläser, K., Fox-Kämper, R., Siedentop, S. and Zakrzewski, P., 2014. Demography-driven suburban decline? At the crossroads: mature single-family housing estates in Germany. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, *7*(3), pp.286-306.

Blanco, F., Martín, V. and Vazquez, G., 2016. Regional house price convergence in Spain during the housing boom. *Urban Studies*, *53*(4), pp.775-798.

Bleck, M. and Wagner, M., 2006. Stadt-Umland-Wanderung in Nordrhein-Westfalen - eine Meta-Analyse. *Raumforschung und Raumordnung*, 64(2), pp.104-115.

BMWI, 2020. *Das gesamtdeutsche Fördersystem für strukturschwache Regionen*. Available at: https://www.bmwi.de/Redaktion/DE/Downloads/G/gesamtdeutsches-foerdersystem.pdf?__blob=publicationFile&v=16 (Accessed: 06.07.20)

Bpb_a n.d.. *Kreis*. Available at: https://www.bpb.de/nachschlagen/lexika/recht-a-z/22488/kreis (Accessed: 30.06.20)

Bpb_b n.d. *Kaufkraft*. Available at: https://www.bpb.de/nachschlagen/lexika/lexikon-der-wirtschaft/19958/kaufkraft (Accessed: 03.07.20)

Borcherdt, C., 1977: Versorgungsorte und Versorgungsbereiche. *Zentralitätsforschungen in Nordwürttemberg*.

Brooks, C. and Tsolacos, S., 2010. *Real Estate Modelling and Forecasting*. 1st edn. United States of America: Cambridge University Press.

Buch, T., Hamann, S., Niebuhr, A. and Rossen, A., 2014. What makes cities attractive? The determinants of urban labour migration in Germany. *Urban Studies*, 51(9), pp.1960-1978.

Buhr, W., 1981. *Lectures on regional stagnation*. Baden-Baden: Nomos Verlagsgesellschaft.

Cadil, J., Mazouch, P., Musil, P. and Kramulova, J., 2014. True regional purchasing power: evidence from the Czech Republic. *Post-Communist Economies*, *26*(2), pp.241-256.

Capozza, D.R., Hendershott, P.H., Mack, C. and Mayer, C.J., 2002. *Determinants of real house price dynamics* (No. w9262). National Bureau of Economic Research.

Chen, M.C., Tsai, I.C. and Chang, C.O., 2007. House prices and household income: Do they move apart? Evidence from Taiwan. *Habitat International,* 31(2), pp.243-256.

Chow, G.C.,1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3), pp.591-605.

Deschermeier, P., Henger, R., Seipelt, B. and Voigtländer, M., 2016. *Zuwanderung, Wohnungsnachfrage und Baubedarfe. Aktualisierte Ergebnisse des IW Wohnungsbedarfsmodells* (18/2016). IW-Report.

Destatis, 2020. *Preisindizes für Wohnimmobilien nach Kreistypen*. Available at: https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Baupreise-Immobilienpreisindex/Tabellen/haeuserpreisindex-kreistypen.html;jsessionid=266F02EBE8F3314ED0F61EE5DFE94C8C.internet8721 (Accessed: 09.09.20)

Diekmann, F., 2019. *Die Landflucht der jungen Deutschen*. Available at: https://www.spiegel.de/wirtschaft/soziales/deutschland-die-extreme-landflucht-der-jungen-und-ihre-gruende-a-1292981.html (Accessed: 29.02.2020)

DiPasquale, D. and Wheaton, W.C., 1994. Housing market dynamics and the future of house prices. *Journal of Urban Economics*, 35(1), pp.1-27.

Efthymiou, D. and Antoniou, C., 2013. How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice*, 52, pp.1-22.

Eilinghoff, D., 2019. *So viel bringt eine Immobilie als Kapitalanlage*. Available at: https://www.finanztip.de/baufinanzierung/immobilie-kapitalanlage/ (Accessed: 22.07.20)

Einig, K., 2003. Baulandpolitik und Siedlungsflächenentwicklung durch regionales Flächenmanagement. *BBR: Bauland- und Immobilienbericht, 18(2),* pp.111-140.

Fabricious, M., 2015. *Warum der deutsche Wohnungsmarkt versagt*. Available at: https://www.welt.de/finanzen/immobilien/article141525947/Warum-der-deutsche-Wohnungsmarkt-versagt.html (Accessed: 29.02.2020)

Fingleton, B., 2006. A cross-sectional analysis of residential property prices: the effects of income, commuting, schooling, the housing stock and spatial interaction in the English regions. *Papers in Regional Science*, 85(3), pp.339-361.

Fischer, RA., 1950. Nig J Paediatr. *Statistical methods for research workers*. 5th edn. London: Oliver and Boyd.

Gallin, J., 2006. The long-run relationship between house prices and income: evidence from local housing markets. *Real Estate Economics*, 34(3), pp.417-438.

Goodman, A.C., 1988. An econometric model of housing price, permanent income, tenure choice, and housing demand. *Journal of Urban Economics*, *23*(3), pp.327-353.

Grabka, M.M. and Goebel, J., 2018. Einkommensverteilung in Deutschland: Realeinkommen sind seit 1991 gestiegen, aber mehr Menschen beziehen Niedrigeinkommen. *DIW-Wochenbericht*, 85(21), pp.449-459.

Gradmann, R., 1931. *Süddeutschland: Die einzelnen Landschaften.* 1st edn. Darmstadt: J. Engelhorns Verlag.

Hair, J.F., Anderson, R.E., Babin, B.J. and Black, W.C., 2010. Multivariate data analysis: a global perspective. 7th edn. Upper Saddle River, NJ (USA): Pearson.

Henger, R.M. and Oberst, C., 2019. Alterung der Gesellschaft im Stadt-Land Vergleich. *IW-Kurzbericht.* No. 16/2019.

Hiller, N. and Lerbs, O. W., 2016. Aging and urban house prices. *Regional Science and Urban Economics*, 60(1), pp.276-291.

Hoogstra, G.J., van Dijk, J. and Florax, R.J., 2017. Do jobs follow people or people follow jobs? A meta-analysis of Carlino–Mills studies. *Spatial Economic Analysis*, 12(4), pp.357-378.

Jacobs, J., 1969. The economy of cities. *Random House*, New York.

Kempermann, H., Ricci, A., Sagner, P. and Lang, T., 2019. *Wohnen in Deutschland 2019*. Available at: https://www.sparda-wohnen2019.de/fileadmin/files/pdf/Sparda-Studie_Wohnen_in_Deutschland_2019.pdf (Accessed: 17.05.2020)

Kulikauskas, D., 2015. Measuring fundamental house prices in the Baltic States: empirical approach. *Proceeding of 22nd Annual European Real Estate Society Conference,* pp.15-40.

Mankiw, N.G. and Weil, D.N., 1989. The baby boom, the baby bust, and the housing market. *Regional Science and Urban Economics*, 19(2), pp.235-258.

Marquaridt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*(3), pp.591-612.

Meen, G., 2002. The time-series behavior of house prices: a transatlantic divide?. *Journal of Housing Economics*, 11(1), pp.1-23.

Milosevich, M., 2015. *World Press Trends 2015*. Available at: https://wan-ifra.org/reports/2015/10/01/world-press-trends-report-2015 (Accessed: 20.05.2020)

Mulder, C.H., 2006. Home-ownership and family formation. *Journal of Housing and the Built Environment*, 21(3), pp.281-298.

OECD, 2016. *Housing Tenures*. 1.3. Paris: OECD Social Policy Division.

Papon, K., 2019. *So teuer ist Wohnen in Deutschland*. Available at: https://www.faz.net/aktuell/finanzen/meine-finanzen/mieten-und-wohnen/f-a-z-exklusiv-wo-wohnen-in-deutschland-am-teuersten-ist-16521147.html (Accessed: 29.02.2020)

Philiponnet, N. and Turrini, A., 2017. *Assessing house price developments in the EU* (No. 048). Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Pomogajko, K. and Voigtländer, M., 2012. Demografie und Immobilien: Der Einfluss der erwarteten Flächennachfrage auf die heutigen Wohnimmobilienpreise. *IW-Trends-Vierteljahresschrift zur empirischen Wirtschaftsforschung*, 39(2), pp.55-69.

Poterba, J.M., 1984. Tax subsidies to owner-occupied housing: an asset-market approach. *Quarterly Journal of Economics,* 99(4), pp.729-752.

Reichert, A.K., 1990. The impact of interest rates, income, and employment upon regional house prices. *The Journal of Real Estate Finance and Economics*, *3*(4), pp.373-391.

Renkow, M., 2003. Employment growth, worker mobility, and rural economic development. *American Journal of Agricultural Economics*, 85(2), pp.503-513.

Röhl, K.H., 2018. Regionale Konvergenz: Der ländliche Raum schlägt sich gut. *Wirtschaftsdienst*, *98*(6), pp.433-438.

Rosenthal, S.S. and Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics*, pp.2119-2171.

Shiller, R.J., 2007. *Understanding recent trends in house prices and home ownership* (No. w13553). National Bureau of Economic Research.

Statista, 2019. *Einwohner in den Metropolregionen in Deutschland 2017*. Available at: https://de.statista.com/statistik/daten/studie/603747/umfrage/einwohner-in-den-metropolregionen-in-deutschland/ (Accessed: 09.03.2020)

Takáts, E., 2012. Aging and house prices. *Journal of Housing Economics*, 21(2), pp.131-141.

Trading Economics, 2020. *Germany Home Ownership Rate*. Available at: https://tradingeconomics.com/germany/home-ownership-rate (Accessed: 04.07.20)

UND, 2020. *Heteroskedasticity*. Available at: https://www3.nd.edu/~rwilliam/stats2/l25.pdf (Accessed: 30.07.20)

Vermeulen, W. and Van Ommeren, J., 2009. Does land use planning shape regional economies? A simultaneous analysis of housing supply, internal migration and local employment growth in the Netherlands. *Journal of Housing Economics*, *18*(4), pp.294-310.

Watkins, C.A., 2001. The definition and identification of housing submarkets. *Environment and Planning A*, 33(12), pp.2235-2253.

Wiersma, 2020. *Singles in the City: The association between single-person households and property transaction prices.* Master thesis. University of Groningen, Groningen.

Wolf, A., 2016. *Das wirtschaftliche Süd-Nord-Gefälle in Deutschland: Aktuelle Befunde und Ursachen* (No. 99). Hamburg: HWWI.

World Bank, 2020. *Urban population (% of total population)*. Available at: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2018&name_desc=false&start=1994 (Accessed: 17.04.2020)

Xlstat, n.d. *Breusch-Pagan & White Heteroscedasticity tests in Excel*. Available at: https://help.xlstat.com/s/article/breusch-pagan--white-heteroscedasticity-tests-in-excel?language=en (Accessed: 30.07.20)

Zady, M. F, 2000. *Correlation and Simple Least Squares Regression*. Available at: https://www.westgard.com/lesson42.htm (Accessed: 05.08.20)

Zhang, C., Jia, S. and Yang, R., 2016. Housing affordability and housing vacancy in China: The role of income inequality. *Journal of Housing Economics*, *33*, pp.4-14.

Zumbro, T., 2014. The relationship between homeownership and life satisfaction in Germany. *Housing Studies*, *29*(3), pp.319-338.

Appendix A (Definitions)

Table A1: Descriptive Statistics of Regressor Variables

Name (English)	Name (German)	Abbreviation	Definition	Source
County	"Kreisfreie Stadt" or "Landkreis"	Region, county	Officially defined territorial entity	Bpb_a (n.y.)
Frankfurt General Newspaper	Frankfurter Allgemeine Zeitung	FAZ	FAZ is a centre-right, liberal-conservative German newspaper, founded in 1949	Milosevich (2015)
FRG	BRD	-	Federal Republic of Germany	-
GDR	DDR	-	German Democratic Republic	-
Gross Domestic Product	Bruttoinlands-produkt	GDP	Gross Domestic Product (GDP) is the monetary value of all finished goods and services made within a country during a specific period	-
Improvement of the regional economic structure	Verbesserung der regionalen Wirtschaftsstruktur	GRW	German federal economic growth program for regional disparities	-

The table supplies definitons, abrevaiations and translation of used terms

Table A2**:** Defintions and units of controlled variables

Category	in Reg.	with LN	Variable	Unit	Description	Source	# of regions
dependent	X	X	houseprice (all dwellings)	€/m²	median mean by the price of all existing sold dwellings divded by the total sold m² in the county	1	401
economic			GDP	€/ inhabitant	the gross domestic product was x euros per inhabitant	4	3362
	X		GVA	€/ inhabitant	the gross value added was y euros per inhabitant	4	3362
	X	X	Inc	€/ inhabitant	anual average income of a region per inhabtitant	3	401
	X		Tourist	/1000 inhabtitants	The number of beds for tourist per 1.000 inhabtiants	3	401
			purchasing power	%	the purchasing power compares a counties spending power to the german average (=100)	3	401
socio-cultural	X	X	Age	years	The average age of all people is x years	2	3362
	X	X	Pdensity	ratio	Inhabitants per km² of settlement and traffic area	3	3362
	X		Mig	/1000 inhabtitants	There were x people moving for every 1,000 people in the existing population more than	3	3362

					they left (or vice versa if the balance is negative)		
			high education	% of population	x% of employees subject to social security insurance have an academic professional qualification	2	3362
infrastructral			recreation	km²/1000 inh	the recreation area per 1,000 inhabitants is x km²	2	3362
			highway	minutes	Average car travel time to the next motorway junction in minutes	3	3362
			airport	minutes	Car travel time to the nearest international airport in minutes	3	3362
	X		Train	minutes	Car travel time to the nearest train station (long-distance traffic) in minutes	3	3362
	X	X	Bb	% of households	Share of households with a broadband connection of at least 50 mbit / s in all households in%	4	3362
real estate	X		rent	€/m²	In the region, the average net rent price was x€/m²	1	401
	X		Sfhd	%	X% of the dwellings in the municipal housing stock consist of dwellings in one and two-family houses	2	3362
	X	X	Sqmliv	m²/person	An average of x square meters of living space is available per person. Without the living space in dormitories	2	3362
regulatory			land tax	%	The county y charges a x% land tax on the buying price for a property	2	3362
	X	X	Blp	€/m²	Average purchase values for building land in € per m²	3	401
	X		Building permissions	/1000 inhabtitants	Regional building permissions for new dwelling units of any residential kind		401
	X		Economic aid	€/ inhabitant	Granted joint task: “Improvement of the regional economic structure"		401
	X		Infrastructure aid	€/ inhabitant	Contains program areas: renovation development and urban monument protection		401

Note: The abreviation of sources are as follows: 1=Postbank Wohnatlas 2019; 2=Wegweiser-Kommune; 3=INKAAR; 4=Regionaldatenbank Deutschland

Table A3**:** federal states with surface area and population

Federal state	Surface (km²)	Population	Population/km²
Bavaria	70,542	12,997,204	184.25
Lower Saxony	47,710	7,962,775	166.90
Baden-Württemberg	35,748	11,023,425	308.36
Northrine-Westphalia	34,112	17,912,134	525.09
Brandenburg	29,654	2,504,040	84.44
Mecklenburg Western Pomerania	23,295	1,611,119	69.16
Hessen	21,116	6,142,161	290.88
Saxony-Anhalt	20,454	2,223,081	108.69
Rhineland Palatinate	19,858	4,073,679	205.14
Saxony	18,450	4,081,308	221.21
Thuringia	16,202	2,151,205	132.77
Schleswig Holstein	15,804	2,889,821	182.85
Saarland	257	994,187	3866.78
Berlin	891	3,613,495	4055.00
Hamburg	755	1,830,584	2424.33
Bremen	419	681,032	1623.98
Germany	357,582	82,691,250	231.25

Appendix B (Descriptive Analysis and Ratios)

Table A4: Descriptive Statistics of Regressor Variables

Variable	Obs	Mean	Std.Dev.	Min	Max
Gross value added	401	62.438	10.887	46.73	147.49
Ln(Income)	401	10.014	.111	9.704	10.569
Tourist beds	401	41.867	49.249	3.7	405.7
ln(Age)	401	3.795	.044	3.684	3.916
ln(Population density)	401	5.623	1.105	3.587	8.452
Migration	401	-10.361	29.725	-149.37	62.66
ln(Broadband access)	401	4.316	.225	3.311	4.601
Train distance	401	21.925	15.379	0	79
Rent	401	7.256	1.802	4.63	17.36
ln(Single family home)	401	4.411	.14	3.914	4.565
ln(Floor space)	401	3.833	.099	3.555	4.212
ln(Building land price)	401	4.732	.894	2.442	7.795

Note: The table summarizes the mean, standard deviation, minimum and maximum of the independent variables for the regression models.

Table A5: Descriptive statistics about regional characteristics

			Mean			
Variable	Urban	Rural	East	West	North	South
Gross value added	64.912	61.526	54.399	64.349	59.112	65.164
Ln(Income)	9.97	10.03	9.893	10.042	9.951	10.073
Tourist beds	24.448	48.287	45.416	41.023	59.195	41.148
ln(Age)	3.77	3.805	3.849	3.783	3.816	3.78
ln(Population density)	7.091	5.082	5.063	5.756	5.182	5.613
Migration	22.078	-22.319	-11.36	-10.124	-13.728	-10.245
ln(Broadband access)	4.524	4.239	4.101	4.367	4.252	4.347
Train distance	8.593	26.84	26.351	20.873	21.961	22.621
Rent	7.898	7.019	5.991	7.557	6.556	7.94
ln(Single family home)	4.225	4.48	4.378	4.419	4.441	4.431
ln(Floor space)	3.84	3.831	3.848	3.83	3.846	3.837
ln(Building land price)	5.348	4.505	3.832	4.946	4.252	5.041

Note: The table sums the means of the regressor variables and groups the output by their location.

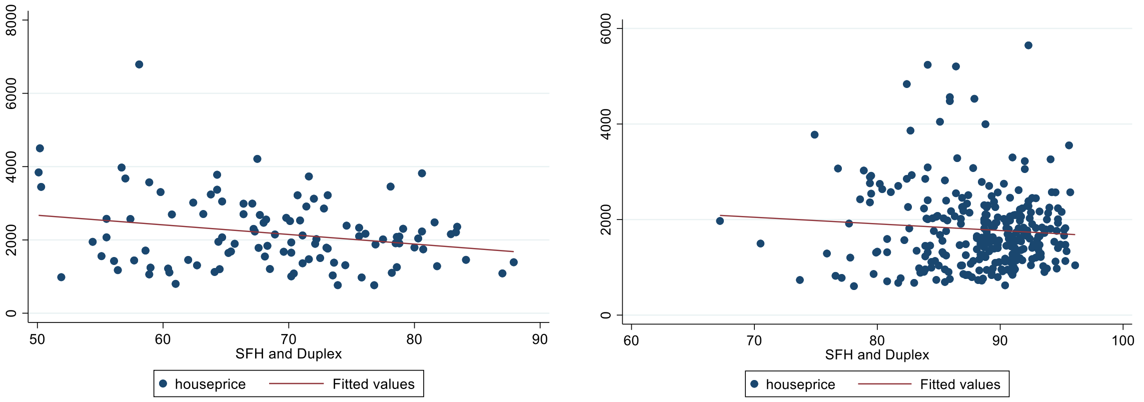


Figure A1: Relation between Single Family and Duplex rate per county with house prices n m² in 2017. The left figure accounts for urban regions, the right figure accounts for rural regions.

Table A6: regional house prices on the categorized zones

House price	Obs	Mean	Std.Dev.	Min	Max
Urban	108	2176.981	966.089	763.93	6789.44
Rural	293	1793.012	830.514	606.42	5646.81
West	324	2034.148	878.848	734.02	6789.44
East	77	1316.918	643.052	606.42	3676.41
South	182	2238.891	932.677	755.95	6789.44
North	103	1661.304	772.965	675.77	5646.81

Note: The highest house prices per square meter in rural counties are Nordfriesland (with the corresponding islands) and the rural county of Munich.

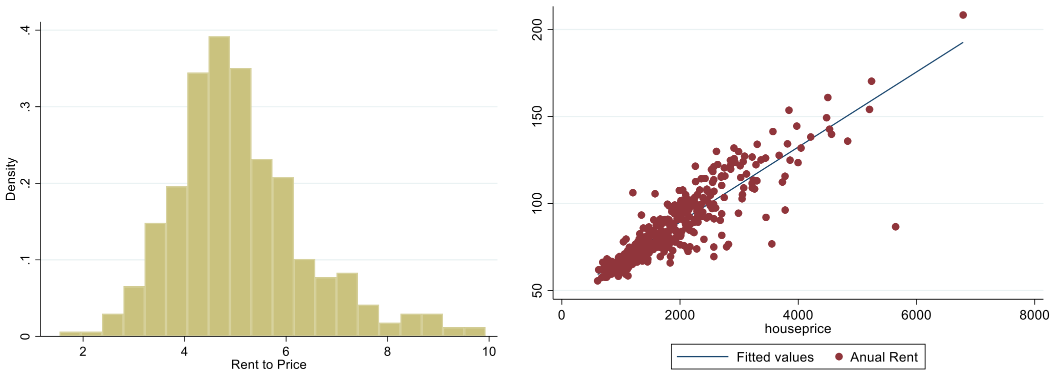


Figure A2: Histogram of the Rent-to-Price ratio and the scatter plot of the annual rent compared to the house price in square meter.

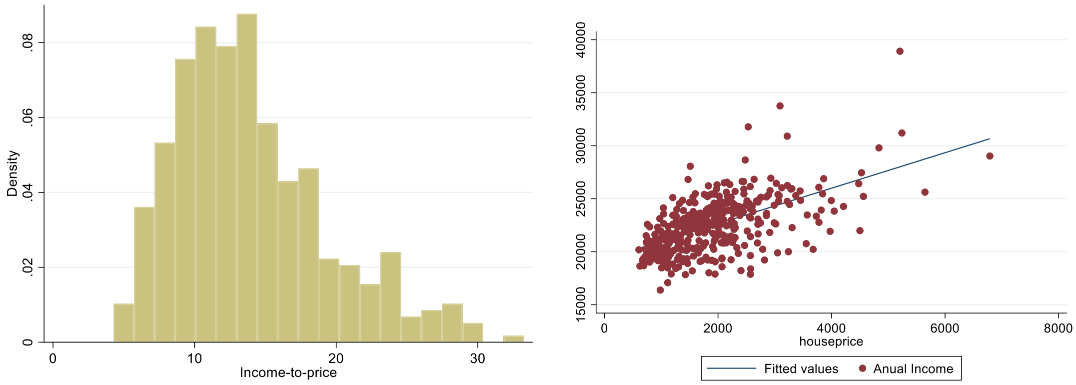


Figure A3: Histogram of the Income-to-Price ratio and the scatter plot of the annual income compared to the house price in square meter.

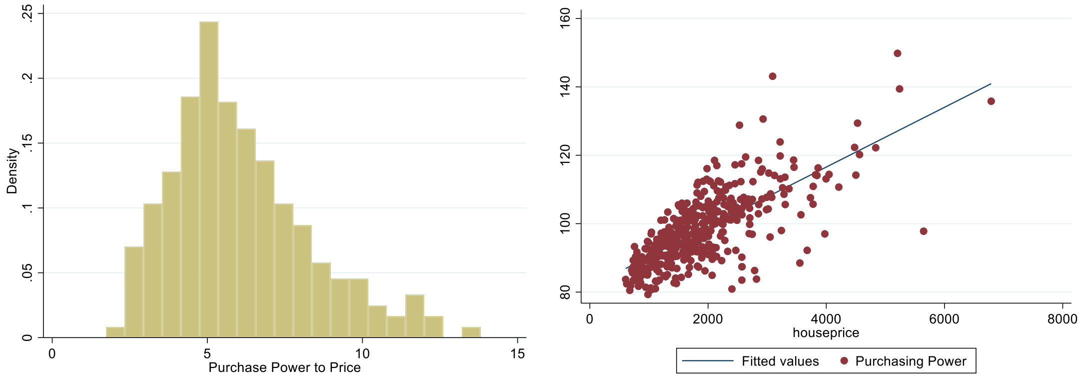


Figure A4: Histogram of Purchase-Power-to-Price ratio and the scatter plot of the purchase power ratio compared to the house price in square meter.

Table A7: Real Estate Ratios as mean value sorted by spatial characteristic

Regions	Obs	RtP	ItP.	PptP
Urban	108	4.76	11.50	5.424
Rural	293	5.207	14.86	6.553
East	77	6.189	18.03	8.968
West	324	4.825	12.98	5.603
North	103	5.356	14.98	6.99
South	182	4.585	12.17	5.071

Table A8: Overview on Real Estate Ratios over the 401 German counties

County	House price	RtP	ItP	PPtP
Aichach-Friedberg	1935	4.17	12.190	5.15
Alb-Donau-Kreis	3056	3.44	8.310	3.56
Altenburger Land	2095	4.61	11.300	4.99
Altenkirchen (Westerwald)	780	7.78	25.450	10.86
Altmarkkreis Salzwedel	1025	6.43	20.970	9.02
Altötting	837	7.10	24.290	10.24
Alzey-Worms	2015	4.06	11.570	5.14
Amberg	1720	4.82	13.290	5.87
Amberg-Sulzbach	2101	4.10	10.950	4.84
Ammerland	1383	5.21	16.520	6.78
Anhalt-Bitterfeld	2237	3.72	10.080	4.36
Ansbach	738	8.82	26.710	11.32

Aschaffenburg	1909	4.09	11.370	5.36
Augsburg	2681	3.37	9.240	4.00
Aurich	3050	3.37	6.520	3.15
Bad Dürkheim	3553	2.16	5.840	2.49
Bad Kissingen	1831	4.74	14.310	6.12
Bad Kreuznach	1391	5.01	16.600	6.83
Bad Tölz-Wolfratshausen	1819	4.43	12.200	5.29
Baden-Baden, Stadtkreis	3862	3.24	6.960	3.01
Bamberg	3219	3.47	9.590	3.85
Barnim	2707	3.02	8	3.76
Bautzen	1629	5.07	12.520	5.76
Bayreuth	957	6.84	21.260	9.09
Berchtesgadener Land	2124	3.47	10.030	4.71
Bergstraße	2747	3.77	8.260	3.53
Berlin, Stadt	2015	4.70	11.850	5.24
Bernkastel-Wittlich	3676	3.47	5.5	2.51
Biberach	1477	4.85	14.890	6.47
Bielefeld, Stadt	2063	4.59	11.730	5.12
Birkenfeld	1652	5.39	13.730	5.76
Bochum, Stadt	888	6.82	24.360	10.00
Bodenseekreis	1441	5.65	13.960	6.49
Bonn, Stadt	3027	3.80	8.730	3.79
Borken	2556	4.63	9.1	4.39
Bottrop, Stadt	1618	4.77	13.980	5.74
Brandenburg a. d. Havel, Stadt	1473	5.31	13.790	6.35
Braunschweig, Stadt	1547	4.75	11.750	5.45
Breisgau-Hochschwarzwald	2149	4.49	10.440	4.95
Bremen, Stadt	2704	4.08	9.120	3.93
Bremerhaven, Stadt	2015	5.04	10.860	4.67
Burgenlandkreis	1032	6.50	18.080	7.87
Böblingen	708	8.69	27.530	12.20
Börde	2853	4.37	8.820	4.15
Calw	882	7.20	23.650	10.24
Celle	1555	5.83	15.200	6.76
Cham	1145	6.54	18.230	8.29
Chemnitz, Stadt	1288	5.40	17.330	7.05
Cloppenburg	1060	5.87	19.260	8.45
Coburg	1667	4.17	12.610	5.11
Cochem-Zell	1876	3.89	13.540	5.76
Coesfeld	1833	3.59	12.410	5.17
Cottbus, Stadt	1563	4.95	15.220	6.63
Cuxhaven	1362	5.19	14.800	6.58
Dachau	2283	3.24	9.680	4.17
Dahme-Spreewald	4563	3.06	5.520	2.63
Darmstadt, Wissenschaftsstadt	2091	4.52	9.950	4.45
Darmstadt-Dieburg	2990	4.34	7.6	3.60
Deggendorf	2181	4.63	10.950	4.93
Delmenhorst, Stadt	2211	3.80	10.190	4.36
Dessau-Roßlau, Stadt	1391	5.77	13.900	6.26
Diepholz	764	8.92	26.180	11.53
Dillingen ad-Donau	1647	4.72	14.300	6.02
Dingolfing-Landau	1930	4.35	12.730	5.16
Dithmarschen	2161	3.92	11.370	4.80
Donau-Ries	1808	4.18	12.650	5.08
Donnersbergkreis	2169	4.05	11.610	4.77
Dortmund, Stadt	1217	5.86	17.810	7.67
Dresden, Stadt	1455	5.85	13.290	6.26
Duisburg, Stadt	2071	4.35	9.360	4.41
Düren	1114	6.54	15.330	7.47
Düsseldorf, Stadt	1335	5.68	15.840	7.16
Ebersberg	3447	3.66	7.460	3.44

Eichsfeld	4529	3.15	6.060	2.86
Eichstätt	1139	5.81	17.580	7.29
Eifelkreis Bitburg-Prüm	3262	3.32	7.490	3.39
Eisenach, Stadt	2001	3.81	11.130	4.55
Elbe-Elster	1125	6.35	17.370	7.80
Emden, Stadt	1110	5.61	17.120	7.30
Emmendingen	1087	7.32	17.380	8.15
Emsland	2574	4.16	9	3.99
Ennepe-Ruhr-Kreis	1477	4.92	14.900	6.13
Enzkreis	1498	5.04	16.440	7.04
Erding	1938	4.94	13.120	5.65
Erfurt, Stadt	3996	3.09	6.210	2.83
Erlangen	1895	4.66	10.120	4.80
Erlangen-Höchstadt	3220	3.94	7.680	3.72
Erzgebirgskreis	2568	3.88	10.180	4.58
Essen, Stadt	823	7.22	24.570	9.93
Esslingen	1557	5.43	13.040	6.23
Euskirchen	2891	4.29	8.770	3.98
Flensburg, Stadt	1654	4.66	13.160	5.85
Forchheim	1893	4.63	10.120	4.75
Frankenthal, kreisfreie Stadt	2374	3.90	10.210	4.50
Frankfurt (Oder), Stadt	1794	5.07	12.420	5.27
Frankfurt am Main, Stadt	1839	3.78	9.790	4.75
Freiburg i. Breisgau, Stadtkreis	4501	3.58	4.880	2.54
Freising	3975	3.64	5.510	2.44
Freudenstadt	4047	3.26	5.880	2.83
Freyung-Grafenau	1318	6.26	18.600	7.84
Friesland	903	7.03	23.510	9.64
Fulda	2136	3.39	9.830	4.35
Fürstenfeldbruck	1677	4.99	13.450	5.56
Fürth	4479	3.33	5.890	2.73
Garmisch-Partenkirchen	2603	3.74	9.340	4.14
Gelsenkirchen, Stadt	3776	3.06	6.9	2.80
Gera, Stadt	983	6.97	16.660	8.07
Germersheim	801	7.64	24.530	10.93
Gießen	1980	4.65	11.460	5.09
Gifhorn	2059	4.80	10.120	4.51
Goslar	1794	4.66	13.330	5.86
Gotha	772	8.63	27.300	11.70
Grafschaft Bentheim	1134	6.14	17.210	7.63
Greiz	1472	5.03	13.940	5.90
Groß-Gerau	958	6.08	21.420	8.89
Göppingen	2263	4.97	9.270	4.55
Görlitz	2078	4.86	11.590	5.06
Göttingen	676	8.51	28.310	11.92
Günzburg	1347	6.93	15.490	6.91
Gütersloh	1905	4.45	12.600	5.26
Hagen	1758	4.58	14.440	5.85
Halle (Saale), Stadt	1219	5.48	16.500	7.45
Hamburg, Hansestadt	1425	5.21	12.510	5.80
Hameln-Pyrmont	4212	3.28	5.760	2.63
Hamm, Stadt	993	6.60	22.120	9.52
Harburg	1257	5.84	14.660	6.75
Harz	2148	4.75	11.920	5.45
Havelland	887	7.21	22.560	9.59
Haßberge	1926	4.40	10.500	4.80
Heidekreis	1409	4.86	15.730	6.49
Heidelberg, Stadtkreis	1061	6.69	19.890	8.64
Heidenheim	3572	3.96	6.560	2.87
Heilbronn	1617	5.40	13.870	6.32
Heilbronn, Stadtkreis	2294	4.58	10.650	4.79

Heinsberg	2532	4.72	12.550	5.09
Helmstedt	1302	5.82	15.610	6.92
Herford	1265	5.54	18.440	7.99
Herne, Stadt	1255	5.55	18.140	7.66
Hersfeld-Rotenburg	1175	6.11	15.230	7.10
Herzogtum Lauenburg	1222	5.54	17.330	7.42
Hildburghausen	1796	5.03	13.060	5.92
Hildesheim	929	7.08	22.140	9.34
Hochsauerlandkreis	1249	5.95	17.060	7.69
Hochtaunuskreis	1038	6.23	23.240	9.34
Hof	3092	4.11	10.910	4.63
Hohenlohekreis	1084	5.47	18.820	8.34
Holzminden	1861	5.13	12.980	5.65
Höxter	734	7.85	29.310	12.17
Ilm-Kreis	952	6.21	23.160	9.72
Ingolstadt	1313	5.64	14.890	6.48
Jena, Stadt	3818	3.52	6.260	2.99
Jerichower Land	2071	4.91	8.970	4.33
Kaiserslautern	765	8.34	27.050	11.43
Kaiserslautern, kreisfreie Stadt	1173	6.25	18.080	7.95
Karlsruhe	1380	6.23	13.820	6.40
Karlsruhe, Stadtkreis	2262	4.42	10.540	4.75
Kassel	3018	4.03	7.490	3.46
Kassel, documenta-Stadt	1564	4.82	14.220	6.23
Kaufbeuren	1950	4.67	9.770	4.76
Kelheim	2083	4.33	10.920	4.79
Kempten (Allgäu)	2386	3.98	9.810	4.24
Kiel, Landeshauptstadt	2515	4.17	8.730	4.10
Kitzingen	2230	4.43	8.620	4.00
Kleve	1856	4.24	12.710	5.27
Koblenz, kreisfreie Stadt	1438	5.42	14.490	6.29
Konstanz	2301	4.21	9.270	4.44
Krefeld, Stadt	3069	4.04	8.070	3.52
Kronach	1506	5.55	14.340	6.53
Kulmbach	1204	6.02	19.450	7.83
Kusel	1391	5.22	17.160	6.89
Kyffhäuserkreis	1133	5.60	18.320	7.80
Köln, Stadt	624	9.93	29.900	13.22
Lahn-Dill-Kreis	3306	4.06	6.730	3.20
Landau i.d. Pfalz	1606	4.91	14.440	5.96
Landkreis Rostock	2231	4.58	9.620	4.49
Landsberg am Lech	2788	2.70	7.250	3.10
Landshut	3301	3.43	7.860	3.44
Leer	3455	2.66	7.190	3.37
Leipzig	2571	2.92	7.630	3.25
Leipzig, Stadt	1229	5.38	17.240	7.37
Leverkusen, Stadt	1948	4.27	9.180	4.43
Lichtenfels	2022	4.71	10.660	5.02
Limburg Weilburg	1731	4.17	12.890	5.31
Lindau (Bodensee)	1651	4.64	13.240	5.83
Lippe	2425	4.72	10.980	4.38
Ludwigsburg	1279	5.57	17.740	7.46
Ludwigshafen a. Rhein,	2917	4.31	8.830	3.98
Ludwigslust-Parchim	1788	5.57	10.890	5.20
Lörrach	976	6.94	20.610	8.71
Lübeck, Hansestadt	2636	4.64	10.170	4.53
Lüchow Dannenberg	2332	4.19	8.730	3.94
Lüneburg	975	6.13	21.690	8.84
Magdeburg, Landeshauptstadt	2255	4.60	9.250	4.33
Main-Kinzig-Kreis	1202	6.08	15.660	7.33
Main-Spessart	2026	4.78	11.620	5.10

Main-Tauber-Kreis	1554	4.83	15.090	6.36
Main-Taunus-Kreis	1631	4.75	14.130	6.05
Mainz, kreisfreie Stadt	2931	4.22	9.180	4.46
Mainz-Bingen	2912	4.53	7.490	3.68
Mannheim, Stadtkreis	2109	4.65	12.560	5.62
Mansfeld-Südharz	2696	4.27	7.720	3.60
Marburg-Biedenkopf	946	6.67	20.240	8.60
Mayen-Koblenz	2144	4.72	9.710	4.22
Mecklenburgische Seenplatte	1612	4.65	13.720	6.07
Meißen	1450	4.58	13.100	5.80
Memmingen	1316	5.24	15.860	6.72
Merzig-Wadern	2479	4.06	11.550	4.23
Mettmann	1820	4.24	10.550	4.81
Miesbach	1916	4.92	13.280	5.87
Miltenberg	4836	2.81	6.160	2.53
Minden-Lübbecke	1740	4.59	13.400	5.58
Mittelsachsen	1213	6.02	19.320	7.81
Märkisch-Oderland	861	7.18	23.620	9.98
Märkischer Kreis	1694	4.81	11.970	5.42
Mönchengladbach, Stadt	1203	5.78	20.860	8.39
Mühldorf a. Inn	1310	6.08	15.800	7.13
Mülheim an der Ruhr, Stadt	2093	4.31	11.360	4.90
München	1645	5.05	14.050	6.45
München, Landeshauptstadt	5240	3.25	5.950	2.66
Münster, Stadt	6789	3.07	4.270	2.00
Neckar-Odenwald-Kreis	2857	4.19	8.260	3.70
Neu-Ulm	1386	5.82	16.090	6.97
Neuburg-Schrobenhausen	2334	4.62	10.360	4.50
Neumarkt i.d. O-Pf	2446	4.17	9.720	4.27
Neumünster, Stadt	2250	4.04	10.660	4.44
Neunkirchen	1283	6.07	15.020	6.59
Neustadt a.d. Aisch-Bad W.	1120	6.27	18.020	8.02
Neustadt ad-Waldnaab	1666	4.49	14.050	5.80
Neustadt a.d.W., Stadt	1362	4.92	16.750	6.71
Neuwied	2157	4.28	12.300	5.08
Nienburg (Weser)	1556	4.87	14.780	6.23
Nordfriesland	1148	5.85	18.600	7.86
Nordhausen	5647	1.53	4.530	1.73
Nordsachsen	692	9.57	26.990	11.83
Nordwestmecklenburg	1038	6.35	18.940	8.41
Northeim	2069	3.66	9.340	4.10
Nürnberg	936	6.73	23.830	9.91
Nürnberger Land	2703	4.28	8.350	3.87
Oberallgäu	2301	4.05	11.160	4.82
Oberbergischer Kreis	2542	3.82	9.660	4.05
Oberhausen, Stadt	1261	5.81	18.400	7.74
Oberhavel	1307	5.68	14.210	6.70
Oberspreewald-Lausitz	1950	4.87	10.880	4.97
Odenwaldkreis	906	7.12	21.600	9.58
Oder-Spree	1402	5.53	15.590	6.57
Offenbach	1591	5.08	12.750	5.62
Offenbach am Main, Stadt	2360	4.84	10.040	4.71
Oldenburg	2571	4.71	6.950	3.51
Oldenburg (Oldenburg), Stadt	1789	4.38	13	5.56
Olpe	2359	4.22	8.410	4.29
Ortenaukreis	1511	5.00	18.570	6.96
Osnabrück	2008	4.60	11.650	5.04
Osnabrück, Stadt	1452	5.11	15.960	6.48
Ostalbkreis	1932	4.99	10.870	4.98
Ostallgäu	1947	4.92	12.500	5.39
Osterholz	2490	3.72	9.790	4.06

Ostholstein	1536	5.38	14.860	6.58
Ostprignitz-Ruppin	2705	3.48	8.530	3.69
Paderborn	1465	5.00	12.890	5.63
Passau	1761	4.75	12.360	5.29
Peine	2168	3.47	9.860	4.65
Pfaffenhofen a.d. Ilm	1263	5.98	17.240	7.63
Pforzheim, Stadtkreis	3223	3.38	8.140	3.51
Pinneberg	1971	5.03	11.570	4.99
Pirmasens, kreisfreie Stadt	2096	5.01	11.750	5.30
Plön	765	7.57	25.620	11.25
Potsdam, Stadt	2235	3.99	10.280	4.49
Potsdam-Mittelmark	3241	3.50	6.170	3.02
Prignitz	2286	4.18	9.910	4.44
Rastatt	860	7.19	22.480	9.62
Ravensburg	2027	4.68	11.430	5.20
Recklinghausen	2396	4.31	9.850	4.33
Regen	1290	5.62	15.830	7.26
Regensburg	1004	6.40	21.240	8.87
Region Hannover	3779	2.55	6.020	2.94
Regionalverband Saarbrücken	2091	4.70	10.310	4.89
Rems-Murr-Kreis	1570	5.31	12.480	5.80
Remscheid, Stadt	2760	4.20	9.1	4.07
Rendsburg-Eckernförde	1207	5.95	18.690	8.05
Reutlingen	1784	4.66	13.250	5.68
Rhein-Erft-Kreis	2551	4.40	9.610	4.20
Rhein-Hunsrück-Kreis	1910	5.19	11.770	5.48
Rhein-Kreis Neuss	1274	5.21	18.820	7.58
Rhein-Lahn-Kreis	2021	4.80	12.480	5.57
Rhein-Neckar-Kreis	1378	5.28	16.260	6.94
Rhein-Pfalz-Kreis	2232	4.61	11.260	4.90
Rhein-Sieg-Kreis	1977	4.60	13.110	5.72
Rheingau-Taunus-Kreis	2004	4.71	11.930	5.32
Rheinisch-Bergischer Kreis	2175	4.61	11.440	5.17
Rhön-Grabfeld	1982	4.86	13.270	5.86
Rosenheim	1403	5.02	16.480	6.61
Rostock	3732	3.01	6.250	2.88
Rotenburg (Wümme)	2575	3.54	7.130	3.39
Roth	1403	5.30	16.180	6.81
Rottal-Inn	2094	4.20	11.740	4.99
Rottweil	1764	4.33	13.230	5.38
Saale-Holzland-Kreis	1559	5.27	16.130	6.80
Saale-Orla-Kreis	1075	6.27	18.720	8.14
Saalekreis	843	7.91	23.730	9.94
Saalfeld-Rudolstadt	837	7.83	24.660	10.73
Saarlouis	981	7.10	20.700	8.71
Saarpfalz-Kreis	1470	5.27	14.330	6.37
Salzgitter, Stadt	1470	5.45	15.490	6.83
Salzlandkreis	976	6.95	20.290	9.14
Schaumburg	676	9.16	28.530	12.16
Schleswig-Flensburg	1133	6.04	19.280	8.42
Schmalkalden-Meiningen	1650	4.71	13.850	5.73
Schwabach	958	6.96	21.580	9.06
Schwalm-Eder-Kreis	2208	4.47	11.920	5.08
Schwandorf	1192	5.46	18.080	7.88
Schwarzwald-Baar-Kreis	1817	4.21	12.150	5.27
Schweinfurt	1593	5.63	15.440	6.56
Schwerin	1783	4.19	11.680	5.36
Schwäbisch Hall	1709	4.65	11.200	5.22
Segeberg	1666	5.47	14.140	6.05
Siegen Wittgenstein	2068	4.93	11.210	5.08
Sigmaringen	1530	5.38	15.360	6.54

Soest	1598	5.00	14.280	6.07
Solingen, Klingenstadt	1485	5.00	15.310	6.55
Sonneberg	1677	4.95	13.930	6.06
Speyer, kreisfreie Stadt	903	7.34	23.050	9.65
Spree-Neiße	2389	4.32	10.030	4.50
St Wendel	841	7.26	24.100	10.39
Stade	1338	5.25	16.060	7.39
Starnberg	1749	5.19	13.100	5.97
Steinburg	5205	2.96	7.470	2.88
Steinfurt	1275	6.02	16.970	7.39
Stendal	1545	4.87	14.430	6.07
Stormarn	1040	6.08	18.350	7.76
Straubing	2457	4.37	10.360	4.77
Straubing-Bogen	2304	3.88	9.660	4.40
Stuttgart, Stadtkreis	1043	7.48	22.580	9.35
Städteregion Aachen	3843	4.00	6.620	2.97
Suhl, Stadt	1907	4.94	10.750	4.89
Sächsische Osterzgebirge	1102	6.15	19.540	8.51
Sömmerda	1309	5.35	15.710	6.68
Südliche Weinstraße	960	7.19	20.350	8.88
Südwestpfalz	1824	4.63	13.190	5.61
Teltow-Fläming	975	6.30	23.720	9.73
Tirschenreuth	1652	4.91	12.030	5.48
Traunstein	1120	5.22	19.980	8.12
Trier, kreisfreie Stadt	2511	3.93	9.870	4.13
Trier-Saarburg	2469	4.33	8.340	3.73
Tuttlingen	2027	4.14	11.310	4.60
Tübingen	1814	4.99	13.940	6.01
Uckermark	2852	4.27	8.2	3.61
Uelzen	1113	5.98	16.600	7.48
Ulm, Stadtkreis	1183	6.08	18.010	7.69
Unna	3124	3.75	8.330	3.65
Unstrut-Hainich-Kreis	1328	5.49	16.180	7.20
Unterallgäu	795	8.12	23.930	10.47
Vechta	2451	3.74	10.860	4.16
Verden	2007	4.04	11.500	4.80
Viersen	1688	5.00	14	6.14
Vogelsbergkreis	1621	5.00	14.250	6.23
Vogtlandkreis	1043	6.10	21.720	8.73
Vorpommern-Greifswald	606	9.16	33.270	13.80
Vorpommern-Rügen	2404	3.30	7.570	3.37
Vulkaneifel	2820	2.72	6.810	2.97
Waldeck-Frankenberg	1122	5.77	19.320	8.29
Waldshut	1095	5.84	19.840	8.31
Warendorf	1815	5.44	13.880	6.13
Wartburgkreis	1483	5.12	15.800	6.68
Weiden i d O-Pf	843	7.74	24.840	10.30
Weilheim Schongau	1763	4.55	12.450	5.60
Weimar, Stadt	3078	3.54	8.210	3.50
Weimarer Land	1684	5.22	11.330	5.18
Weißenburg-Gunzenhausen	1167	5.93	16.960	7.36
Werra-Meißner-Kreis	1931	4.03	11.940	4.96
Wesel	917	6.54	22.090	9.48
Wesermarsch	1455	5.38	15.150	6.82
Westerwaldkreis	1180	5.56	18.360	7.76
Wetteraukreis	1376	5.37	18.050	7.09
Wiesbaden, Landeshauptstadt	2271	4.34	10.390	4.65
Wilhelmshaven, Stadt	3370	3.71	7.5	3.27
Wittenberg	1006	6.71	18.360	8.99
Wittmund	723	8.85	27.980	11.85
Wolfenbüttel	2573	2.70	8.320	3.40

Wolfsburg, Stadt	1419	5.58	15.800	7.16
Worms, kreisfreie Stadt	2041	4.99	12.130	5.50
Wunsiedel i Fichtelgebirge	1743	5.13	11.770	5.40
Wuppertal, Stadt	756	7.62	29.870	12.34
Würzburg	1244	6.15	16.990	7.62
Zollernalbkreis	2989	3.16	8.150	3.48
Zweibrücken, kreisfreie Stadt	1600	5.48	15.910	6.50
Zwickau	1456	4.67	13.850	6.37
Observations	401			

The variable "house price" is the median value for residential housing measured in Euro per square meter in the respected county.

Appendix C (OLS Results and Tests)

Table A9: Regression of counties fixed effects

County	Coef	Std. Err.
Aichach-Friedberg	0.457	(0.279)
Alb-Donau-Kreis	0.0794	(0.279)
Altenburger Land	-0.908***	(0.279)
Altenkirchen (Westerwald)	-0.635**	(0.279)
Altmarkkreis Salzwedel	-0.838***	(0.279)
Altötting	0.0405	(0.279)
Alzey-Worms	-0.117	(0.279)
Amberg	0.0822	(0.279)
Amberg-Sulzbach	-0.336	(0.279)
Ammerland	0.145	(0.279)
Anhalt-Bitterfeld	-0.964***	(0.279)
Ansbach	-0.0505	(0.242)
Aschaffenburg	0.177	(0.242)
Augsburg	0.403	(0.242)
Aurich	0.608**	(0.279)
Bad Dürkheim	-0.0551	(0.279)
Bad Kissingen	0.509*	(0.279)
Bad Kreuznach	-0.33	(0.279)
Bad Tölz-Wolfratshausen	-0.0616	(0.279)
Baden-Baden, Stadtkreis	0.691**	(0.279)
Bamberg	0.196	(0.242)
Barnim	-0.172	(0.279)
Bautzen	-0.704**	(0.279)
Bayreuth	-0.191	(0.242)
Berchtesgadener Land	0.351	(0.279)
Bergstraße	0.0409	(0.279)
Berlin, Stadt	0.642**	(0.279)
Bernkastel-Wittlich	-0.27	(0.279)
Biberach	0.0642	(0.279)
Bielefeld, Stadt	-0.158	(0.279)
Birkenfeld	-0.779**	(0.279)
Bochum, Stadt	-0.295	(0.279)
Bodenseekreis	0.447	(0.279)
Bonn, Stadt	0.278	(0.279)
Borken	-0.785**	(0.279)
Bottrop, Stadt	-0.179	(0.279)
Brandenburg an der Havel, Stadt	-0.273	(0.279)
Braunschweig, Stadt	-0.223	(0.279)
Breisgau-Hochschwarzwald	0.105	(0.279)

Bremen, Stadt	0.335	(0.279)
Bremerhaven, Stadt	0.0406	(0.279)
Burgenlandkreis	-0.629**	(0.279)
Böblingen	-1.005***	(0.279)
Börde	-0.218	(0.279)
Calw	-0.525*	(0.279)
Celle	-0.407	(0.279)
Cham	-0.602**	(0.279)
Chemnitz, Stadt	-0.149	(0.279)
Cloppenburg	-0.152	(0.279)
Coburg	-0.0539	(0.242)
Cochem-Zell	-0.213	(0.279)
Coesfeld	7.568***	(0.279)
Cottbus, Stadt	-0.351	(0.279)
Cuxhaven	0.166	(0.279)
Dachau	0.858***	(0.279)
Dahme-Spreewald	0.0776	(0.279)
Darmstadt, Wissenschaftsstadt	0.435	(0.279)
Darmstadt-Dieburg	0.12	(0.279)
Deggendorf	0.133	(0.279)
Delmenhorst, Stadt	-0.33	(0.279)
Dessau-Roßlau, Stadt	-0.929***	(0.279)
Diepholz	-0.161	(0.279)
Dillingen ad-Donau	-0.00273	(0.279)
Dingolfing-Landau	0.111	(0.279)
Dithmarschen	-0.0678	(0.279)
Donau-Ries	0.114	(0.279)
Donnersbergkreis	-0.464	(0.279)
Dortmund, Stadt	-0.285	(0.279)
Dresden, Stadt	0.0679	(0.279)
Duisburg, Stadt	-0.552*	(0.279)
Düren	-0.371	(0.279)
Düsseldorf, Stadt	0.577*	(0.279)
Ebersberg	0.850***	(0.279)
Eichsfeld	-0.530*	(0.279)
Eichstätt	0.522*	(0.279)
Eifelkreis Bitburg-Prüm	0.0339	(0.279)
Eisenach, Stadt	-0.542*	(0.279)
Elbe-Elster	-0.556*	(0.279)
Emden, Stadt	-0.576*	(0.279)
Emmendingen	0.286	(0.279)
Emsland	-0.27	(0.279)
Ennepe-Ruhr-Kreis	-0.256	(0.279)
Enzkreis	0.00152	(0.279)
Erding	0.725**	(0.279)
Erfurt, Stadt	-0.0207	(0.279)

Erlangen	0.509*	(0.279)
Erlangen-Höchstadt	0.283	(0.279)
Erzgebirgskreis	-0.855***	(0.279)
Essen, Stadt	-0.217	(0.279)
Esslingen	0.401	(0.279)
Euskirchen	-0.157	(0.279)
Flensburg, Stadt	-0.0216	(0.279)
Forchheim	0.205	(0.279)
Frankenthal (Pfalz), kreisfreie Stadt	-0.0755	(0.279)
Frankfurt (Oder), Stadt	-0.0505	(0.279)
Frankfurt am Main, Stadt	0.844***	(0.279)
Freiburg im Breisgau, Stadtkreis	0.720**	(0.279)
Freising	0.738**	(0.279)
Freudenstadt	-0.384	(0.279)
Freyung-Grafenau	-0.762**	(0.279)
Friesland	0.0988	(0.279)
Fulda	-0.143	(0.279)
Fürstenfeldbruck	0.840***	(0.279)
Fürth	0.267	(0.242)
Garmisch-Partenkirchen	0.669**	(0.279)
Gelsenkirchen, Stadt	-0.677**	(0.279)
Gera, Stadt	-0.881***	(0.279)
Germersheim	0.0232	(0.279)
Gießen	0.0621	(0.279)
Gifhorn	-0.0758	(0.279)
Goslar	0.0714	(0.279)
Gotha	-1.052***	(0.279)
Grafschaft Bentheim	-0.919***	(0.279)
Greiz	-0.535*	(0.279)
Groß-Gerau	-0.362	(0.279)
Göppingen	-0.273	(0.279)
Görlitz	-0.703**	(0.279)
Göttingen	0.157	(0.279)
Günzburg	-0.0156	(0.279)
Gütersloh	-0.0956	(0.279)
Hagen, Stadt der Fern-Universität	-0.462	(0.279)
Halle (Saale), Stadt	-0.306	(0.279)
Hamburg, Freie und Hansestadt	0.778**	(0.279)
Hameln-Pyrmont	-0.667**	(0.279)
Hamm, Stadt	-0.432	(0.279)
Harburg	0.105	(0.279)
Harz	-0.780**	(0.279)
Havelland	-0.317	(0.279)
Haßberge	-0.00477	(0.279)
Heidekreis	-0.601**	(0.279)
Heidelberg, Stadtkreis	0.613**	(0.279)

Heidenheim	-0.179	(0.279)
Heilbronn	0.17	(0.279)
Heilbronn, Stadtkreis	0.269	(0.279)
Heinsberg	-0.396	(0.279)
Helmstedt	-0.425	(0.279)
Herford	-0.433	(0.279)
Herne, Stadt	-0.498*	(0.279)
Hersfeld-Rotenburg	-0.46	(0.279)
Herzogtum Lauenburg	-0.0746	(0.279)
Hildburghausen	-0.734**	(0.279)
Hildesheim	-0.438	(0.279)
Hochsauerlandkreis	-0.623**	(0.279)
Hochtaunuskreis	0.469	(0.279)
Hof	-0.733***	(0.242)
Hohenlohekreis	-0.0391	(0.279)
Holzminden	-0.969***	(0.279)
Höxter	-0.709**	(0.279)
Ilm-Kreis	-0.388	(0.279)
Ingolstadt	0.680**	(0.279)
Jena, Stadt	0.0679	(0.279)
Jerichower Land	-0.929***	(0.279)
Kaiserslautern	-0.500*	(0.279)
Kaiserslautern, kreisfreie Stadt	-0.338	(0.279)
Karlsruhe	0.156	(0.279)
Karlsruhe, Stadtkreis	0.445	(0.279)
Kassel	-0.213	(0.279)
Kassel, documenta-Stadt	0.0077	(0.279)
Kaufbeuren	0.0737	(0.279)
Kelheim	0.209	(0.279)
Kempten (Allgäu)	0.262	(0.279)
Kiel, Landeshauptstadt	0.142	(0.279)
Kitzingen	-0.0416	(0.279)
Kleve	-0.297	(0.279)
Koblenz, kreisfreie Stadt	0.173	(0.279)
Konstanz	0.536*	(0.279)
Krefeld, Stadt	0.461	(0.279)
Kronach	-0.25	(0.279)
Kulmbach	-0.474	(0.279)
Kusel	-0.33	(0.279)
Kyffhäuserkreis	-0.535*	(0.279)
Köln, Stadt	-1.132***	(0.279)
Lahn-Dill-Kreis	-0.186	(0.279)
Landau in der Pfalz, kreisfreie Stadt	0.142	(0.279)
Landkreis Rostock	0.365	(0.279)
Landsberg am Lech	0.534*	(0.279)
Landshut	0.432*	(0.242)

Leer	0.284	(0.279)
Leipzig	-0.454	(0.279)
Leipzig, Stadt	0.00667	(0.279)
Leverkusen, Stadt	0.0442	(0.279)
Lichtenfels	-0.112	(0.279)
Limburg Weilburg	-0.158	(0.279)
Lindau (Bodensee)	0.226	(0.279)
Lippe	-0.414	(0.279)
Ludwigsburg	0.309	(0.279)
Ludwigshafen am Rhein, kreisfreie Stadt	0.187	(0.279)
Ludwigslust-Parchim	-0.685**	(0.279)
Lörrach	0.41	(0.279)
Lübeck, Hansestadt	-0.0788	(0.279)
Lüchow Dannenberg	-0.685**	(0.279)
Lüneburg	0.153	(0.279)
Magdeburg, Landeshauptstadt	-0.476	(0.279)
Main-Kinzig-Kreis	0.0463	(0.279)
Main-Spessart	-0.219	(0.279)
Main-Tauber-Kreis	-0.171	(0.279)
Main-Taunus-Kreis	0.415	(0.279)
Mainz, kreisfreie Stadt	0.409	(0.279)
Mainz-Bingen	0.0864	(0.279)
Mannheim, Stadtkreis	0.332	(0.279)
Mansfeld-Südharz	-0.715**	(0.279)
Marburg-Biedenkopf	0.103	(0.279)
Mayen-Koblenz	-0.475	(0.279)
Mecklenburgische Seenplatte	-0.133	(0.279)
Meißen	-0.182	(0.279)
Memmingen	-0.288	(0.279)
Merzig-Wadern	-0.385	(0.279)
Mettmann	0.248	(0.279)
Miesbach	-0.0613	(0.279)
Miltenberg	-0.00959	(0.279)
Minden-Lübbecke	0.916***	(0.279)
Mittelsachsen	-0.106	(0.279)
Märkisch-Oderland	-0.467	(0.279)
Märkischer Kreis	-0.810**	(0.279)
Mönchengladbach, Stadt	-0.39	(0.279)
Mühldorf a. Inn	0.0785	(0.279)
Mülheim an der Ruhr, Stadt	-0.162	(0.279)
München	0.996***	(0.279)
München, Landeshauptstadt	1.255***	(0.279)
Münster, Stadt	0.39	(0.279)
Neckar-Odenwald-Kreis	-0.333	(0.279)
Neu-Ulm	0.234	(0.279)
Neuburg-Schrobenhausen	0.151	(0.279)

Neumarkt i.d. O-Pf	-0.411	(0.279)
Neumünster, Stadt	-0.547*	(0.279)
Neunkirchen	-0.149	(0.279)
Neustadt a.d. Aisch-Bad Windsheim	-0.351	(0.279)
Neustadt ad-Waldnaab	0.109	(0.279)
Neustadt a.d.W., kreisfreie Stadt	0.187	(0.279)
Neuwied	-0.218	(0.279)
Nienburg (Weser)	-0.522*	(0.279)
Nordfriesland	1.071***	(0.279)
Nordhausen	-1.028***	(0.279)
Nordsachsen	-0.623**	(0.279)
Nordwestmecklenburg	0.0672	(0.279)
Northeim	-0.726**	(0.279)
Nürnberg	0.334	(0.279)
Nürnberger Land	0.173	(0.279)
Oberallgäu	0.273	(0.279)
Oberbergischer Kreis	-0.428	(0.279)
Oberhausen, Stadt	-0.392	(0.279)
Oberhavel	0.00802	(0.279)
Oberspreewald-Lausitz	-0.758**	(0.279)
Odenwaldkreis	-0.322	(0.279)
Oder-Spree	-0.195	(0.279)
Offenbach	0.199	(0.279)
Offenbach am Main, Stadt	0.284	(0.279)
Oldenburg	-0.0786	(0.279)
Oldenburg (Oldenburg), Stadt	0.198	(0.279)
Olpe	-0.247	(0.279)
Ortenaukreis	0.0371	(0.279)
Osnabrück	-0.287	(0.279)
Osnabrück, Stadt	-0.00126	(0.279)
Ostalbkreis	0.00644	(0.279)
Ostallgäu	0.252	(0.279)
Osterholz	-0.231	(0.279)
Ostholstein	0.335	(0.279)
Ostprignitz-Ruppin	-0.278	(0.279)
Paderborn	-0.0944	(0.279)
Passau	-0.0451	(0.242)
Peine	-0.427	(0.279)
Pfaffenhofen a.d. Ilm	0.510*	(0.279)
Pforzheim, Stadtkreis	0.0185	(0.279)
Pinneberg	0.0799	(0.279)
Pirmasens, kreisfreie Stadt	-0.928***	(0.279)
Plön	0.144	(0.279)
Potsdam, Stadt	0.516*	(0.279)
Potsdam-Mittelmark	0.167	(0.279)
Prignitz	-0.811**	(0.279)

Rastatt	0.0466	(0.279)
Ravensburg	0.214	(0.279)
Recklinghausen	-0.406	(0.279)
Regen	-0.656**	(0.279)
Regensburg	0.486*	(0.242)
Region Hannover	-0.209	(0.279)
Regionalverband Saarbrücken	0.0775	(0.279)
Rems-Murr-Kreis	-0.472	(0.279)
Remscheid, Stadt	0.355	(0.279)
Rendsburg-Eckernförde	-0.0812	(0.279)
Reutlingen	0.276	(0.279)
Rhein-Erft-Kreis	-0.0127	(0.279)
Rhein-Hunsrück-Kreis	0.117	(0.279)
Rhein-Kreis Neuss	-0.418	(0.279)
Rhein-Lahn-Kreis	0.0243	(0.279)
Rhein-Neckar-Kreis	0.0438	(0.279)
Rhein-Pfalz-Kreis	-0.339	(0.279)
Rhein-Sieg-Kreis	0.143	(0.279)
Rheingau-Taunus-Kreis	0.0214	(0.279)
Rheinisch-Bergischer Kreis	0.0353	(0.279)
Rhön-Grabfeld	-0.321	(0.279)
Rosenheim	0.593**	(0.242)
Rostock	0.286	(0.279)
Rotenburg (Wümme)	-0.321	(0.279)
Roth	0.0791	(0.279)
Rottal-Inn	-0.0922	(0.279)
Rottweil	-0.216	(0.279)
Saale-Holzland-Kreis	-0.588*	(0.279)
Saale-Orla-Kreis	-0.838***	(0.279)
Saalekreis	-0.830**	(0.279)
Saalfeld-Rudolstadt	-0.679**	(0.279)
Saarlouis	-0.274	(0.279)
Saarpfalz-Kreis	-0.275	(0.279)
Salzgitter, Stadt	-0.391	(0.279)
Salzlandkreis	-0.684**	(0.279)
Schaumburg	-1.052***	(0.279)
Schleswig-Flensburg	-0.535*	(0.279)
Schmalkalden-Meiningen	-0.159	(0.279)
Schwabach	-0.702**	(0.279)
Schwalm-Eder-Kreis	0.132	(0.279)
Schwandorf	-0.15	(0.279)
Schwarzwald-Baar-Kreis	-0.484	(0.279)
Schweinfurt	-0.0627	(0.242)
Schwerin	-0.194	(0.279)
Schwäbisch Hall	-0.131	(0.279)
Segeberg	-0.124	(0.279)

Siegen Wittgenstein	0.0667	(0.279)
Sigmaringen	-0.235	(0.279)
Soest	-0.191	(0.279)
Solingen, Klingenstadt	-0.265	(0.279)
Sonneberg	-0.143	(0.279)
Speyer, kreisfreie Stadt	-0.701**	(0.279)
Spree-Neiße	-0.762**	(0.279)
St Wendel	0.211	(0.279)
Stade	-0.833***	(0.279)
Starnberg	-0.101	(0.279)
Steinburg	-0.0147	(0.279)
Steinfurt	0.990***	(0.279)
Stendal	-0.417	(0.279)
Stormarn	-0.225	(0.279)
Straubing	-0.621**	(0.279)
Straubing-Bogen	0.239	(0.279)
Stuttgart, Stadtkreis	0.175	(0.279)
Städteregion Aachen	-0.618**	(0.279)
Suhl, Stadt	0.686**	(0.279)
Sächsische Schweiz-Osterzgebirge	-0.369	(0.279)
Sömmerda	-0.0588	(0.279)
Südliche Weinstraße	-0.686**	(0.279)
Südwestpfalz	-0.563*	(0.279)
Teltow-Fläming	-0.158	(0.279)
Tirschenreuth	-0.547*	(0.279)
Traunstein	0.261	(0.279)
Trier, kreisfreie Stadt	0.244	(0.279)
Trier-Saarburg	0.0468	(0.279)
Tuttlingen	0.388	(0.279)
Tübingen	-0.0644	(0.279)
Uckermark	-0.553*	(0.279)
Uelzen	-0.492*	(0.279)
Ulm, Stadtkreis	0.479	(0.279)
Unna	-0.376	(0.279)
Unstrut-Hainich-Kreis	-0.890***	(0.279)
Unterallgäu	0.237	(0.279)
Vechta	0.0368	(0.279)
Verden	-0.136	(0.279)
Viersen	-0.177	(0.279)
Vogelsbergkreis	-0.618**	(0.279)
Vogtlandkreis	-1.160***	(0.279)
Vorpommern-Greifswald	0.217	(0.279)
Vorpommern-Rügen	0.377	(0.279)
Vulkaneifel	-0.545*	(0.279)
Waldeck-Frankenberg	-0.569*	(0.279)
Waldshut	-0.0641	(0.279)

Warendorf	-0.266	(0.279)
Wartburgkreis	-0.830**	(0.279)
Weiden i d O-Pf	-0.093	(0.279)
Weilheim Schongau	0.464	(0.279)
Weimar, Stadt	-0.139	(0.279)
Weimarer Land	-0.506*	(0.279)
Weißenburg-Gunzenhausen	-0.00185	(0.279)
Werra-Meißner-Kreis	-0.746**	(0.279)
Wesel	-0.285	(0.279)
Wesermarsch	-0.495*	(0.279)
Westerwaldkreis	-0.341	(0.279)
Wetteraukreis	0.16	(0.279)
Wiesbaden, Landeshauptstadt	0.555*	(0.279)
Wilhelmshaven, Stadt	-0.654**	(0.279)
Wittenberg	-0.985***	(0.279)
Wittmund	0.285	(0.279)
Wolfenbüttel	-0.31	(0.279)
Wolfsburg, Stadt	0.0536	(0.279)
Worms, kreisfreie Stadt	-0.104	(0.279)
Wunsiedel i Fichtelgebirge	-0.940***	(0.279)
Wuppertal, Stadt	-0.441	(0.279)
Würzburg	0.295	(0.242)
Zollernalbkreis	-0.19	(0.279)
Zweibrücken, kreisfreie Stadt	-0.284	(0.279)
Zwickau	-0.967***	(0.279)
Observations	401	
R-squared	0.993	

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A10: Regression on ln house prices – federal state level

VARIABLES	(1) Federal States	(2) InputVar	(3) InputVar & Institution
Hamburg	0.709**	-0.0350	-0.0352
	(0.359)	(0.159)	(0.157)
Lower Saxony	-0.316***	-0.112**	-0.0900*
	(0.104)	(0.0466)	(0.0465)
Bremen	-0.363	-0.249**	-0.239**
	(0.262)	(0.117)	(0.116)
North Rhine-Westphalia	-0.301***	-0.183***	-0.145***
	(0.102)	(0.0487)	(0.0492)
Hesse	-0.0644	-0.133***	-0.0983*
	(0.113)	(0.0511)	(0.0514)
Rhineland-Palatinate	-0.251**	-0.133***	-0.103**
	(0.107)	(0.0483)	(0.0484)
Baden-Württemberg	0.0840	-0.133***	-0.101**
	(0.104)	(0.0493)	(0.0495)
Bavaria	0.0762	-0.0218	-0.0105
	(0.0966)	(0.0453)	(0.0451)
Saarland	-0.358**	-0.0393	-0.00336
	(0.168)	(0.0762)	(0.0760)
Berlin	0.573	0.0851	0.0415
	(0.359)	(0.160)	(0.159)
Brandenburg	-0.308**	0.248***	0.174**
	(0.122)	(0.0615)	(0.0681)
Mecklenburg-Western Pomerania	-0.0419	0.265***	0.206**
	(0.152)	(0.0732)	(0.0815)
Saxony	-0.679***	0.000154	-0.0284
	(0.132)	(0.0683)	(0.0708)
Saxony-Anhalt	-0.870***	-0.127*	-0.152**
	(0.129)	(0.0677)	(0.0703)
Thuringia	-0.700***	-0.0335	-0.0596
	(0.115)	(0.0579)	(0.0627)
Gva		-0.00293***	-0.00283***
		(0.000938)	(0.000928)
ln(Inc)		0.646***	0.598***
		(0.130)	(0.131)
Tourist		0.00190***	0.00174***
		(0.000204)	(0.000209)
ln(age)		-2.475***	-1.971***
		(0.352)	(0.384)
ln(Pdensity)		0.0319	0.0248
		(0.0228)	(0.0227)
Mig		0.000336	0.000367
		(0.000377)	(0.000382)
ln(Bb)		0.0322	0.0304
		(0.0603)	(0.0596)
Train		-0.00222***	-0.00232***
		(0.000699)	(0.000697)
Rent		0.0914***	0.0933***
		(0.0103)	(0.0102)
ln(SfhD)		0.491***	0.397***

		(0.133)	(0.135)
ln(sqmliv)		0.0736	0.0703
		(0.0789)	(0.0780)
ln(Blp)		0.156***	0.149***
		(0.0253)	(0.0254)
Building permissions			0.0186***
			(0.00558)
Economic aid			9.02e-05
			(9.87e-05)
Infrastructure aid			5.50e-05
			(0.000212)
Constant	7.637***	6.431***	5.415***
	(0.0898)	(1.555)	(1.564)
Observations	401	401	401
R-squared	0.393	0.889	0.893

Note: Regression (1) solely focuses on the association of federal states, while (2) includes input variables and (3) includes input variables and institutional factors. The dependent variable, house price is transformed into the ln log. Similarly, the independent variables annual income (Inc), average age (Age), population density (Pdensity), broadband access (Bb), single family and duplex homes (Sfhd), average living area (sqmliv) and building land price (Blp) is expressed as ln values, and present elasticities for the ln house price. Infrastructure and economic aids are granted federal subsidy programs for regional developments in the long term (aggregated for 2008-2017). The federal state of Schleswig-Holstein is the reference category for location. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A11: Regression models with robust standard error

VARIABLES	(1) All Counties	(2) U-R	(3) U-R & U-E	(4) U-R, U-E & N-S
Gva	-0.00368***	-0.00341***	-0.00298***	-0.00290***
	(0.000899)	(0.000850)	(0.000823)	(0.000834)
ln(Inc)	0.267**	0.307**	0.527***	0.541***
	(0.123)	(0.120)	(0.138)	(0.139)
Tourist	0.00175***	0.00176***	0.00201***	0.00201***
	(0.000314)	(0.000311)	(0.000299)	(0.000295)
ln(age)	-1.445***	-1.418***	-2.228***	-2.318***
	(0.299)	(0.299)	(0.357)	(0.353)
ln(Pdensity)	-0.0553**	-0.0744***	-0.0507**	-0.0253
	(0.0218)	(0.0219)	(0.0210)	(0.0217)
Mig	0.00102***	0.000618	0.000352	0.000290
	(0.000379)	(0.000389)	(0.000382)	(0.000378)
ln(Bb)	0.0993	0.0813	0.154*	0.145*
	(0.0770)	(0.0771)	(0.0792)	(0.0804)
Train	-0.00259***	-0.00237***	-0.00210**	-0.00204**
	(0.000853)	(0.000834)	(0.000821)	(0.000813)
Rent	0.122***	0.124***	0.109***	0.102***
	(0.0116)	(0.0118)	(0.0122)	(0.0123)
ln(SfhD)	0.307**	0.407***	0.589***	0.530***
	(0.131)	(0.140)	(0.135)	(0.135)
ln(sqmliv)	0.104	0.0893	0.0919	0.0652
	(0.0900)	(0.0899)	(0.0877)	(0.0870)
ln(Blp)	0.182***	0.184***	0.188***	0.184***
	(0.0296)	(0.0292)	(0.0280)	(0.0275)
Urban		0.115***	0.114***	0.0679*
		(0.0376)	(0.0376)	(0.0390)
East			0.160***	0.184***
			(0.0375)	(0.0376)
South				0.0728***
				(0.0206)
North				0.0670***
				(0.0251)
Constant	6.866***	6.085***	5.718***	6.187***
	(1.647)	(1.632)	(1.624)	(1.649)
Observations	401	401	401	401
R-squared	0.853	0.855	0.862	0.865

Note: The dependent variable, house price is transformed into the ln log. Similarly, the independent variables annual income (Inc), average age (Age), population density (Pdensity), broadband access (Bb), single family and duplex homes (Sfhd), average living area (Sqmliv) and building land price (Blp) is expressed as ln values, and present elasticities for the ln house price. Urban, East, South and North are spatial variables for the location of the county. A constant term has been included in all regression. The results show no improvement compared to a model without robust standard errors. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A12: Matrix of correlation for all considered Variables

	Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(1)	*ln (house price)	1.000																			
(2)	ln (GDP)	0.471	1.000																		
(3)	*Gross value added	0.450	0.715	1.000																	
(4)	*ln (Income)	0.565	0.265	0.433	1.000																
(5)	*Tourist beds	0.129	-0.128	-0.176	0.001	1.000															
(6)	Purchasing Power	0.071	0.043	0.027	-0.032	-0.099	1.000														
(7)	*ln (Age)	-0.669	-0.574	-0.460	-0.320	0.200	-0.081	1.000													
(8)	*ln (Popdensity)	0.370	0.573	0.324	-0.012	-0.355	0.106	-0.530	1.000												
(9)	*Migration	0.181	0.805	0.267	-0.111	-0.031	0.066	-0.306	0.553	1.000											
(10)	High edu residents	0.565	0.553	0.376	0.168	-0.166	0.092	-0.503	0.656	0.456	1.000										
(11)	Recreation areas	-0.036	-0.050	-0.088	-0.052	0.030	-0.049	0.018	-0.070	0.022	-0.047	1.000									
(12)	Highway distance	-0.228	-0.336	-0.231	-0.034	0.349	-0.113	0.307	-0.614	-0.279	-0.362	0.045	1.000								
(13)	Airport distance	-0.273	-0.188	-0.251	-0.106	0.357	-0.107	0.315	-0.548	-0.108	-0.411	0.099	0.515	1.000							
(14)	*Train distance	-0.422	-0.390	-0.275	-0.003	0.151	-0.116	0.446	-0.615	-0.336	-0.524	-0.004	0.408	0.436	1.000						
(15)	*Broadband access	0.425	0.531	0.327	0.155	-0.266	0.019	-0.515	0.774	0.440	0.486	-0.016	-0.52	-0.413	-0.498	1.000					
(16)	*Rent	0.860	0.491	0.559	0.565	-0.068	0.081	-0.652	0.447	0.176	0.681	-0.031	-0.29	-0.381	-0.430	0.428	1.000				
(17)	*Single dwellings	-0.205	-0.474	-0.128	0.232	0.161	-0.077	0.281	-0.816	-0.617	-0.652	0.019	0.416	0.358	0.513	-0.538	-0.301	1.000			
(18)	*ln (Floor ratio)	0.020	0.019	-0.003	-0.070	0.077	0.069	0.022	-0.025	0.043	-0.009	0.220	0.009	0.078	-0.017	-0.066	-0.001	-0.032	1.000		
(19)	Land tax	-0.259	-0.184	-0.127	-0.340	-0.083	-0.034	0.168	0.129	0.014	-0.008	-0.008	-0.07	-0.259	-0.189	0.064	-0.200	-0.107	-0.029	1.000	
(20)	*ln(Land Price)	0.805	0.556	0.530	0.487	-0.158	0.102	-0.730	0.690	0.273	0.668	-0.058	-0.42	-0.490	-0.521	0.655	0.831	-0.461	-0.029	-0.163	1.000

The correlation matrix shows the regarded explanatory variables and those have not been included, partly reasoned because high correlation values. right hand side variables that are used in the model are marked with a star, “ * ”.

Transformation of independent variables

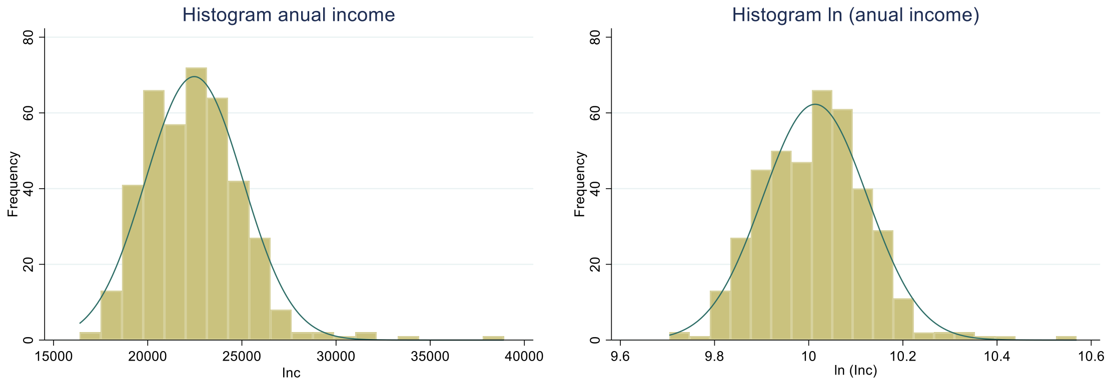


Figure A5: Transformation of income variable

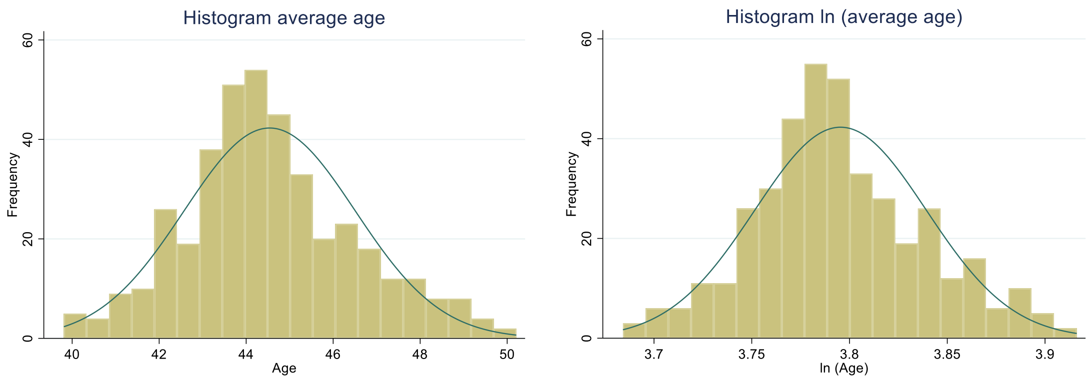


Figure A6 Transformation of age variable

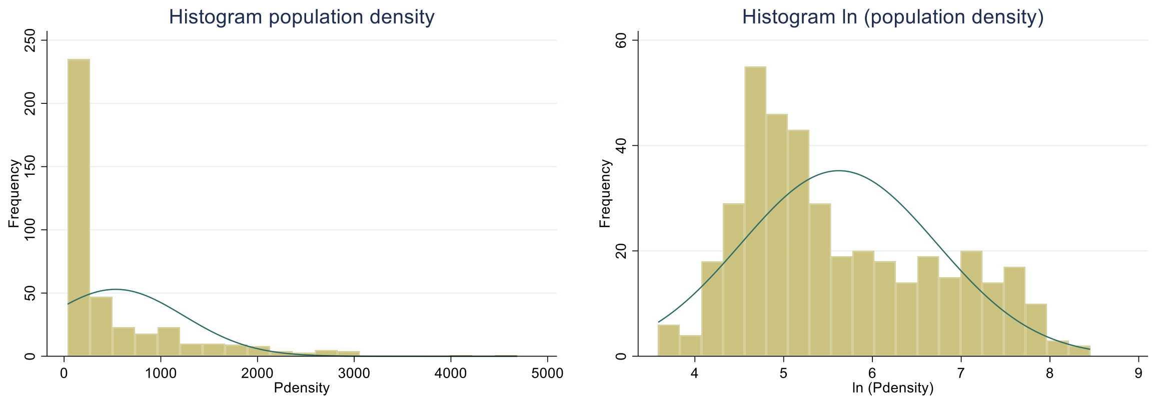


Figure A7: Transformation of population density variable

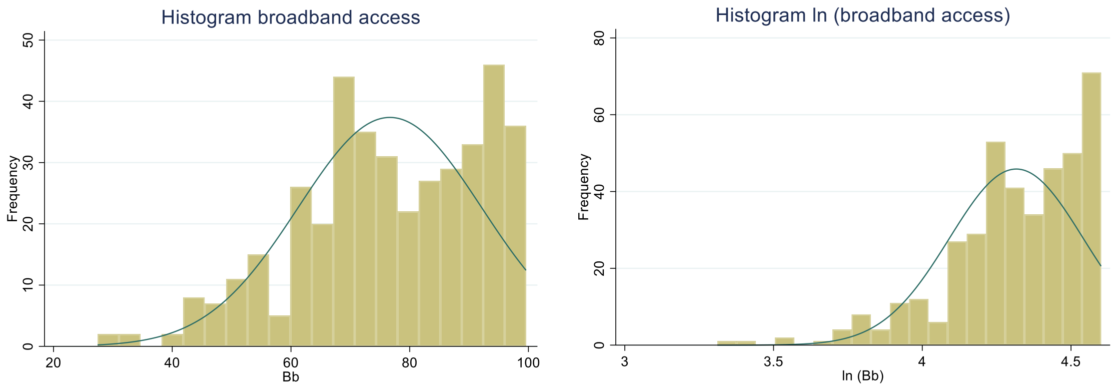


Figure A8: Transformation of broadband access variable

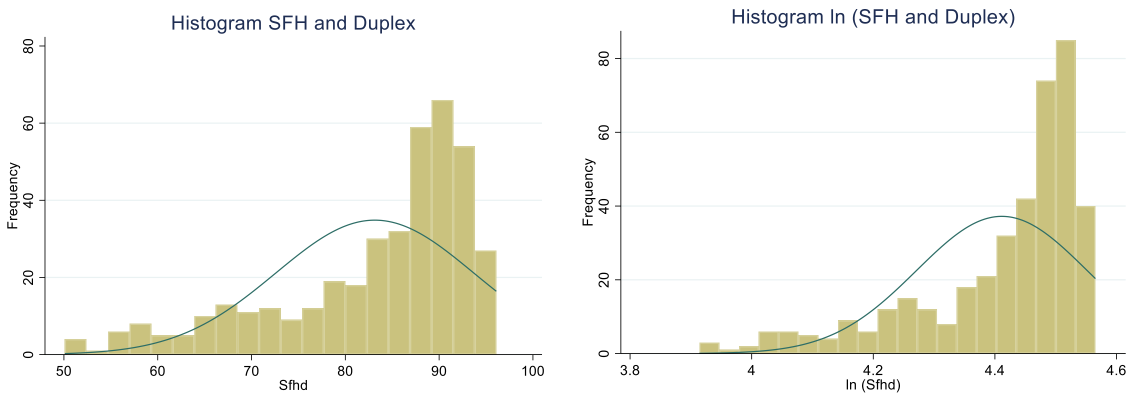


Figure A9: Transformation of single-family-and-duplex-ratio variable

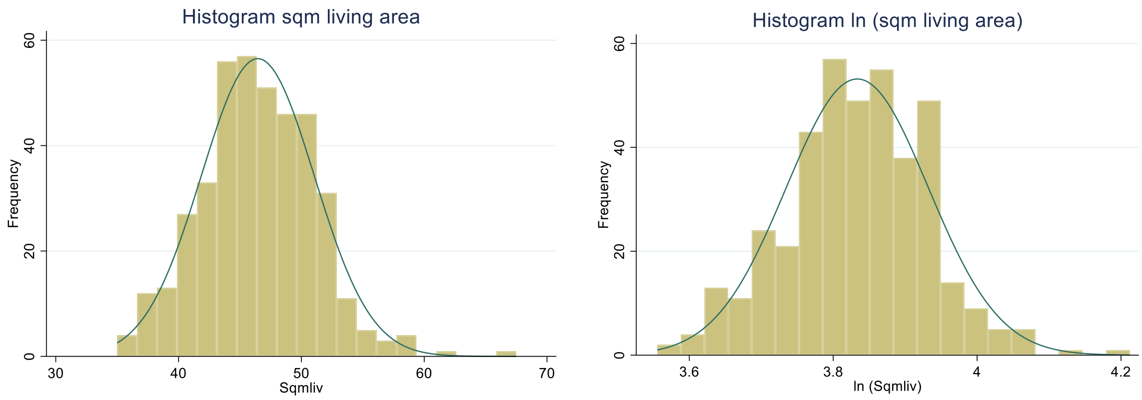


Figure A10: Transformation of square meter per person variable

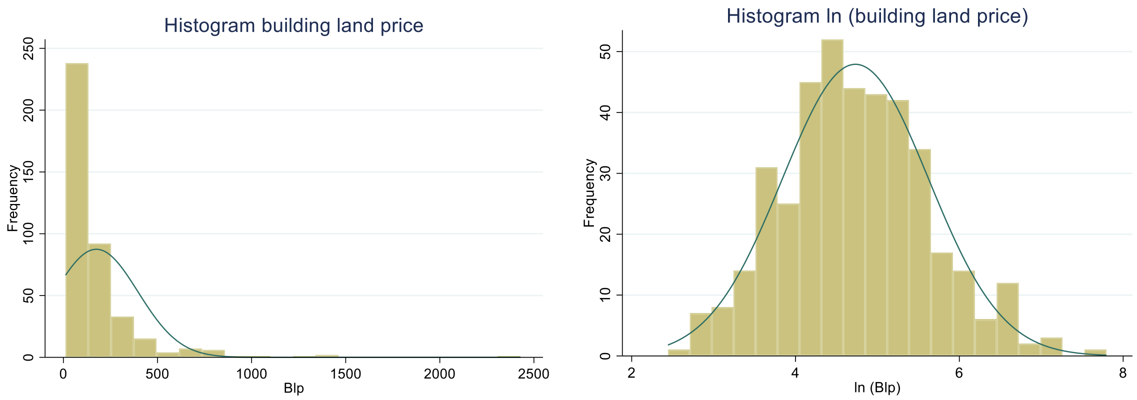


Figure A11: Transformation of building land price variable

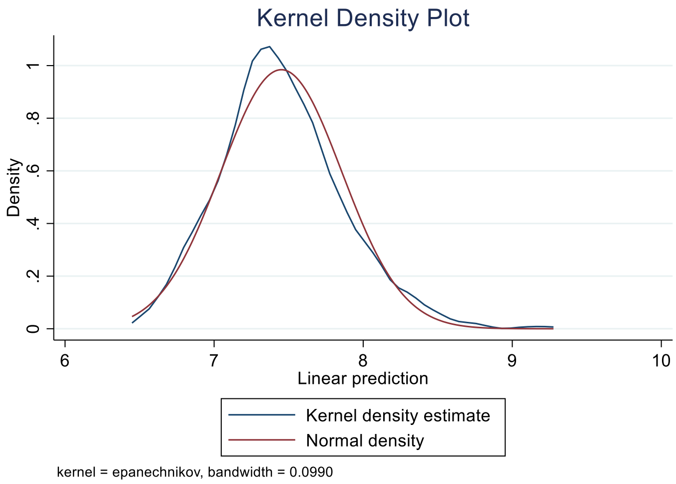


Figure A12: Kernel Density Plot

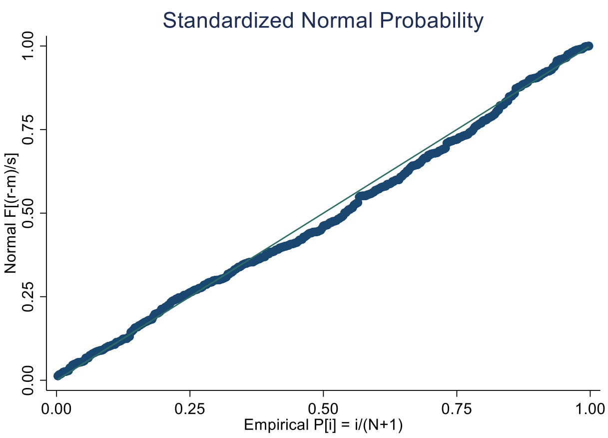


Figure A13: Standardized Normal Probability Plot

Appendix D (Assumption testing)

The following appendix provides tests on the model expressed with commonly used key figures and showcasing in tables and plots, referring to Brooks & Tsolacos (2010). The conceptual order of tests is derived from Wiersma (2020) and Brooks & Tsolacos (2010) and includes several concerns:

- Correlation of variables
- Multicollinearity
- Homoscedasticity
- Parameter stability
- Relation between variables and error term
- Chow test
- Normality of Residuals

Correlation of variables

For a statistically significant regression model, the correlation of variables can lead to unfavourable outcomes (Hair et al., 2010). Therefore, Stata provides the *correlate* function to test the correlation between each pair of variables. The outcomes for the regression model are presented in table 11**,** while the full correlation table can be found in appendix C, table A13.

Table A13: Matrix of correlations of regressors

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Gva	1.000							
(2) ln (Inc)	0.433	1.000						
(3) Tourist	-0.176	0.001	1.000					
(4) ln (Age)	-0.460	-0.320	0.200	1.000				
(5) ln (Pdensity)	0.324	-0.012	-0.355	-0.530	1.000			
(6) Mig	0.267	-0.111	-0.031	-0.306	0.553	1.000		
(7) ln (Bb)	0.316	0.182	-0.246	-0.516	0.726	0.395	1.000	
(8) Train	-0.275	-0.003	0.151	0.446	-0.615	-0.336	-0.467	1.000
(9) Rent	0.559	0.565	-0.068	-0.652	0.447	0.176	0.426	-0.430
(10) ln (Sfhd)	-0.127	0.237	0.156	0.293	-0.809	-0.601	-0.476	0.510
(11) ln (Sqmliv)	-0.003	-0.070	0.077	0.022	-0.025	0.043	-0.081	-0.017
(12) ln (Blp)	0.530	0.487	-0.158	-0.730	0.690	0.273	0.640	-0.521

Variables	(9)	(10)	(11)	(12)
(9) Rent	1.000			
(10) ln (Sfhd)	-0.301	1.000		
(11) ln (Sqmliv)	-0.001	-0.034	1.000	
(12) ln (Blp)	0.831	-0.458	-0.029	1.000

Based on the literature research, several variables were initially considered. However, due to high correlation values or redundancies excluded in the final regression model. For instance, the data set contained multiple variables for migration. The decision for choosing only the Mig (=total migration balance) over $family\ migration$ balance, $mid-age\ migration$ or $educational\ migration$ proved to deliver a more significant outcome in the model and also the problem of intersecting migration flows was avoided. Although family migration and educational migration are expected to correlate with the house prices, the inclusion in the given model does not show sound evidence. The rule of thumb for assessing the correlation values is given by Zady (2000). Accordingly, a value between 0.00 and 0.29 gives little to no evidence of correlation, while 0.30 to 0.49 indicates some correlation, 0.50 to 0.69 conforms medium correlation, 0.70 to 0.89 gives high and 0.90 to 1.00 indicates very high correlation. Values exceeding the given threshold are conducted further for variance inflation factors.

Multicollinearity

To quantify the multicollinearity of the regression models used, the variance inflation factors (VIF) are determined as the quotient of the variance of the regressors. Table A14 highlights the used independent variables for determining the LN of the regional house prices. Previous regressions lead to the transformation of variables towards its natural logarithm, which can lead to lower VIF.

Table A14: Variance inflation factors

Variable	(1)	(2) UR	(3) UR, EW	(3) UR, EW, NS
ln (Population density)	8.44	9.23	9.810	10.95
ln (Building land price)	7.83	7.83	7.840	7.91
ln (Single Family and Duplex homes)	4.63	4.98	5.520	5.69
Urban		4.67	4.670	5.28
East			3.31	3.54
North				1.92
South				2.20
Rent	4.42	4.43	5.07	5.30
ln (Broadband access)	2.51	2.53	2.76	2.80
ln (average Age)	2.51	2.51	3.57	3.60
ln (annual income)	2.46	2.49	2.99	3.15
Migration balance	1.98	2.23	2.28	2.30
Train distance	1.76	1.78	1.80	1.85
Gross value added	1.75	1.76	1.78	1.79
Tourist beds	1.37	1.36	1.50	1.54
Square meter living area	1.03	1.02	1.03	1.04
Mean VIF	3.39	3.61	3.58	3.67

This reasoned the decision of transforming Age and Income. Some other variables, for instance GDP proved by the literature review and the correlation table to be useful for explaining the dependent variable, however, delivered too high VIF values. Decisive for the exclusion of independent variables is the threshold of $VIF < 10$ for further consideration (Hair et al., 2010; Marquaridt, 1970). The threshold of Mean $VIF < 4$ further proves that no redundant information about the response was included and thus multicollinearity of the model can be precluded (Hair et al., 2010). Merely the ln(Population density) variable exceeds the threshold.

Homoscedasticity

Homoscedasticity is a required assumption for the model that describes a constant variance of the error terms, meaning that the random disturbance along the observations of independent and dependent variables are the same. Hence, the model output compares the residual values with the fitted values to check whether the standard errors are close to zero. Figure 7 and 8 graphically highlights the findings.

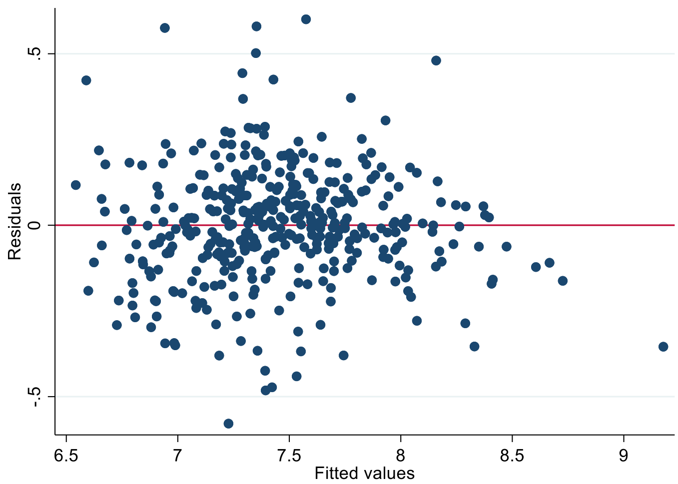


Figure 7: Residual-versus-Fitted Values regression (2)

The main share of the residual errors is situated close to zero, especially the observation that is close to the mean of the fitted values of 7.45. However, a small convergence towards the right tail is visible. As the OLS regression used aims to minimize these standard errors and gives equal weight to the residuals outliers that associate with heteroscedasticity results more than a model with weighted least squares would do and therefore a different visualization is deliberated.

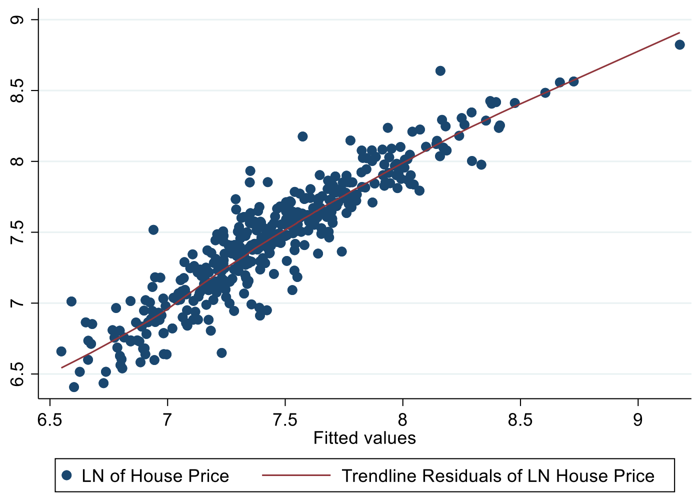


Figure 8: Histogram of Correlation between ln (P) and residuals of ln (P) for regression (2)

Figure 8 compares the predicted residuals with the actual regional ln house prices (per square meter) in a scatter plot. The data points are closely tied together, but some outliers can be observed, that is why two tests for heteroskedasticity are conducted.

Table A15: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance
Variables: fitted values of ln house price chi2(1) = 3.06 Prob > chi2 = 0.0800

The Breusch-Pagan test controls for any linear form of heteroskedasticity in a regression. The null hypothesis tests whether the error variance is a multiplicative function of one or more variables (UND, 2020)

Table A16: Cameron & Trivedi's decomposition of IM-test

White's test for Ho: homoskedasticity

	chi2	df	p
Heteroskedasticity	201.57	103	0.000
Skewness	15.35	13	0.286
Kurtosis	9.14	1	0.003
Total	226.05	117	0.000

The White test version of the Cameron & Trivedi's decomposition test relaxes the assumption if normally distributed errors.

The Breusch-Pagan (Table A15) test concludes a p-value of 0.08 which lies above the threshold of $p<0.05$ (xlstat, n.y.; UND, 2020; Fischer, 1950), allowing to assume that the regression is considered homoscedastic. Admitting that the distance between the threshold and p-value is not very large, and noise is still present, we can nevertheless not reject the null hypothesis of constant variance. Additionally, table A16 approves this finding with the Cameron

& Trivedi's White test and the no rejection of the null hypothesis of homoskedasticity in the data. Nevertheless, the regressions are controlled for the output with robust standard errors, assuming that the data is after all heteroscedastic. Table A11 of appendix C shows the results, while comparing with the results of table 9 no significant differences can be observed.

Chow test

This part reveals context to the Chow test in chapter 4.4.

One assumption for sound academic research is to address the quality and distribution of the dataset. A parameter stability test as introduced by Chow (1960) checks whether the data is constant across the observations. The test has been used frequently in existing literature, e.g. by Watkins (2001) who tests 6 housing submarkets in a pairwise process. For this work, I recall the null hypothesis.

$$H_0 = There\ is\ no\ signifanct\ improvement\ in\ fit\ from\ separating\ urban\ and\ rural\ regressions.$$

Hence, by the official German categorization, urban-rural, east-west and north-south areas in the dataset are prepared to control continuity. The Chow test is an F-Test with the following F-statistic:

$$F = \frac{(RSS_p - RSS_1 - RSS_2)/k}{(RSS_1 + RSS_2)/(n-2k)} \qquad (6)$$

Therefore, equation (6) provides the formula with F being the F statistic value of the test and RSS meaning the residual sum of squares of each model, where the subscripts p , 1 and 2 indicate pooled, subgroup 1 and subgroup 2, respectively. k is the number of regressors and n is the number of observations, which is similar to the number of counties in this work. If the distinction into the two subgroups delivers a high F value, we can assume a larger improvement in fit, than without testing for urban-rural disparities. Thus, if the value F equals to zero, then we do not have an improvement in fit. A F value is close to zero, would imply under the H_0 hypotheses that the F-statistics follows an F-Distribution with k and $n-2k$ degrees of freedom and we cannot reject the null hypothesis. So, in order to check the F-statistics, we have to control whether the F value is on the F-distribution. Therefore, the critical value is calculated and compared to the F-value. H_0 is rejected when the F value is greater than the critical value.

Normality of Residuals

One assumption of the residuals is that the residuals need to follow a normal distribution, which is controlled by a Kernel Density Plot and a standardized normal probability plot. Figure A12 and A13 in appendix C show the plotted error terms of the residuals, indicating the differences between the observed dependent and predicted values. Both figures merely show small deviations from the normal target values, leading to the assumption that the residuals follow a normal distribution.

Appendix E (Stata Syntax)

*1. Generating new variables

*Dependent Variable

label variable housepriceexistingmedian "houseprice"

gen houseprice = housepriceexistingmedian

gen ln_houseprice = ln(houseprice)

*Independent Variables

gen ln_populationdensity = ln(populationdensity)

gen ln_buildinglandprice = ln(buildinglandprice)

gen ln_averageage = ln(averageage)

gen ln_sfhandduplex = ln(sfhandduplex)

gen ln_sqmlivingarea2 = ln(sqmlivingarea2)

gen anualincome =averageavailablehousholdincome*12

gen ln_anualincome = ln(anualincome)

gen ln_sharenewbuildhome = ln(sharenewbuildhome)

gen ln_broadbandaccess = ln(broadbandaccess)

gen ln_buildingpermissionappartment = ln(buildingpermissionappartment)

gen ln_betteninfvbetrieben = ln(betteninfvbetrieben)

gen buildpermissiondummy = (buildingpermissionappartment<3.2)

gen purchasingpowerdummy = (purchasingpower<100)

gen ln_purchasingpowerdummy = ln(purchasingpower>4.6)

*Dummy Variables

encode nameruralandurbancounty, gen (namecounty)

gen Urban = (typofregion=="Urban area")

gen Rural = (typofregion=="Rural county")

gen West = (federalstatenumber<=10)

gen East = (federalstatenumber>=11)

gen South = (federalstatenumber>=7 &federalstatenumber<=10)

gen North = (federalstatenumber<5|federalstatenumber==12|federalstatenumber==13|federalstatenumber==15)

*Variables for Ratios

gen rentanual = (rentpriceaverage*12)

gen rtp = ((rentpriceaverage*12)/houseprice*100)

gen ln_rtp = ln(rtp)

gen pti= (houseprice/(averageavailablehousholdincome*12)*100)

```
gen ln_pti = ln(pti)
gen pptp = (purchasingpower/houseprice*100)
gen ln_pptp = ln(pptp)
```

***2. Descriptive Analysis**

```
global inputvar gva ln_anualincome betteninfvbetrieben ln_averageage ln_populationdensity
    migrationbalance ln_broadbandaccess accessibilityiceciceinminutes rentpriceaverage
    ln_sfhandduplex ln_sqmlivingarea2 ln_buildinglandprice

sum houseprice, d
hist houseprice, graphregion(color(white)) bgcolor(white) normal
hist ln_houseprice, graphregion(color(white)) bgcolor(white) normal

asdoc sum $inputvar buildingpermissionappartment grwinfrastrukturlangfristig baufoerderunglf

scatter ln_houseprice ln_populationdensity, graphregion(color(white)) bgcolor(white)
scatter ln_houseprice ln_population, graphregion(color(white)) bgcolor(white)

graph twoway (scatter ln_houseprice ln_populationdensity) (lowess ln_houseprice
    ln_populationdensity), graphregion(color(white)) bgcolor(white)
twoway (lfit ln_houseprice ln_populationdensity) (scatter ln_houseprice ln_populationdensity)
twoway (scatter houseprice sfhandduplex) (lfit houseprice sfhandduplex), graphregion(color(white))
    bgcolor(white)

toway (scatter houseprice sfhandduplex) (lfit houseprice sfhandduplex) if Urban==0,
    graphregion(color(white)) bgcolor(white)
twoway (scatter houseprice sfhandduplex) (lfit houseprice sfhandduplex) if Urban==1,
    graphregion(color(white)) bgcolor(white)

scatter Urban houseprice
scatter Rural houseprice
scatter East houseprice
scatter West houseprice
scatter North houseprice
scatter South houseprice

asdoc sum houseprice if Urban==1
asdoc sum houseprice if Rural==1
asdoc sum houseprice if West==1
asdoc sum houseprice if East==1
```

asdoc sum houseprice if South==1
asdoc sum houseprice if North==1

Sum of variable means for regional disparities

egen gva_u = mean(gva) if Urban==1
egen gva_r = mean(gva) if Rural==1
egen gva_e = mean(gva) if East==1
egen gva_w = mean(gva) if West==1
egen gva_n = mean(gva) if North==1
egen gva_s = mean(gva) if South==1

egen ln_anualincome_u = mean(ln_anualincome) if Urban==1
egen ln_anualincome_r = mean(ln_anualincome) if Rural==1
egen ln_anualincome_e = mean(ln_anualincome) if East==1
egen ln_anualincome_w = mean(ln_anualincome) if West==1
egen ln_anualincome_n = mean(ln_anualincome) if North==1
egen ln_anualincome_s = mean(ln_anualincome) if South==1

egen betteninfvbetrieben_u = mean(betteninfvbetrieben) if Urban==1
egen betteninfvbetrieben_r = mean(betteninfvbetrieben) if Rural==1
egen betteninfvbetrieben_e = mean(betteninfvbetrieben) if East==1
egen betteninfvbetrieben_w = mean(betteninfvbetrieben) if West==1
egen betteninfvbetrieben_n = mean(betteninfvbetrieben) if North==1
egen betteninfvbetrieben_s = mean(betteninfvbetrieben) if South==1

egen ln_averageage_u = mean(ln_averageage) if Urban==1
egen ln_averageage_r = mean(ln_averageage) if Rural==1
egen ln_averageage_e = mean(ln_averageage) if East==1
egen ln_averageage_w = mean(ln_averageage) if West==1
egen ln_averageage_n = mean(ln_averageage) if North==1
egen ln_averageage_s = mean(ln_averageage) if South==1

egen ln_populationdensity_u = mean(ln_populationdensity) if Urban==1
egen ln_populationdensity_r = mean(ln_populationdensity) if Rural==1
egen ln_populationdensity_e = mean(ln_populationdensity) if East==1
egen ln_populationdensity_w = mean(ln_populationdensity) if West==1
egen ln_populationdensity_n = mean(ln_populationdensity) if North==1
egen ln_populationdensity_s = mean(ln_populationdensity) if South==1

egen migrationbalance_u = mean(migrationbalance) if Urban==1

```
egen migrationbalance_r = mean(migrationbalance) if Rural==1
egen migrationbalance_e = mean(migrationbalance) if East==1
egen migrationbalance_w = mean(migrationbalance) if West==1
egen migrationbalance_n = mean(migrationbalance) if North==1
egen migrationbalance_s = mean(migrationbalance) if South==1

egen ln_broadbandaccess_u = mean(ln_broadbandaccess) if Urban==1
egen ln_broadbandaccess_r = mean(ln_broadbandaccess) if Rural==1
egen ln_broadbandaccess_e = mean(ln_broadbandaccess) if East==1
egen ln_broadbandaccess_w = mean(ln_broadbandaccess) if West==1
egen ln_broadbandaccess_n = mean(ln_broadbandaccess) if North==1
egen ln_broadbandaccess_s = mean(ln_broadbandaccess) if South==1

egen accessibilityiceciceinminutes_u = mean(accessibilityiceciceinminutes) if Urban==1
egen accessibilityiceciceinminutes_r = mean(accessibilityiceciceinminutes) if Rural==1
egen accessibilityiceciceinminutes_e = mean(accessibilityiceciceinminutes) if East==1
egen accessibilityiceciceinminutes_w = mean(accessibilityiceciceinminutes) if West==1
egen accessibilityiceciceinminutes_n = mean(accessibilityiceciceinminutes) if North==1
egen accessibilityiceciceinminutes_s = mean(accessibilityiceciceinminutes) if South==1

egen rentpriceaverage_u = mean(rentpriceaverage) if Urban==1
egen rentpriceaverage_r = mean(rentpriceaverage) if Rural==1
egen rentpriceaverage_e = mean(rentpriceaverage) if East==1
egen rentpriceaverage_w = mean(rentpriceaverage) if West==1
egen rentpriceaverage_n = mean(rentpriceaverage) if North==1
egen rentpriceaverage_s = mean(rentpriceaverage) if South==1

egen ln_sfhandduplex_u = mean(ln_sfhandduplex) if Urban==1
egen ln_sfhandduplex_r = mean(ln_sfhandduplex) if Rural==1
egen ln_sfhandduplex_e = mean(ln_sfhandduplex) if East==1
egen ln_sfhandduplex_w = mean(ln_sfhandduplex) if West==1
egen ln_sfhandduplex_n = mean(ln_sfhandduplex) if North==1
egen ln_sfhandduplex_s = mean(ln_sfhandduplex) if South==1

egen ln_sqmlivingarea2_u = mean(ln_sqmlivingarea2) if Urban==1
egen ln_sqmlivingarea2_r = mean(ln_sqmlivingarea2) if Rural==1
egen ln_sqmlivingarea2_e = mean(ln_sqmlivingarea2) if East==1
egen ln_sqmlivingarea2_w = mean(ln_sqmlivingarea2) if West==1
egen ln_sqmlivingarea2_n = mean(ln_sqmlivingarea2) if North==1
egen ln_sqmlivingarea2_s = mean(ln_sqmlivingarea2) if South==1
```

```
egen ln_buildinglandprice_u = mean(ln_buildinglandprice) if Urban==1
egen ln_buildinglandprice_r = mean(ln_buildinglandprice) if Rural==1
egen ln_buildinglandprice_e = mean(ln_buildinglandprice) if East==1
egen ln_buildinglandprice_w = mean(ln_buildinglandprice) if West==1
egen ln_buildinglandprice_n = mean(ln_buildinglandprice) if North==1
egen ln_buildinglandprice_s = mean(ln_buildinglandprice) if South==1

asdoc sum gva_u gva_r gva_e gva_w gva_n gva_s ln_anualincome_u ln_anualincome_r
    ln_anualincome_e ln_anualincome_w ln_anualincome_n ln_anualincome_s
    betteninfvbetrieben_u betteninfvbetrieben_r betteninfvbetrieben_e betteninfvbetrieben_w
    betteninfvbetrieben_n betteninfvbetrieben_s ln_averageage_u ln_averageage_r
    ln_averageage_e ln_averageage_w ln_averageage_n ln_averageage_s ln_populationdensity_u
    ln_populationdensity_r ln_populationdensity_e ln_populationdensity_w ln_populationdensity_n
    ln_populationdensity_s migrationbalance_u migrationbalance_r migrationbalance_e
    migrationbalance_w migrationbalance_n migrationbalance_s ln_broadbandaccess_u
    ln_broadbandaccess_r ln_broadbandaccess_e ln_broadbandaccess_w ln_broadbandaccess_n
    ln_broadbandaccess_s accessibilityiceciceinminutes_u accessibilityiceciceinminutes_r
    accessibilityiceciceinminutes_e accessibilityiceciceinminutes_w accessibilityiceciceinminutes_n
    accessibilityiceciceinminutes_s rentpriceaverage_u rentpriceaverage_r rentpriceaverage_e
    rentpriceaverage_w rentpriceaverage_n rentpriceaverage_s ln_sfhandduplex_u
    ln_sfhandduplex_r ln_sfhandduplex_e ln_sfhandduplex_w ln_sfhandduplex_n ln_sfhandduplex_s
    ln_sqmlivingarea2_u ln_sqmlivingarea2_r ln_sqmlivingarea2_e ln_sqmlivingarea2_w
    ln_sqmlivingarea2_n ln_sqmlivingarea2_s ln_buildinglandprice_u ln_buildinglandprice_r
    ln_buildinglandprice_e ln_buildinglandprice_w ln_buildinglandprice_n ln_buildinglandprice_s
```

<u>*3. Real Estate Ratios*</u>

```
label variable anualincome "Anual Income"
label variable rtp "Rent to Price"
label variable itp "Income to Price"
label variable pptp "Purchase Power to Price"
label variable rentanual "Anual Rent"
label variable purchasingpower "Purchasing Power"

hist rtp, graphregion(color(white)) bgcolor(white)
hist itp, graphregion(color(white)) bgcolor(white)
hist pptp, graphregion(color(white)) bgcolor(white)

graph twoway (lfit rentanual houseprice) (scatter rentanual houseprice), graphregion(color(white))
    bgcolor(white)
```

```
graph twoway (lfit anualincome houseprice) (scatter anualincome houseprice),
    graphregion(color(white)) bgcolor(white)
graph twoway (lfit purchasingpower houseprice) (scatter purchasingpower houseprice),
    graphregion(color(white)) bgcolor(white)

asdoc sum rtp itp pptp, append
```

Ratios on spatial level

```
egen rtp_mean_u = mean(rtp) if Urban==1
egen rtp_mean_r = mean(rtp) if Rural==1
egen rtp_mean_e = mean(rtp) if East==1
egen rtp_mean_w = mean(rtp) if West==1
egen rtp_mean_n = mean(rtp) if North==1
egen rtp_mean_s = mean(rtp) if South==1

egen itp_mean_u = mean(itp) if Urban==1
egen itp_mean_r = mean(itp) if Rural==1
egen itp_mean_e = mean(itp) if East==1
egen itp_mean_w = mean(itp) if West==1
egen itp_mean_n = mean(itp) if North==1
egen itp_mean_s = mean(itp) if South==1

egen pptp_mean_u = mean(pptp) if Urban==1
egen pptp_mean_r = mean(pptp) if Rural==1
egen pptp_mean_e = mean(pptp) if East==1
egen pptp_mean_w = mean(pptp) if West==1
egen pptp_mean_n = mean(pptp) if North==1
egen pptp_mean_s = mean(pptp) if South==1

asdoc sum rtp_mean_u rtp_mean_r rtp_mean_e rtp_mean_w rtp_mean_n rtp_mean_s itp_mean_u
    itp_mean_r itp_mean_e itp_mean_w itp_mean_n itp_mean_s pptp_mean_u pptp_mean_r
    pptp_mean_e pptp_mean_w pptp_mean_n pptp_mean_s

asdoc tabdisp namecounty, cell (houseprice rtp itp pptp) append
```

4. Regression

Data transformation presentation (normal vs ln)

```
hist anualincome, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram anual
    income")
```

```
hist ln_anualincome, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram ln (anual income)")
hist averageage, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram average age")
hist ln_averageage, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram ln (average age)")
hist populationdensity, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram population density")
hist ln_populationdensity, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram ln (population density)")
hist broadbandaccess, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram broadband access")
hist ln_broadbandaccess, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram ln (broadband access)")
hist sfhandduplex, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram SFH and Duplex")
hist ln_sfhandduplex, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram ln (SFH and Duplex)")
hist sqmlivingarea2, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram sqm living area")
hist ln_sqmlivingarea2, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram ln (sqm living area)")
hist buildinglandprice, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram building land price")
hist ln_buildinglandprice, graphregion(color(white)) bgcolor(white) frequency normal title("Histogram ln (building land price)")
```

1 Regression: Fixed effect model on Federal State house price level

```
reg ln_houseprice i.federalstatenumber
estimates store federalstates
outreg2 using federalstates.doc, replace ctitle (federalstates)
reg ln_houseprice i.federalstatenumber $inputvar
estimates store federalstatesinputv
outreg2 using federalstates.doc, append ctitle (federalstatesinputv)
reg ln_houseprice i.federalstatenumber $inputvar buildingpermissionappartment grwinfrastrukturlangfristig baufoerderunglf
estimates store federalstatesinpins
outreg2 using federalstates.doc, append ctitle (federalstatesinpins)
```

2 Regression: Fixed effect model on county house price level

```
reg ln_houseprice i.namecounty
estimate store FEcounty
outreg2 using FEcounty.doc, replace ctitle (FEcounty)
asdoc tabdisp namecounty, cell (_est_FEcounty) append

*3 Regression: Spatial Effect on houseprice c.p.
reg ln_houseprice Urban
estimates store URonly
outreg2 using myregonlyspattial.doc, replace ctitle (URonly)
reg ln_houseprice Urban East East#Urban
estimates store UREWonly
outreg2 using myregonlyspattial.doc, append ctitle (UREWonly)
reg ln_houseprice Urban East South North Urban#East Urban#North Urban#South
estimates store all
outreg2 using myregonlyspattial.doc, append ctitle (all)

*4 Regression: All counties
reg ln_houseprice $inputvar
estimates store all_counties
outreg2 using myreg.doc, replace ctitle (All Counties)
predict resid1

*5 Regression: Urban Rural comparison
reg ln_houseprice $inputvar Urban
estimates store UR
outreg2 using myreg.doc, append ctitle (U-R)
predict resid2

*6 Regression: East West comparison
reg ln_houseprice $inputvar Urban East
estimates store UREW
outreg2 using myreg.doc, append ctitle (U-R_E-W)
predict resid3

*7 Regression: North South comparison
reg ln_houseprice $inputvar Urban East South North
estimates store UREWNS
outreg2 using myreg.doc, append ctitle (U-R_E-W_N-S)
predict resid4

*8 Regression: Robust Regression
```

```
regress ln_houseprice $inputvar,r
estimates store all_counties_robust
outreg2 using robustreg.doc, replace ctitle (All Counties)
reg ln_houseprice $inputvar Urban,r
estimates store Urban_Rural_robust
outreg2 using robustreg.doc, append ctitle (Urban_Rural)
reg ln_houseprice $inputvar Urban East,r
estimates store UE_robust
outreg2 using robustreg.doc, append ctitle (U-E)
reg ln_houseprice $inputvar Urban East South North,r
estimates store NSUE_robust
outreg2 using robustreg.doc, append ctitle (N-S-U-E)
```

*General dominance statistics

```
asdoc domin ln_houseprice $inputvar
asdoc domin ln_houseprice $inputvar Urban
asdoc domin ln_houseprice $inputvar Urban East
asdoc domin ln_houseprice $inputvar Urban East South West
```

***5. Checking Conditions**

*Correlation

```
asdoc corr $inputvar
asdoc corr ln_houseprice ln_gdp1 gva ln_anualincome betteninfvbetrieben purchasingpower
    ln_averageage ln_populationdensity migrationbalance higheducationpersonatjoblocation
    recreationalareakm2per1000inhabi accessibilityhighwayinminutes accessibilityairportinminutes
    accessibilityiceciceinminutes broadbandaccess rentpriceaverage sfhandduplex ln_sqmlivingarea2
    landtransfertax ln_buildinglandprice buildingpermissionappartment grwinfrastrukturlangfristig
    baufoerderunglf
```

*Multicollinearity :

```
reg ln_houseprice $inputvar
asdoc estat vif
reg ln_houseprice Urban $inputvar
asdoc estat vif
reg ln_houseprice East $inputvar
asdoc estat vif
reg ln_houseprice South North $inputvar
asdoc estat vif
reg ln_houseprice Urban East $inputvar
asdoc estat vif
```

*Heteroskedasticity for all models:

```
reg ln_houseprice $inputvar
asdoc estat hettest
asdoc estat imtest, white
reg ln_houseprice Urban $inputvar
asdoc estat hettest
asdoc estat imtest, white
reg ln_houseprice Urban East $inputvar
asdoc estat hettest
asdoc estat imtest, white
reg ln_houseprice Urban East South North $inputvar
asdoc estat hettest
asdoc estat imtest, white
```

*Checking standard errors of urban-rural regression

```
graph twoway (scatter ln_houseprice resid2) (lowess ln_houseprice resid2), graphregion(color(white))
    bgcolor(white)
rvfplot, yline(0) graphregion(color(white)) bgcolor(white)
```

*Sensitivity analysis: Chow-f test

```
*pooled reg
reg ln_houseprice $inputvar
outreg2 using chowreg.doc, replace ctitle (All Counties)
ereturn list
```

*Urban-Rural

```
*u reg
reg ln_houseprice $inputvar if typofregion=="Urban area"
outreg2 using chowreg.doc, append ctitle (Urban)
ereturn list
*r reg
reg ln_houseprice $inputvar if typofregion=="Rural district"
outreg2 using chowreg.doc, append ctitle (Rural)
ereturn list
```

*East-West

```
*e reg
reg ln_houseprice $inputvar if East==1
outreg2 using chowreg.doc, append ctitle (East)
```

```
ereturn list
*w reg
reg ln_houseprice $inputvar if West==1
outreg2 using chowreg.doc, append ctitle (West)
ereturn list

*North-South
*pooled for North-South (n=285)
reg ln_houseprice $inputvar if South==1|North==1
outreg2 using chowreg.doc, append ctitle (NS-Pooled)
*N reg
reg ln_houseprice $inputvar if North==1
outreg2 using chowreg.doc, append ctitle (North)
ereturn list
*S reg
reg ln_houseprice $inputvar if South==1
outreg2 using chowreg.doc, append ctitle (South)
ereturn list

*Normality of Residuals
kdensity r, normal name(kdensity1, replace) title("Kernel Density Plot") graphregion(color(white))
    bgcolor(white)
pnorm r, name(pnorm1, replace) title("Standardized Normal Probability") graphregion(color(white))
    bgcolor(white)

clear
```