**Bachelor Thesis**

**Influencing factors on real estate prices in Germany**

Mercator Chair of Demand Management & Sustainable Transport

Prof. Dr. Arne Strauss

Gideon Gottschalg

Vallendar, May 11th, 2023

**Philipp von Seydlitz-Kurzbach**

20004321

June 7th, 2001 in Hamburg, Germany

Weitersburger Weg 1

56179 Vallendar

# Abstract

# Table of Contents

[Abstract I](#_Toc133941196)

[Table of Contents II](#_Toc133941197)

[List of Tables and Figures IV](#_Toc133941198)

[List of Appendices V](#_Toc133941199)

[1. Introduction 6](#_Toc133941200)

[1.1 Background 6](#_Toc133941201)

[1.2 Purpose and Outline 7](#_Toc133941202)

[2. Literature Review 9](#_Toc133941203)

[2.1 Macroeconomic factors influencing real estate prices 9](#_Toc133941204)

[2.2 Characteristics of the building influencing real estate prices 14](#_Toc133941205)

[2.3 Regional factors influencing real estate prices 16](#_Toc133941206)

[3. Methodology 20](#_Toc133941207)

[3.1 Sample and Data Collection 20](#_Toc133941208)

[3.2 Empirical approach 22](#_Toc133941209)

[3.2.1 Dependent Variable 22](#_Toc133941210)

[3.2.2 Independent Variables 22](#_Toc133941211)

[3.2.2.1 Model 1 22](#_Toc133941212)

[3.2.2.2 Model 2 24](#_Toc133941213)

[3.2.2.3 Model 3 25](#_Toc133941214)

[3.2.2.4 Model 4 26](#_Toc133941215)

[3.3 Data Analysis 26](#_Toc133941216)

[4. Results and Discussion 31](#_Toc133941217)

[4.1 Detailed Interpretation 31](#_Toc133941218)

[4.1.1 Model 1 31](#_Toc133941219)

[4.1.2 Model 2 32](#_Toc133941220)

[4.1.3 Model 3 34](#_Toc133941221)

[4.1.4 Model 4 35](#_Toc133941222)

[4.2 Predictions Model 1 – 4 37](#_Toc133941223)

[4.3 Implications 40](#_Toc133941224)

[4.3.1 Theoretical Implications 40](#_Toc133941225)

[4.3.2 Practical Implications 42](#_Toc133941226)

[5. Conclusion 44](#_Toc133941227)

[5.1 Main Findings and Answer to the Research Question 44](#_Toc133941228)

[5.2 Limitations 44](#_Toc133941229)

[5.3 Future Outlook 44](#_Toc133941230)

[Appendix 45](#_Toc133941231)

[Bibliography 49](#_Toc133941232)

[Affidavit 53](#_Toc133941233)

[AI Declaration 54](#_Toc133941234)

# List of Tables and Figures

[Table 1 – Overview of datasets 20](#_Toc134006872)

[Table 2 – Independent variables Model 1 23](#_Toc134006873)

[Table 3 – Regression output Model 1 31](#_Toc134006874)

[Table 4 – Regression output Model 2 33](#_Toc134006875)

[Table 5 – Regression output Model 3 35](#_Toc134006876)

[Table 6 – Regression output Model 4 36](#_Toc134006877)

[Figure 1 - Predictions Model 1 38](#_Toc134006878)

[Figure 2 – Predictions Model 2 38](#_Toc134006879)

[Figure 3 – Predictions Model 3 39](#_Toc134006880)

[Figure 4 - Predictions Model 4 39](#_Toc134006881)

[Figure 5 – Correlation heat map Model 2 41](#_Toc134006882)

# List of Appendices

[Appendix 1 – Regression results Model 1 houses for sale 45](#_Toc134006883)

[Appendix 2 – Regression results Model 1 apartments for sale 45](#_Toc134006884)

[Appendix 3 – Regression results Model 2 houses for sale 46](#_Toc134006885)

[Appendix 4 – Regression results Model 2 apartments for sale 46](#_Toc134006886)

[Appendix 5 – Regression results Model 3 houses for sale 47](#_Toc134006887)

[Appendix 6 – Regression results Model 3 apartments for sale 47](#_Toc134006888)

[Appendix 7 - Regression results Model 4 houses for sale 48](#_Toc134006889)

[Appendix 8 - Regression results Model 4 apartments for sale 48](#_Toc134006890)

# 1. Introduction

## 1.1 Background

The real estate market is crucial in macroeconomic policy as real estate as an asset class has received much attention recently ([Belke & Keil, 2018, p. 25](#_ENREF_1)). The residential real estate market and its developments are highly interesting to investors, especially private persons, as they need housing. Therefore, private persons are highly affected by price developments when considering investing in their own real estate property. The German real estate market can be described as a stable market as prices remained flat throughout the financial crisis in 2008 ([Voigtländer, 2014, p. 584](#_ENREF_33)) & ([Kröhnert, 2012, p. 4](#_ENREF_20)). Moreover, it is known to be a “safe haven, thanks to the stable economic outlook, the relatively low volatility of real estate yields and rents, and also thanks to the low interest rates” ([Just & Maennig, 2012, p. 6](#_ENREF_18)). This makes real estate investments in Germany attractive for investors and more secure for private persons because market prices are not as volatile as, for example, stock prices. However, the recent increase in interest rates following high inflation rates is not considered and may lead to disturbances in the real estate market soon.

The residential real estate sector is one of the most significant parts of the German economy, with a gross value added of 280 billion euros yearly. Therefore, the development of this industry impacts the macroeconomic situation in Germany ([Kröhnert, 2012, p. 3](#_ENREF_20)). The real estate industry contributes nearly 20% of Germany’s GDP and offers around 3.8 million jobs ([Just & Maennig, 2012, p. 5](#_ENREF_18)). This underlines the importance of the real estate sector for the German economy and conveys that price fluctuations can negatively affect the whole economic situation. Between 2008 and 2012, the prices experienced a drastic increase after a long period of stable prices ([Kröhnert, 2012, p. 4](#_ENREF_20)). Price increases can have several consequences, for example, an increasing ability of private households to borrow against their real estate as collateral ([Belke & Keil, 2018, p. 26](#_ENREF_1)). Therefore, the question arises if the German real estate market price increases can be fundamentally justified. If they are mainly driven by speculation and cannot be explained by macroeconomic developments, there is a greater danger of real estate price bubbles ([Belke & Keil, 2018, p. 26](#_ENREF_1)).

Since the real estate market is so significant for the whole economy, “observed prices must be examined in relation to underlying factors which determine fundamentally justified prices in order to make a judgment about the development of prices and to guide reasonable economic policy decisions” ([Belke & Keil, 2018, pp. 26, 27](#_ENREF_1)). There are several factors potentially influencing real estate market prices. For example, the demographic and economic development between 2000 and 2008 may have been significant for a stable real estate price level. In the years afterward, a decreasing unemployment rate happened simultaneously with increased real estate prices ([Kröhnert, 2012, p. 4](#_ENREF_20)). As there is a trend towards urbanization which keeps cities in Germany growing ([Hiller & Lerbs, 2016, p. 276](#_ENREF_14)), the price development in Germany showed regional differences in the last years. In East German and rural areas, prices remained at the same level or have fallen, while in the other parts and especially in cities, prices experienced a rapid increase ([Kröhnert, 2012, pp. 4, 5](#_ENREF_20)). Besides macroeconomic and regional factors, characteristics “which are correlated with the dwelling itself and those which are correlated with the locations and the surrounding area” ([Wittowsky et al., 2020, p. 44](#_ENREF_34)) impact real estate prices.

## 1.2 Purpose and Outline

As the background information shows, the real estate market is significant for the German economy. Therefore, real estate professionals and investors must understand better how real estate prices are determined. Using quantitative analysis, this paper aims to identify Germany's main influencing factors of real estate prices. The analyzed drivers of real estate prices are sorted into three categories: macroeconomic factors, regional factors and property characteristics.

The first part covers a comprehension of literature findings where the influence of several factors on real estate prices is evaluated. Therefore, a status quo of the existing research is provided. The methodology section introduces the techniques used in the quantitative analysis based on the literature findings and methods. The literature provides several data analysis techniques to find key drivers and uses many independent variables in the regression models. Therefore, under the constraint of data availability and applicability for this thesis, the data analysis techniques and independent variables were chosen based on the literature.

Chapter 4 includes a detailed interpretation of all findings from the regression analyses based on four different datasets. Moreover, the regression formulas were used to calculate predictions and compare the predicted values to the test data. After that, the theoretical and practical implications of the results are presented. The thesis will be concluded with a summary and limitations of all findings.

# 2. Literature Review

The literature review will address several studies regarding influencing factors on real estate prices. The primary market analyzed in this section is Germany. However, some findings and methodologies from comparable markets, i.e., UK and Switzerland are also included. The following chapter is divided into three sections: firstly, macroeconomic factors, such as interest rate and unemployment rate, influencing real estate prices are stated. Secondly, the focus will be set on how the characteristics of a building, such as the age and the living area in square meters, influence the purchase value of flats and houses. Lastly, some findings regarding how the regional characteristic of a property influence the price will be presented. Due to the reason that the empirical part of this bachelor thesis covers the residential real estate market only, the literature review neglects the commercial real estate market.

## 2.1 Macroeconomic factors influencing real estate prices

The first factor which influences real estate prices is interest rates. Long-term interest mortgage rates impact the affordability of real estate as a long-term investment asset and therefore should impact housing demand ([Belke & Keil, 2018, p. 31](#_ENREF_1)). The higher the financing costs for real estate, the lower the demand, as financing costs increase with higher interest rates. A 1% decrease in interest rates leads to a 0.5 to 1.5% increase in house prices over time. An increase in housing prices is observed for a reduction in short-term and long-term interest rates, but there is a slightly weaker response for long-term interest rate declines ([Sutton, 2002, p. 49](#_ENREF_32)). Moreover, [Funk and Drechsel (2017)](#_ENREF_11) discovered that decreasing mortgage rates lead to rising housing prices (p. 42). Still, according to their model evaluation, interest rate volatility can explain only 20% of the real estate price movements. [Jäger and Voigtländer (2006)](#_ENREF_16) used a structural vector autoregression (SVAR) approach: Referring to Germany, they discovered that a 1% interest rate increase would lead to a 0.5% decrease in real estate prices in the following two years, while in the UK it would lead to a 6.5% price decrease (p. 19). Therefore, the effect of interest rate movements on real estate prices seems to be low in Germany compared to other countries.

Another relationship between housing prices and interest rates was discovered by [Belke and Keil (2018)](#_ENREF_1) who found in their regression analysis that the interest rate enters positively into the equation (p. 38). According to them, interest rates in their model “do not primarily capture increasing financing costs related to higher interest rates, but they rather reflect the economic environment and cyclical stance of the economy, which in turn drives demand for housing” ([Belke & Keil, 2018, p. 42](#_ENREF_1)). [Gürtler and Rehan (2008)](#_ENREF_12) also found that increasing interest rates positively influence real estate prices as rising interest rates signify a growing economy (p. 20). A growing economy leads to a positive attitude of the population towards consumption and investment as jobs and income seem to be secured. This can explain the increasing demand and housing prices following rising interest rates ([Gürtler & Rehan, 2008, p. 20](#_ENREF_12)). Following the invasion of Russia in Ukraine, interest rates increased in 2022, but apartment purchase prices and rents remained at the same level. Therefore, higher financing costs for house purchases may lead to a decrease in demand as credit institutions are likely to request higher risk premiums for real estate loans. At the same time, real estate construction companies suffer from higher financing costs. In addition, an evident shortage of building supplies due to disturbed supply chain performance puts pressure on construction companies. They will likely stop or postpone construction projects as higher resell values do not equalize higher financing costs. As the construction costs for real estate projects, material costs and property prices are still high, real estate projects are less profitable than in previous years, which leads to a decrease in supply ([Just, 2023, p. 21](#_ENREF_17)).

Secondly, the influence of income on housing demand and prices is surveyed. Average household income is theoretically an essential factor influencing the demand for housing ([Belke & Keil, 2018, p. 31](#_ENREF_1)). Therefore, a rising average income in the economy should be positively related to increasing housing prices as “demand increases with income” ([Quigley, 2002, p. 6](#_ENREF_22)). [Sutton (2002)](#_ENREF_32) discovered that the growth rate of national income leads to higher house prices over time (p. 49). His analysis shows that a 1% increase in the growth rate of national income is associated with an increase in house prices of 1 to 4%. However, the target market for this analysis is six developed countries, Germany not included. Another study found that “income has a positive and significant influence of 0.33% on German housing demand” ([Bischoff, 2012, p. 8](#_ENREF_2)). Therefore, increasing national income leads to a long-term increase in housing prices.

[Quigley (2002)](#_ENREF_22), too, discovered that local economic conditions such as income are essential determinants for the development of house prices over time, and income is positively related to housing prices (p. 8). Income growth positively influences demand ([Hiller & Lerbs, 2016, p. 286](#_ENREF_14)), meaning that higher income increases demand for real estate properties. In his regression analysis, [Kröhnert (2012)](#_ENREF_20) examined the influence of six factors on real estate prices at the state level in Germany (p. 17). The result revealed that the indicators of available income and the proportion of highly qualified employees explain 61% of the variance in their dependent variable house prices and apartment rents. At the district and township level, income was the most substantial influencing factor on real estate price levels ([Kröhnert, 2012, p. 22](#_ENREF_20)).

However, the income effect is very likely affected by the upward bias meaning that the effect of changes in income on housing demand might be, in fact, more negligible. Therefore, the income effect must be interpreted cautiously ([Bischoff, 2012, p. 8](#_ENREF_2)). Moreover, in his research [Kröhnert (2012)](#_ENREF_20) found that the development of income over time does not influence the development of real estate prices, suggesting that relating predictions of real estate prices with income predictions might be challenging to draw (p. 17).

The following section examines the macroeconomic factor unemployment rate and its influence on real estate prices. Unemployment potentially affects the demand for real estate properties as it directly negatively affects disposable household income ([Belke & Keil, 2018, p. 31](#_ENREF_1)). One can expect that high unemployment rates resulting in less income lead to less demand and, at the same time, more real estate supply due to forced sales. Both effects combined would cause declining real estate prices. Moreover, investors could prevent regions with high unemployment rates as high unemployment rates make a region appear unattractive ([Gürtler & Rehan, 2008, p. 20](#_ENREF_12)). [Kröhnert (2012)](#_ENREF_20) conducted three regression models to analyze influencing factors on real estate prices, and in all three models, population changes proved to be the most influential factors (p. 17). The unemployment rate was one of the factors in measuring population change, and the researcher found statistical proof that a decreasing number of unemployed people leads to a positive development in real estate prices ([Kröhnert, 2012, p. 17](#_ENREF_20)). [Belke and Keil (2018)](#_ENREF_1) executed a POLS regression to determine fundamental determinants of real estate prices in Germany (p. 37). For example, they discovered that higher unemployment rates are associated with lower real estate prices, so these two variables have an inverse relationship.

[Irandoust (2019)](#_ENREF_15) investigated the relationship between unemployment and real estate prices in eight major European countries (p. 154). For Germany and Switzerland, he detected a bidirectional or feedback effect meaning that a change in unemployment affects real estate prices and vice versa. The coefficients were statistically significant and negative, which suggests that a decrease in unemployment will lead to housing price increases. Possible explanations for the bidirectional causality are that Germany and Switzerland are among the developed countries with the lowest homeownership rates. At the same time, both countries show a high degree of labor mobility ([Irandoust, 2019, p. 155](#_ENREF_15)).

[Gürtler and Rehan (2008)](#_ENREF_12) set their target markets to three counties and three cities in Germany (p. 20). In their multiple regression models, they did not find statistical proof of the influence of the unemployment rate on real estate prices. One possible reason could be the selection of a small target market and not Germany as a whole. Another reason for the lack of statistical proof could be that the variable unemployment rate was measured with Germany’s total population. Following the researchers, it would seemingly be more interesting to evaluate the unemployment rate of real estate owners – on the other hand, this is possibly too challenging to be measured ([Gürtler & Rehan, 2008, p. 21](#_ENREF_12)).

Another factor potentially influencing real estate prices in Germany is the population’s age structure. In the literature, age structure is often defined as the “dependency ratio, i.e. the number of persons aged under 15 or above 65 divided by the number of persons aged 16 to 65 expressed in percentage points” ([Belke & Keil, 2018, p. 33](#_ENREF_1)). Being a demographic factor, age structure will likely shape the housing demand and influence real estate prices ([Belke & Keil, 2018, p. 31](#_ENREF_1)). A high dependency ratio means more dependent people like children and pensioners than workers. The working population is the generation that is most likely to afford real estate properties. Therefore, the smaller the size of the working population, the smaller the expected demand for real estate. [Kajuth et al. (2013)](#_ENREF_19) estimated the effects of determinants of regional housing prices in Germany based on a random effects regression analysis (p. 16). They found that a larger share of the middle-aged population puts pressure on housing prices, meaning that a high proportion of working people causes housing prices to increase. This finding is coherent with the results from [Belke and Keil (2018)](#_ENREF_1) who discovered that the age structure enters with a negative impact into the regression equation, indicating that a high proportion of the population, which is not in working age between 15 and 65, decreases the demand for real estate (p. 37).

[Hiller and Lerbs (2016)](#_ENREF_14) estimated a mixed-regressive spatial panel housing price model with city-level data for the German market (p. 287). They found “evidence of negative and economically meaningful effects of population aging on real sales prices of condominiums and single-family homes” ([Hiller & Lerbs, 2016, p. 287](#_ENREF_14)). This suggests that the sales price growth of apartments and single-family homes is negatively related to the growth of the old-age dependency ratio. According to the Hiller and Lerbs’ model, sales prices in the condo segment are more heavily affected by aging than the family home segment. However, the relationship between an aging population and rent growth is positively associated and therefore different than for apartment and single-family sales prices. This phenomenon can be explained by the increasing demand for apartments with housing services with an aging population ([Hiller & Lerbs, 2016, p. 287](#_ENREF_14)).

Not only the population's age structure can influence real estate prices, but also the population measured in absolute terms and the development of the population. [Gürtler and Rehan (2008)](#_ENREF_12) discovered an influence of the variable number of inhabitants on apartment and housing prices (p. 21). The positive relationship means more inhabitants per square kilometer cause higher real estate prices. A possible explanation is a higher demand for real estate in an area with a high population density due to the increasing attractivity of the region ([Gürtler & Rehan, 2008, p. 9](#_ENREF_12)). Using a linear regression [Kröhnert (2012)](#_ENREF_20) analyzed influencing factors on real estate price developments in the German market and concluded that the population development from 2005 until 2010 was the most decisive influencing factor for the development of real estate prices between 2007 and 2012 (p. 17).

However, population size determines housing demand only indirectly because the number of households is the decisive factor ([Just & Maennig, 2012, p. 32](#_ENREF_18)). The number of inhabitants in Germany declined by almost 800,000 between 2003 and 2010, but the number of households in Germany continued to rise during these years. This is due to decreasing average household size. In 1991 the average household in Germany comprised almost 2.3 people, but currently, the average household size is just over two people. [Belke and Keil (2018)](#_ENREF_1) analyzed the impact of the number of households in their pooled ordinary least squares estimation for the timeframe of 1995 until 2010 (p. 37). They indicated that the number of households has a positive sign meaning that a higher number of households leads to increasing demand and therefore increasing prices of real estate properties. In addition to that, the living space per person has increased significantly over the last few decades. Currently, each inhabitant occupies an average of around 47m2 living space, around 12 m2 more than in 1980 ([Just & Maennig, 2012, p. 33](#_ENREF_18)). The increasing number of households and an increased average living space per person are two possible reasons for the increasing demand for real estate, although the population in Germany has been declining in the last few years.

## 2.2 Characteristics of the building influencing real estate prices

The literature mentions several characteristics of a house or apartment that significantly influence their retail prices. These characteristics are, for example, the age, living area in sqm, building features and heating type. [Wittowsky et al. (2020)](#_ENREF_34) concluded from their research that the most critical factors for price determination are directly related to the property’s condition, irrespective of the property type (p. 65).

[Gürtler and Rehan (2008)](#_ENREF_12) investigated that the variable age of the building has the most substantial influence in their regression models (p. 17). This specific factor got a negative impact indicating that the older a building, the lower the property’s price. This finding is coherent with the results of [Ferlan et al. (2017)](#_ENREF_9) who also concluded a negative relationship between the building age and the value of an averagely maintained building (p. 140). This means that the price is decreasing with the increasing age of buildings. [Wittowsky et al. (2020)](#_ENREF_34) found that the property's condition shows comparatively high coefficients compared with other independent variables, such as the number of rooms or the living area (p. 63). Therefore, newly built or recently renovated apartments reveal a much higher price per square meter. There are two possible explanations for this phenomenon: Firstly, the age of a building indicates its level of quality because older buildings typically require more maintenance than younger ones ([Gürtler & Rehan, 2008, p. 17](#_ENREF_12)). Depending on the age and the condition of the building, it will to a greater extent be subject to regular maintenance and larger investment repair work. The renovation of wall and roof frontage is the most extensive and expensive maintenance work. Therefore, wall and roof frontage renovation represents the most significant influence on the value of the residential real estate ([Ferlan et al., 2017, p. 137](#_ENREF_9)). This suggests that although the age of the building was proven to be one of the most influential factors on real estate prices, it can be misleading if a building is old but newly and modernly renovated.

The second explanation for decreasing prices with a higher age of the building is that the age includes information about the floor plan of a building. Most buildings were constructed with many small rooms suitable for large families in the 1950s. However, in the 80s, fewer and larger rooms were wanted and became prevalent. These two reasons indicate that many investors and private persons prefer buying a younger house ([Gürtler & Rehan, 2008, p. 17](#_ENREF_12)).

Another essential characteristic of a building that influences the property's price is the property's size, measured in the number of rooms or square meters. The number of rooms in a property defines a type of apartment. It is an essential piece of information for interested apartment buyers because the purchase decision is linked to the size of their household ([Ferlan et al., 2017, p. 137](#_ENREF_9)). [Gürtler and Rehan (2008, p. 17)](#_ENREF_12) and [Soot et al. (2016)](#_ENREF_24) indicated from their regression results that the variable living area is associated with a negative relationship with the property price per square meter, which means that the higher the living area measured in square meters is, the lower the price per square meter. This can be explained by the fact that above a particular living space, individuals are willing to pay less for an additional square meter ([Gürtler & Rehan, 2008, p. 17](#_ENREF_12)). However, [Wittowsky et al. (2020)](#_ENREF_34) conducted a regression analysis for the city of Dortmund and analyzed influential factors on prices for owner-occupied houses as well as apartments and rental apartments (p. 63). Their result was that the size of the living area is positively correlated with the housing price, but this is only relevant for owner-occupied apartments and not for owner-occupied houses and rental apartments. Therefore, the literature suggests different relationships between the living area measured in square meters and property prices.

[Ferlan et al. (2017)](#_ENREF_9) mention that the size of a property measured in the number of rooms is among the first ten factors influencing property prices and therefore an essential characteristic of a property influencing the price (p. 139). The relationship between the number of rooms and the price for all three types of properties – owner-occupied homes, owner-occupied apartments and rental apartments – is negative, indicating that the more rooms, the lower the price ([Wittowsky et al., 2020, p. 63](#_ENREF_34)). A possible explanation could be that a trend is observable for newly built and modern properties, whereas fewer but larger rooms are more attractive to owners and tenants ([Wittowsky et al., 2020, p. 65](#_ENREF_34)).

[Ferlan et al. (2017)](#_ENREF_9) included the floor level of an apartment as well as the availability of an elevator in their analysis (p. 139). They revealed that apartments in a property without an elevator located on the first floor have an average 2.46% value increase compared to apartments on the first floor. For buildings with an elevator, they showed an increase in the value of a property on a higher floor. On average, the value increased by 1.3% per floor, and the most significant increase in value is for first-floor properties.

## 2.3 Regional factors influencing real estate prices

The third category of factors influencing real estate prices comprises regional factors that characterize the property's location. Germany consists of 16 largely autonomous states, each with their own constitution. Moreover, they each have different characteristics, including social and demographic factors ([Just & Maennig, 2012, p. 42](#_ENREF_18)). Therefore, the German real estate market can be described as quite fragmented compared to other European countries ([Just & Maennig, 2012, p. 56](#_ENREF_18)). This chapter will provide information about regional differences in the German real estate market influencing prices. Furthermore, other regional factors, such as infrastructure, will be signposted, potentially influencing real estate prices. For example, schools, roads, or hospitals in the surrounding of the real estate property might increase the demand for housing and thus influence the price ([Belke & Keil, 2018, p. 31](#_ENREF_1)).

As previously mentioned, the demographic structure of a country is a significant factor shaping the demand for real estate properties. Germany faced major disruptions during the last century in the form of the two world wars followed by the partition, due to which ten states belonged to the Federal Republic of Germany (FRG) in the west, now called “old states”. The German Democratic Republic in the east consisted of five states, now called “new states”. The reunification in 1990 merged all 16 states in one country, but social and economic differences between these states influence the regional real estate markets to date ([Just & Maennig, 2012, pp. 41, 42](#_ENREF_18)). Over the last 30 years, the so-called new states experienced a severe demographic change. After the reunification in 1990, the new states lost around 720,000 inhabitants (4% of east Germany’s population), and the population decreased by another 850,000 in the 2000s. Forecasts expect a continuing shrinking population in East Germany until 2030, while the population in West Germany is predicted to remain at the same level ([Henger & Voigtländer, 2015, p. 9](#_ENREF_13)).

[Cajias et al. (2020, p. 1038)](#_ENREF_6) conducted a cluster analysis of the German residential market and assigned the 380 regions of their sample to price and liquidity clusters one or two with the aim to “assign regions to the same cluster so that the dissimilarity within a cluster is minimized and maximized between the clusters” ([Cajias et al., 2020, p. 1037](#_ENREF_6)). Most regions in East Germany are assigned to price and liquidity cluster two. The highest concentration of regions in price and liquidity cluster one is located in southern Germany ([Cajias et al., 2020, p. 1043](#_ENREF_6)). Regions in price cluster one experienced a real estate price increase of 56.75% over the past six years, while regions in price cluster two experienced a price increase of 32%. Therefore, the regions in price cluster one, located mainly in West Germany, can be identified as highly demanded regions relative to regions in price cluster two ([Cajias et al., 2020, p. 1038](#_ENREF_6)). Several factors mentioned in the paper distinguish the two clusters and are possible explanations for the different price development of the two clusters. The first significant difference between both clusters is the population. Cluster one experienced an increase of 3.84% in population and 2.4% in working population, which is way higher than in cluster two. This growing population trend in most parts of West Germany implies a higher demand for living space, resulting in higher prices ([Cajias et al., 2020, p. 1038](#_ENREF_6)).

Another factor that differs between East and West Germany is the GDP which is “commonly used as an indicator of the economic health of a country, as well of a country’s standard of living” ([Just & Maennig, 2012, p. 44](#_ENREF_18)). [Cajias et al. (2020)](#_ENREF_6) found that the regions in cluster one, mainly in West Germany, are characterized by higher real GDP (p. 1039). [Just and Maennig (2012)](#_ENREF_18) stated that the districts in the old states in West Germany have, on average, an almost 50% higher GDP per capita than the districts in East Germany, which is a clear sign of increased economic health in the West German regions (p. 45). Moreover, there exists an east-west gap comparing regional unemployment rates. The average unemployment rate in West Germany in 2015 was 5.6% and in East Germany 9.4% ([Just & Maennig, 2012, p. 47](#_ENREF_18)). In addition, [Cajias et al. (2020)](#_ENREF_6) found out that the regions in cluster two, which is characterized by lower price development of real estate and therefore lower demand, register a higher unemployment rate and added that the disposable income is lower (p. 1039). A lower GDP, higher unemployment rate as well as lower disposable income are possible explanations for the lower demand for real estate properties in East Germany.

The quality of infrastructure in the region of real estate properties also influences demand and price development. People tend to move to economically strong regions with a well-organized infrastructure. Especially the larger German cities, which are highly populated regions, offer a well-organized infrastructure with local transportation services such as subways and public busses ([Just & Maennig, 2012, p. 48](#_ENREF_18)). Moreover, communication infrastructure has become a critical locational factor as it is essential for the development of an economy. Significant differences exist between West and East Germany as well as urban and rural areas regarding broadband connections. On average, 39% of households in West and 29% of households in East Germany have access to broadband connections with a minimum of 50 Mbit/s ([Just & Maennig, 2012, p. 49](#_ENREF_18)). This digital divide may be another factor influencing differences in Germany's real estate price development.

Social infrastructure, such as the availability of healthcare facilities, is another factor shaping the demand for real estate in a region. In their regression analysis, [Belke and Keil (2018)](#_ENREF_1) included the number of hospitals per 1000 inhabitants (p. 38). It is used as a proxy for the quality of available public infrastructure. The number of hospitals with a positive coefficient demonstrates a higher demand in regions with better public infrastructure. [Kröhnert (2012)](#_ENREF_20) used the number of doctors per square kilometer as a proxy for the density of social infrastructure and concluded that districts and communities with a high density of social infrastructure have a higher price level (p. 22). However, analyzing the influence of social infrastructure on the price development of real estate properties, the regression model showed a low explanatory power. Therefore, the regression models examined social infrastructure's influence on price level but not on price development ([Kröhnert, 2012, p. 24](#_ENREF_20)).

The region’s economic factors also shape regional demand for real estate. [Belke and Keil (2018)](#_ENREF_1) analyzed the effect of the economic structure on real estate demand (p. 38). The “economic structure is defined as the number of persons working in the services sector divided by the sum of persons working in the manufacturing and agriculture sector” ([Belke & Keil, 2018, p. 34](#_ENREF_1)). In their regression formula, the economic structure enters with a negative sign indicating that a higher share of industrial and manufacturing labor in a region is linked to a higher demand for real estate and therefore higher prices. [Kröhnert (2012)](#_ENREF_20) included the number of companies per square kilometer in his analysis and discovered a direct link between the number of companies in a region and the price level, meaning that the more people work in a region, the higher the prices (p. 22). [Just and Maennig (2012)](#_ENREF_18) compared West and East German regions by developing company foundations per 1000 inhabitants as a level of innovation (p. 52). Therefore, the economic development of a region can be measured. West German outperform East German regions, and urban outperform rural districts, which indicates less economic development in the lower-performing regions.

The commercial tax rate level is a reliable indicator of a region's locational attractiveness and economic power ([Just & Maennig, 2012, p. 53](#_ENREF_18)). [Gürtler and Rehan (2008)](#_ENREF_12) found small positive correlations between the commercial tax rate and the real estate price level because higher commercial tax rates indicate a growing economy in the region (p. 21). Therefore, the demand for real estate is higher. [Just and Maennig (2012)](#_ENREF_18) concluded that urban areas showed higher tax rates than rural areas, possibly explaining lower real estate demand in rural areas (p. 53).

# 3. Methodology

## 3.1 Sample and Data Collection

The study, which is based on advertising data from the German real estate market, consists of 6 different datasets separated into two categories: panel datasets and cross-sectional datasets. Each category is organized into three separate datasets: houses for sale, apartments for sale, and apartments for rent. The panel datasets cover transactions between 2007 and 2022 and focus on 13 big cities in Germany: Berlin, Cologne, Hamburg, Duisburg, Bremen, Munich, Düsseldorf, Dortmund, Hannover, Frankfurt am Main, Essen Nürnberg and Stuttgart. However, some transactions are also included, which cannot be allocated to one of these cities. The cross-sectional datasets contain transactional data from 2022 for all of Germany. Table 1 provides an overview of the sample sizes in the datasets.

Table 1 – Overview of datasets

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Sample size** | **Sample size after cleaning** |
| Panel data (2007 – 2022) houses for sale | 367, 583 | 132,431 |
| Panel data (2007 – 2022) apartments for sale | 927,319 | 364,240 |
| Cross-sectional data (2022) houses for sale | 259,347 | 71,049 |
| Cross-sectional data (2022) apartments for sale | 149,747 | 71,952 |

The RWI – Leibniz Institut für Wirtschaftsforschung provided the datasets as campus files for research purposes. These are extractions of the Scientific Use Files of RWI-GEO-RED. [Schaffner (2020)](#_ENREF_23) published a detailed data description for the used datasets. However, the data description was published in 2020, meaning the years do not match the datasets. ImmobilienScout24, which “is the largest internet platform on real estate in Germany” ([Schaffner, 2020, p. 6](#_ENREF_23)), is the source of the datasets. It has a market share of about 50% of real estate objects offered for rent or sale in Germany ([Schaffner, 2020, p. 6](#_ENREF_23)) and includes properties marketed by private persons but also by real estate agents or banks ([Soot et al., 2016, p. 4](#_ENREF_24)). Therefore, the dataset has many observations and is a good source of transactional movements in the German real estate market due to ImmobilienScout24’s high market share. ImmobilienScout24 is used for private and commercial entities, but the dataset only includes residential real estate. The datasets contain various information about the property and its characteristics filled out by the property owner who advertises it on ImmobilienScout24. Regression model 1 analyzes how the characteristics influence offering prices.

However, there are some limitations of ImmobilienSout24 as a data source. The owner fills out a property's characteristics, and the information is not verified. Therefore, the data quality depends on the accuracy and honesty of the owner’s presentation and property description ([Wittowsky et al., 2020, p. 50](#_ENREF_34)). Some characteristics are not mandatory to fill out in the registration form when advertising a real estate object on ImmobilienSout24, which leads to many missing values in the feature description of a property. Therefore, “it is not known whether such missing values actually reflect a situation (e.g., no balcony), or whether the owner just did not want to or simply forgot to tick that box while filling in the immobilienshout24 registration form” ([Wittowsky et al., 2020, p. 50](#_ENREF_34)). Therefore, cleaning the dataset before starting the data analysis and developing assumptions to remain with a high sample size was necessary. The assumptions are outlined in the variable description part of this paper, and Table 1 states the sample sizes after the cleaning process. Another limitation of ImmmobilienScout24 is that owners enter an offering price at which they are willing to sell or rent a real estate object. As the offering price is not binding, the dataset does not include actual transaction prices ([Wittowsky et al., 2020, p. 6](#_ENREF_34)) & ([Schaffner, 2020, p. 6](#_ENREF_23)).

To analyze the influence of macroeconomic factors on real estate prices, various macroeconomic sizes were collected from different sources and merged into the transactional data. The data source of each variable is outlined in the variable description. The smallest possible period for the macroeconomic factors is monthly data. Therefore, the average offering price for each quarter between 2007 and 2022 was calculated and merged with the corresponding macroeconomic sizes. Regression model 2 analyzes the influence of macroeconomic factors on average monthly on real estate prices, and model 3 is a combination of macroeconomic factors and characteristics.

The cross-sectional analysis collected macroeconomic factors for 2022 on a state basis. As Germany consists of 16 states, the macroeconomic sizes were merged into the advertisements based on the property’s state. It should be noted that the cross-sectional datasets for 2022 exclude the state of Saxony. Therefore only 15 states are included in the regression analysis. The goal of merging macroeconomic factors on a state basis to the dataset is to find drivers of regional real estate price differences in Germany. Model 4 evaluates the relationship between regional macroeconomic sizes as well as housing characteristics and offering prices.

## 3.2 Empirical approach

### 3.2.1 Dependent Variable

The dependent variable used in the regression models is the sales price of houses or apartments. As mentioned, the sales price is the offering price provided by the property owner and can differ from the actual transaction price. There were no missing values of the offering price in the datasets, as it is a mandatory input field when publishing an advert on ImmobilienScout24. [Soot et al. (2016)](#_ENREF_24) mention that there is likely to be a reduction between the offer and purchase price of about 10 to 15%. However, in this paper, the original offering price from the datasets will be used for the data analysis as a reduction did not influence the accuracy of the regression outputs. For regression model two, the average sales price of all adverts per month was calculated and used as an independent variable. The unit of the dependent variable sales price is €.

### 3.2.2 Independent Variables

#### 3.2.2.1 Model 1

For the regression model 1 the influence of property characteristics on sales prices was analyzed. All the characteristics were part of the adverts and filled in by the real estate objects’ owners when publishing their advert. As not all characteristics are mandatory fields for the publishers, cleaning the dataset was necessary to exclude rows with missing or implausible values. This led to a reduction in the sample sizes, as shown in Table 1. Therefore, a selection of relevant characteristics which do not have many missing values was made. Table 2 provides an overview of all characteristics used for regression model one.

Table 2 – Independent variables Model 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Type of variable** | **Used in dataset houses for sale** | **Used in dataset apartments for sale** |
| Age | Continuous | Yes | Yes |
| Living space in m2 | Discrete | Yes | Yes |
| Number of rooms | Discrete | Yes | Yes |
| Land area in m2 | Discrete | Yes | No |
| Number of bedrooms | Discrete | Yes | Yes |
| Number of bathrooms | Discrete | Yes | Yes |
| Parking space | Nominal | Yes | No |
| Cellar | Nominal | Yes | No |
| Elevator | Nominal | No | Yes |
| Object condition | Ordinal | Yes | Yes |

The variable age was not present in the datasets. Therefore, it was calculated by subtracting the construction year from the date the advert was closed. The end date implies the transaction date.

The discrete variables living space in m2,number of rooms, land area in m2, number of bedrooms and number of bathrooms included some implausible values that seemed too high when analyzing the dataset. Therefore, a range of accepted values was defined for all these variables, and rows with values outside this range were deleted.

The nominal variables parking space, cellar, and elevator hold the values Yes or No in the dataset. These values were replaced with 0 for No and 1 for Yes so that the regression models could include the variables. As these are also voluntary fields when publishing an advert on ImmobilienScout24, there were a lot of missing values for the three variables. In order to maintain a large sample size, the assumption was made that a missing value for one of these three variables means that there is no parking space or cellar. Therefore, all missing values were replaced with No in the data cleaning process. Many rows would have been deleted without this assumption, especially in the years before 2018, suggesting that the assumption was necessary to keep many observations for the entire timeframe.

The ordinal variable object condition had nine different values in the dataset, which specified the property condition at the time of the advert. The categories for this variable are dilapidated, needs renovation, well-kempt, modernized, reconstructed, completely renovated, like new, first occupancy after reconstruction and first occupancy. After having deleted rows with the column value “not specified” or “by arrangement,” these categories were replaced with a number from 0 to 8. Therefore, the higher the number in the column object condition, the better.

#### 3.2.2.2 Model 2

In regression Model 2, the influence of macroeconomic factors on the overall price development was analyzed. Therefore, several macroeconomic sizes were merged to the average prices for every month as this is the smallest common time unit of most datasets. The following list gives an overview of the independent variables used in Model 2:

* Interest rate is defined as the 10-year long-term government bond yield for Germany. The data is sourced from [Organization for Economic Co-operation and Development (April 2023)](#_ENREF_21). This variable shows the development of the long-term interest rate environment in the economy and thereby indicates the financing costs when financing a real estate property ([Belke & Keil, 2018, p. 34](#_ENREF_1)). The data is available every month, so only slight adjustments to the date format had to be made to merge it with the average monthly offering prices from ImmobilienScout24.
* Unemployment rate is defined as the percentage of people in the working population who are not employed. The working population is all people between the age of 15 and 74. Data is taken from [Statistisches Bundesamt (March 2023)](#_ENREF_29). The data is available every month, so only minor adjustments to the date format had to be made in order to merge it with the average monthly offering prices from ImmobilienScout24.
* Income is defined as the total households’ disposable income to capture the household income development in the regression analysis. The data is taken from [Bundesbank (February 2023)](#_ENREF_4). The time unit was quarterly data. Therefore, rows were added to get monthly entries, and the total disposable income per quarter was divided by three under the assumption that the disposable income of each month within a quarter is the same. This step was necessary as no monthly data was available for the disposable income. The income’s unit is billion euros.
* Investment in housing construction is measured as a percentage of GDP and an indicator of the investment and construction activities in the German residential real estate market. The data is taken from [Bundesbank Real Economic Indicators (March 2023)](#_ENREF_5) and the time unit was also quarterly. Therefore, rows were again added to adjust the time unit to monthly entries, and the assumption was made that the percentage values were the same for each month within a quarter.
* Number of construction permissions is a measure of construction activity in Germany as it indicates the number of construction permission for residential apartments in Germany per year. The data entries are simplified and need to be multiplied by 1000. The data was taken from [Statistisches Bundesamt (August 2022)](#_ENREF_27) and [Statistisches Bundesamt (April 2023)](#_ENREF_26). This data was only available yearly. Therefore, the assumption was made that the number of construction permissions is equal for each month within a year, and all yearly values were divided by 12 to get monthly data.
* GDP per capita is a measure of the economic development in Germany. The data was taken from [Statistisches Bundesamt (February 2023)](#_ENREF_28). It was available only every year. Therefore, it was necessary to divide the yearly number by 12 to get monthly data.

#### 3.2.2.3 Model 3

For Model 3, the independent variables of Model 1 and Model 2 were combined into one large set of independent variables. Therefore, the macroeconomic data was merged into the original dataset containing the properties’ characteristics based on the month the advert was closed.

#### 3.2.2.4 Model 4

Model 4 analyzed the cross-sectional dataset for 2022 to explain regional differences influencing real estate prices across Germany. Therefore, the characteristics of Model 1 were used combined with several macroeconomic factors. The macroeconomic factors were collected for 2022 on a state level and merged into the dataset containing the characteristics based on the property's state. The following macroeconomic variables were collected for each of the 16 German states:

* Population density is a measure of the number of inhabitants per square kilometer. The data was taken from [Statistisches Bundesamt (September 2022)](#_ENREF_30). No data was available for 2022, but as mentioned in the literature, population density is an important demographic factor influencing real estate demand. Therefore, the data for 2021 had to be used.
* Unemployment rate is defined as the percentage of unemployed people in the working population. The working population is all people between the age of 15 and 74. Data is taken from [Bundesagentur für Arbeit (January 2023)](#_ENREF_3).
* GDP per capita is a measure of the economic development in Germany. The data is taken from [Statistische Ämter des Bundes und der Länder (March 2023)](#_ENREF_25).
* Income describes the average pre-tax income per full-time worker in each state and reflects the state’s income level. The data was taken from [Stepstone (February 2023)](#_ENREF_31).
* Vacancy rate captures the proportion of unoccupied apartments in 2021 as a percentage of the total number of apartments. It is an indicator of the supply situation in the region. No data for 2022 was publicly available. Therefore, the most recent data was used. The data was taken from [empirica (December 2022)](#_ENREF_8).

## 3.3 Data Analysis

Hedonic price modeling is the theoretical concept used in this study to explain real estate prices. The basic idea of hedonic pricing models is to use individual property characteristics contributing to a property’s value as determinants of real estate prices ([Belke & Keil, 2018, p. 30](#_ENREF_1)). The concept attempts to calculate a property’s price by summing up the individual characteristics. This is done via a regression model that uses residential offering prices as the dependent variable and characteristics, as described in the previous chapter, as independent variables. The hedonic price modeling uses traditionally ordinary least square (OLS) regression models ([Wittowsky et al., 2020, p. 46](#_ENREF_34)). Thereby the effect of the independent variables on the dependent variable offering price is measured with the statistical method of a multiple linear regression model. The data analysis was conducted in Python via the Scikit-learn library to split the datasets into train and test data and the Statsmodels library to calculate all coefficients and relevant parameters of the OLS regression models. The algorithm randomly selected 80% of the observations as training data and 20% as test data. Using the test data, the models were used to calculate predictions based on the independent variables. The predictions were then compared to the test dataset and plotted on a scatter plot graph using the Seaborn library. The regression formulas for Model 1 look as follows:

House prices:

Apartment prices:

For regression Model 2, which tries to explain the average monthly house or apartment prices by fundamental macroeconomic factors an approach used by [Belke and Keil (2018)](#_ENREF_1) is followed. They used a two-way fixed effect panel model, which includes year-specific and region-specific fixed effects. However, due to complexity reasons and as the panel dataset does not cover the whole German market but focuses on the big cities in Germany, the region-specific effects are excluded in this study. Therefore, this model includes year-specific effects which “capture any global factors affecting all regions in the same way ([Belke & Keil, 2018, p. 36](#_ENREF_1)).” Again, an OLS regression algorithm provided by Statsmodels in Python calculates the regression coefficients. The regression formula for model 2 for average house and apartment prices per month looks as follows:

For regression Model 3, the year-specific effects of Model 2 are added to the regression formula of Model 1. Therefore, this model includes both characteristics and macroeconomic sizes, and the regression formulas look as follows:

House prices:

Apartment prices:

Regression Model 4 includes the region-specific fixed effects as a cross-sectional dataset of the German real estate market in 2022 is analyzed. As real estates are very heterogeneous goods, they depend to a large extent on the property’s characteristics, the surroundings, the larger region and its infrastructure, and several other factors. Therefore, it is reasonable to assume that besides the property’s characteristics, region-specific fixed effects influence real estate prices ([Belke & Keil, 2018, p. 36](#_ENREF_1)). However, it is impossible to include several surrounding features, e.g., infrastructure features, in the analysis because this data is not publicly available on the district level. Therefore, macroeconomic sizes at Germany's state level were added to the model as region-specific fixed effects. The regression formulas look as follows:

House prices:

Apartment prices:

The four models applied to the houses and apartments for sales dataset will be subject to a deeper analysis and interpretation in the next chapter. Firstly, the value R2 is studied, the coefficient of determination. It shows how well the independent variables' variance explains the dependent variable's variance. The higher the value of R2, the better the model to explain the dependent variable. The following significant value of the regression output is the F Statistic which examines if all independent variables combined are statistically significant. A significant model indicates that the regression model, including the coefficients of the independent variables, fits better to the dataset than a model without coefficients and only an intercept. The number of stars next to the regression output shows at which p-level the model is significant.

Looking at the explanatory variables, the coefficients, marked as betas in the regression formula, are subject to interpretation. They indicate how according to the model, the dependent variable responds if the independent variable changes by one unit. Moreover, the coefficients’ prefixes expose if the independent and dependent variables have a linear or inverse relationship. Lastly, the p-values of the coefficients, which are calculated using a t-test, are used to test the null hypothesis that the coefficient has no effect and is equal to zero. A statistically significant coefficient (p-value below 0.1) indicates that changes in the variable impact the dependent variable, and therefore the null hypothesis can be rejected ([Flandorfer, 2023](#_ENREF_10)).

# 4. Results and Discussion

## 4.1 Detailed Interpretation

### 4.1.1 Model 1

Table 3 – Regression output Model 1

Ein Bild, das Text, Quittung, Screenshot enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

Model 1 tries to explain house and apartment prices by their characteristics. Table 3 provides the regression output for Model 1. The R2 implies that the model only explains 45% of the house price’s variance and 47% of the apartment price’s variance. This can be explained by the fact that housing is a multidimensional heterogeneous good and as real estate objects are immobile, locational characteristics are essential when determining the price of a real estate object ([Wittowsky et al., 2020, p. 46](#_ENREF_34)). Therefore, in theory, more neighborhood characteristics and regional features would be needed to explain the house or apartment prices accurately. The f-statistic suggests that the whole model is highly statistically significant, and in addition, all coefficients’ p-values are below the significance level of 0.01. Thus, the regression model fits the observed data quite well.

The variable age has the expected negative effect on house prices, indicating that the older the property, the lower its value. However, it is surprising that there is a positive sign in the regression formula for apartment prices. A possible explanation for this is that, as mentioned in the literature, the age of a building can be misleading as it becomes less relevant if the apartment is newly and modernly renovated. This is underlined by the positive object condition coefficient, meaning that the higher the object condition’s category, the higher the property’s price level.

The living space measured in square meters enters the regression formula with a positive sign indicating that the larger the property, the higher its price level. The negative sign of the number of rooms is in line with findings from the literature. A trend towards fewer but larger rooms can possibly explain it. The variable land area also enters the regression formula of the housing prices with a positive sign. However, the coefficient is comparably small, demonstrating that a higher land area has a negligible influence on the price level.

The variables number of bedrooms and number of bathrooms both have positive coefficients, which can be surprising since the total number of rooms has a negative coefficient. However, the number of rooms excludes by definition bathrooms ([Schaffner, 2020, p. 12](#_ENREF_23)). The positive coefficient of the number of bedrooms can be due to a higher attractivity of the property for families with many children who need many bedrooms but not necessarily many other rooms.

The categorical variables parking space and elevator for house prices and elevator for apartment prices all hold a positive sign which suggests that the presence of this feature increases the property’s value. The positive sign of the elevator is coherent with findings from the literature.

### 4.1.2 Model 2

The purpose of Model 2 is to explain the monthly price development of real estate prices by macroeconomic sizes. Table 4 provides the regression output for model 2. The model’s R2 is 0.98 for average monthly house and apartment prices, thereby nearly optimal. Therefore, according to this model, the monthly development can be explained very well by the choice of independent variables. However, this must be interpreted cautiously as multicollinearity between the variables might exist. This would mean that the variables have a similar timely development, although they influence each other only to a small extent. The f-statistic of the model and the coefficients’ p-values suggest that the whole model and individual coefficients are statistically significant on a significance level of at least p < 0.1.

Table 4 – Regression output Model 2

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

The interest rate’s negative sign in the regression formula states that an increasing interest rate leads to lower average real estate prices. This finding is coherent with the literature and can be explained by increasing financing costs for real estate prices. Therefore, real estate properties get less affordable for private persons, and the demand falls.

The unemployment rate enters the regression formula with a positive sign which is not in line with findings from the literature. The unemployment rate was expected to have a negative sign as, according to the literature, a higher unemployment rate results in less income and, thereby, less demand for real estate. Therefore, it is unclear if this model, including the unemployment rate, can be applied in praxis, although the coefficient is highly statistically significant.

Disposable income and GDP per capita hold a positive sign suggesting that they are positively related to house and apartment prices. Most literature articles mentioned the same relationship between income and price development. GDP per capita and disposable income are indicators of national economic development, meaning that a positive development of these factors indicates a rising economy. Therefore, a rising economy leads to a positive price development of real estate objects.

Investment in housing construction and the number of construction permissions reflect the supply side of real estate and enter with a positive sign. The positive relationship could be a supply-side reaction to increasing demand in the big cities meaning that construction companies and the government increase their construction activities following higher demand.

### 4.1.3 Model 3

Model 3 combines Model 1 and Model 2 as the macroeconomic factors used in Model 2 are merged into the dataset containing all characteristics based on the month the advert was published. Table 5 provides an overview of the regression output for house and apartment prices as dependent variables. The R2 values of 0.57 for house prices and 0.64 for apartment prices have improved compared to Model 1, where the R2 values were 0.45 and 0.47. Therefore, including the macroeconomic variables as year-specific fixed effects in the model has improved its explanatory power. This indicates that real estate prices cannot only be explained by their characteristics but also by macroeconomic sizes that influence the price level. Again, this model and all coefficients are highly significant on a significance level of p < 0.01.

Most coefficients kept the same signs as in the models for characteristics and macroeconomic factors only. However, some variables changed their sign in this model. The number of construction permission now enters with a negative sign into the regression formula. This could be explained by more supply relative to the demand, which can lead to falling prices. The variable age now has for house prices and apartment prices is a negative sign. As mentioned, the object condition is often more relevant because a newly and modernly renovated house is worth more than an unrenovated younger house.

Table 5 – Regression output Model 3

Ein Bild, das Text, Screenshot, Quittung enthält.

Automatisch generierte Beschreibung

*Source: own calculation*

### 4.1.4 Model 4

The purpose of Model 4 is to explain regional differences in real estate price levels in the German market. As discovered in model 1, characteristics of a real estate property influence prices largely. Therefore, this model combines the properties’ characteristics and macroeconomic sizes of the real estate’s state. Table 6 provides an overview of the regression output of model 4. The R2 value with house prices as the dependent variable is 0.42, and apartment prices as the dependent variable is 0.57. Therefore, the model's explanatory power is much higher for apartment prices than for house prices. The f-statistic suggests that the model is statistically significant. The coefficients’ p-values of the macroeconomic sizes are below the significance level of at least p < 0.05. However, for the characteristics, one variable (Number of bedrooms for apartment prices) is not statistically significant, and the variable age for apartment prices is only statistically significant on a significance level p < 0.01. This is surprising as the characteristics in model 1 and model 3 were all highly significant.

Table 6 – Regression output Model 4

Ein Bild, das Text, Quittung, Screenshot enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

The positive coefficient of the population density indicates a higher demand for real estate properties in regions with a higher population density. This can be due to a higher attractivity of a region which attracts more people seeking a house or apartment. This finding is in line with the literature where it was mentioned that the price level of real estate properties is higher in regions with more inhabitants per square kilometer.

In contrast to Model 2, where the unemployment rate showed a positive sign, in this model the unemployment rate enters with a negative sign into the regression formula. This result is coherent with the literature findings, where the unemployment rate was higher in regions with lower price levels. A higher unemployment rate can result in less disposable income in a region and therefore less demand for real estate. In combination with the income per person and the GDP per capita, the unemployment rate indicates a region’s economic health. The GDP per capita has a positive sign which is in line with findings from the literature. Findings from the literature cannot explain the income’s negative sign and might be a misleading result from the regression analysis. A high unemployment rate and a low GDP per capita might explain lower demand in certain regions inside Germany.

The vacancy rate is an indicator of the supply side of the real estate market and enters the regression formula with a negative sign. Therefore, regions with an excess supply of real estate properties due to vacant houses and apartments have lower prices. Excess supply means less demand than supply for real estate, so the prices are lower than in other regions with lower vacancy rates.

## 4.2 Predictions Model 1 – 4

As mentioned, 20% of the dataset was kept aside to test the regression results. The independent variables of the test dataset were then inserted into the regression formula to get the predicted offering price. The following scatter plots show how the predicted and test values differ. The x-axis shows the test data, and the y-axis the predicted values. The blue circles highlight each calculated predicted value and compare it to the actual value. The red line outlines the trend line of the predicted values. A perfect prediction model would show a diagonal of 45°, meaning that every predicted value is the same as the actual value from the test data set.

Figure 1 shows the predicted values from Model 1. The blue circles are widely spread, and the trendline has no 45° angle. This indicates that there is a difference between predicted and actual values. The concentration of blue circles in all parts of the graph indicates that false values occur in the high and low price ranges. However, the trendline highlights that the model underestimates the real estate prices.

Figure 1 - Predictions Model 1

Ein Bild, das Diagramm enthält.

Automatisch generierte BeschreibungEin Bild, das Diagramm enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

Figure 2 displays the predicational values derived from regression formula 2. The trendline has a 45° angle, and the predicted values are all located close to the trendline. This indicates that the predictions are very accurate. However, as mentioned before and will be discussed in more detail later, the interpretation of this model must be made cautiously as the results can be misleading due to the multicollinearity of the variables.

Figure 2 – Predictions Model 2

Ein Bild, das Diagramm enthält.

Automatisch generierte Beschreibung Ein Bild, das Diagramm enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

Figure 3 displays the prediction results from Model 3 and Figure 4 from Model 4. The trendline in Model 3 does not have a 45° angle indicating that the predictions are inaccurate. However, the house price predictions seem more accurate than for apartment prices. Figure 4 indicates a similar picture regarding the trendlines and the predictive accuracy.

Figure 3 – Predictions Model 3

Ein Bild, das Diagramm enthält.

Automatisch generierte Beschreibung Ein Bild, das Diagramm enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

Figure 4 - Predictions Model 4

Ein Bild, das Diagramm enthält.

Automatisch generierte Beschreibung Ein Bild, das Diagramm enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

## 4.3 Implications

### 4.3.1 Theoretical Implications

Overall, most findings of this thesis's empirical part align with previous research. Having evaluated four different approaches to explaining house prices, it can be concluded that they are very complex to derive and explain. Models 1, 3 and 4, which tried to explain individual house prices by their characteristics and macroeconomic sizes, proved that housing is a multidimensional heterogeneous good, as several factors are essential when determining the price of a real estate object ([Wittowsky et al., 2020, p. 46](#_ENREF_34)). The R2s of Models 1, 3 and 4 were between 0.45 and 0.64. Therefore, the independent variables could explain around half of the prices. However, this also implies that house prices depend on several additional factors than the ones in the models. To get higher explanatory power, including more variables in the models would be necessary. As the literature has suggested, identifying all influential factors is impossible for real estate objects as it is a highly complex good and dependent on market movements. Moreover, finding reliable data sources for influential factors on real estate prices is challenging.

This study supports the statement from the literature that a property’s characteristics are very significant determinants for house and apartment prices. Around 50% of the prices in Model 1 were explained by only using the properties’ characteristics. The literature evaluated the influence of building age, object condition, size measured in square meters as well as the number of rooms and presence of an elevator in an apartment house. This study uses the mentioned factors and adds the factors of land area, number of bedrooms, number of bathrooms, parking space and cellar. Therefore, it provides a more complex model which tries to explain housing prices when only the property’s characteristics are known. As mentioned by the literature, several different factors are needed to explain housing prices more accurately. This is proven by Model 3 which combines characteristics and macroeconomic sizes. The explanatory power of Model 3 increased in comparison to Model 1, suggesting that adding macroeconomic sizes leads to a more accurate prediction of housing prices.

Model 2 is used to analyze the monthly price development of real estate objects. The interest rate, unemployment rate, disposable income and GDP per capita are determinants of the demand side and investment in housing construction as well as the number of construction permissions are from the supply side. Besides the unemployment rate, all variables show the same relationships as mentioned in the literature. However, the nearly optimal R2s of 0.98 and 0.97 seem unrealistic, considering housing prices were described as multicomplex and complex to forecast in the literature. One reason for this can be a multicollinearity of the independent variables. This appears when two or more independent variables are correlated, leading to a misleading regression result ([Daoud, 2017, p. 1](#_ENREF_7)). Figure 1 shows the correlation matrix between the independent variables. A high correlation between the variables leads to a less reliable model and misleading results, although the R2 is very high. However, data availability restrictions made it impossible to collect other data as dependent and independent variables, reducing the correlation between the predictors.

Figure 5 – Correlation heat map Model 2

Ein Bild, das Diagramm enthält.

Automatisch generierte Beschreibung

*Source: own calculations*

Model 4 seeks for reasons for regional price differences in the German market. All macroeconomic variables apart from the income per person, which were analyzed in the literature and added to the model, showed the same results. Therefore, this study supports the literature findings for regional differences shaping the real estate demand in Germany. Moreover, the vacancy rate as a supply-side factor was added to the model, suggesting that a higher supply in regions leads to a lower price level. The main drivers of regional price differences identified in this analysis are population density, local unemployment rate, GDP per capita and the vacancy rate.

### 4.3.2 Practical Implications

Some implications of the results can be transferred into the practical work of real estate valuation. For example, real estate agents could use Model 1 to estimate a property’s value when the characteristics are known. When estimating a fair market value of a real estate property, they could derive a rough estimation using the regression formula 1. Moreover, Model 1 concluded that the object’s condition is a significant determinator for real estate prices which underlines that renovation investments can increase a property’s value no matter how old it is. This can be an essential insight for investors when debating investments into renovations. However, as mentioned before, the explanatory power of Model 1 is around 50% suggesting that price estimation using the coefficients would not be that accurate. Therefore, an actual application of the findings by real estate professionals will probably be on rare occasions. Only rough estimations can be derived using Model 1.

Models 3 and 4 have slightly higher R2s and therefore result in more accurate estimations, so real estate professionals could utilize one for price estimations. Model 3 is applicable if macroeconomic sizes and characteristics are known for the specific year. Model 4 can be utilized when macroeconomic factors for the property’s state are known. Therefore, it would result in a more accurate estimation than Model 1 because some location-specific variables are considered in this model. Again, it is questionable if an application in the practical work is reasonable because housing prices depend on more factors than the ones included in the regression models, and the regression formulas would only result in rough estimates.

When trying to forecast monthly real estate price developments, Model 2 could be used. If predictions of the independent macroeconomic factors are known, real estate professionals could derive a price development forecast using the regression formula 2. However, as discussed earlier, the model’s accuracy is misleading due to multicollinearity between the independent variables. Therefore, it is again questionable if applying Model 2 in practical work leads to accurate results.

As shown by the scatter plots in the previous section the Models 1, 3 and 4 have a wide spread of predictions, indicating a lack of accuracy for predictions. Model 2 is very accurate, but its limitations can make it unreliable in practical work. However, for rough estimations, the models might be applicable.

# 5. Conclusion

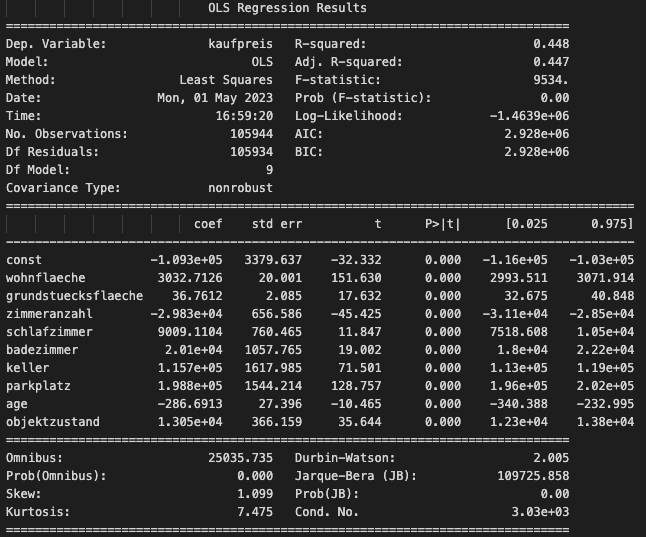
## 5.1 Main Findings and Answer to the Research Question

## 5.2 Limitations

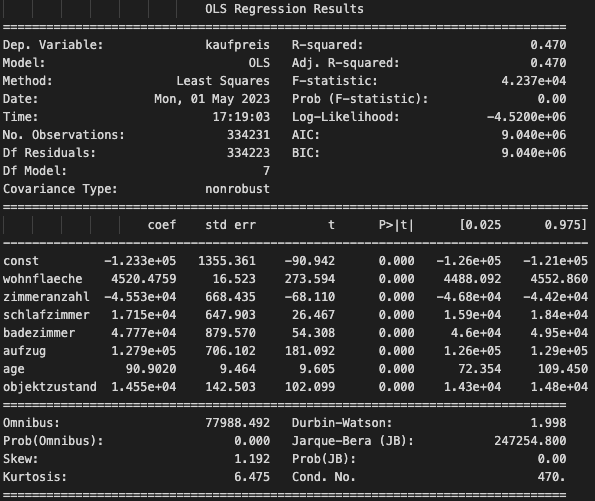
## 5.3 Future Outlook

# Appendix

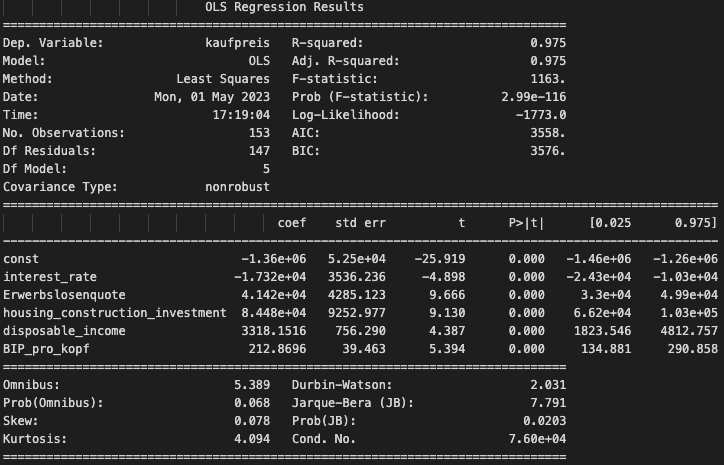
Appendix 1 – Regression results Model 1 houses for sale

 *Source: own calculations*

Appendix 2 – Regression results Model 1 apartments for sale

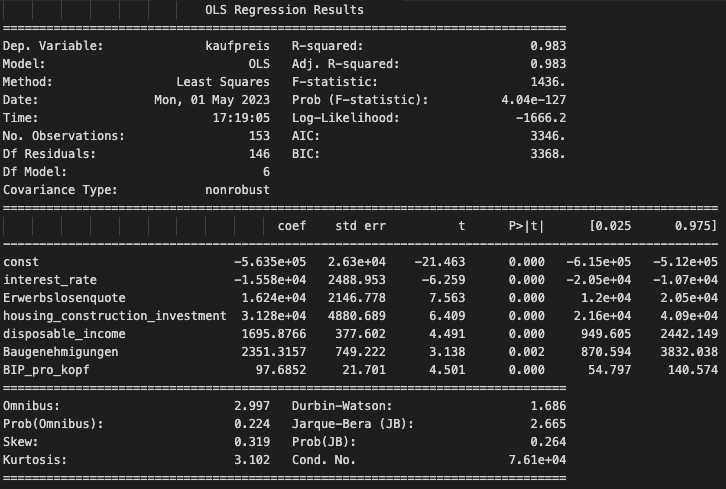
 *Source: own calculations*

Appendix 3 – Regression results Model 2 houses for sale



*Source: own calculations*

Appendix 4 – Regression results Model 2 apartments for sale



*Source: own calculations*

Appendix 5 – Regression results Model 3 houses for sale

Ein Bild, das Text, Platte enthält.

Automatisch generierte Beschreibung *Source: own calculations*

Appendix 6 – Regression results Model 3 apartments for sale

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung *Source: own calculations*

Appendix 7 - Regression results Model 4 houses for sale

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung *Source: own calculations*

Appendix 8 - Regression results Model 4 apartments for sale

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung *Source: own calculations*

# Bibliography

Belke, A., & Keil, J. (2018). Fundamental determinants of real estate prices: A panel study of German regions. *International Advances in Economic Research*, *24*, 25-45. <https://doi.org/10.1007/s11294-018-9671-2>

Bischoff, O. (2012). Explaining regional variation in equilibrium real estate prices and income. *Journal of Housing Economics*, *21*(1), 1-15. <https://doi.org/10.1016/j.jhe.2011.11.002>

Bundesagentur für Arbeit (January 2023). *Arbeitslosenquote nach Bundesländern im Jahr 2022*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/2192/umfrage/durchschnittliche-arbeitslosenquote-nach-bundeslaendern/>

Bundesbank. (February 2023). *Germany / National accounts / Households' income / Disposable income / Including the increase in claims on company pension funds, Table: BBNZ1.Q.DE.N.G.0325.A*. <https://www.bundesbank.de/en/statistics/macroeconomic-accounting-systems/-/households-income-651524>

Bundesbank Real Economic Indicators (March 2023). *Housing construction investment in Germany as a percentage of GDP*. <https://www.bundesbank.de/dynamic/action/en/statistics/time-series-databases/time-series-databases/759784/759784?listId=www_siws_iswi_realw2>

Cajias, M., Freudenreich, P., Freudenreich, A., & Schäfers, W. (2020). Liquidity and prices: a cluster analysis of the German residential real estate market. *Journal of Business Economics*, *90*(7), 1021-1056. <https://doi.org/10.1007/s11573-020-00990-2>

Daoud, J. I. (2017). Multicollinearity and regression analysis. Journal of Physics: Conference Series,

empirica. (December 2022). *Marktaktive Leerstandsquote auf dem deutschen Wohnungsmarkt im Jahr 2021 nach Bundesländern*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/258755/umfrage/marktaktive-leerstandsquote-von-wohnungen-nach-bundeslaendern/>

Ferlan, N., Bastic, M., & Psunder, I. (2017). Influential factors on the market value of residential properties. *Engineering Economics*, *28*(2), 135-144.

Flandorfer, P. (2023). *Durchführung und Interpretation der Regressionsanalyse*. Retrieved 24. April 2023 from <https://www.scribbr.de/statistik/regressionsanalyse/>

Funk, Z. A. K., & Drechsel, D. (2017). Wohnungspreise reagieren wieder stärker auf Zinssenkungen. *Die Volkswirtschaft*, *6*, 41-43.

Gürtler, M., & Rehan, C. (2008). *Preisbildende Faktoren von privaten Immobilien*.

Henger, R., & Voigtländer, M. (2015). *Vereint in regionalen Unterschieden-Der deutsche Wohnungsmarkt 25 Jahre nach der Wiedervereinigung: Ein IW policy paper in Kooperation mit der Schwäbisch Hall Stiftung bauen-leben-wohnen*.

Hiller, N., & Lerbs, O. W. (2016). Aging and urban house prices. *Regional Science and Urban Economics*, *60*, 276-291. <https://doi.org/10.1016/j.regsciurbeco.2016.07.010>

Irandoust, M. (2019). House prices and unemployment: an empirical analysis of causality. *International Journal of Housing Markets and Analysis*, *12*(1), 148-164. <https://doi.org/10.1108/IJHMA-03-2018-0021>

Jäger, M., & Voigtländer, M. (2006). *Immobilienfinanzierung: Hypothekenmärkte und ihre gesamtwirtschaftliche Bedeutung*. IW-Analysen.

Just, T. (2023). Aufschwung vorbei: Zinsen belasten Wohnungsbau schwer. *Wirtschaftsdienst*, *103*(1), 20-23. <https://doi.org/10.2478/wd-2023-0010>

Just, T., & Maennig, W. (2012). *Understanding German real estate markets*. Springer. <https://doi.org/10.1007/978-3-319-32031-1>

Kajuth, F., Knetsch, T., & Pinkwart, N. (2013). Assessing house prices in Germany: Evidence from an estimated stock-flow model using regional data. *Available at SSRN 2796934*.

Kröhnert, S. (2012). Wohnen im demografischen Wandel. *Berlin-Institut für Bevölkerung und Entwicklung. Berlin*.

Organization for Economic Co-operation and Development (April 2023). *Interest Rates: Long-Term Government Bond Yields: 10-Year: Main (Including Benchmark) for Germany [IRLTLT01DEM156N]*. <https://fred.stlouisfed.org/series/IRLTLT01DEM156N>

Quigley, J. M. (2002). Real estate prices and economic cycles. *International Real Estate Review*, *2*(1), 1-20.

Schaffner, S. (2020). *FDZ data description: Real-estate data for Germany Campus Files (RWI-GEO-RED city and RWI-GEO-RED cross)-Advertisments on the internet platform ImmobilienScout24 for teaching purposes*. <http://hdl.handle.net/10419/242995>

Soot, M., Weitkamp, A., Alkhatib, H., Dorndorf, A., & Jeschke, A. (2016). Analysis on different market data for real estate valuation–investigations on german real estate market. FIG Working Week,

Statistische Ämter des Bundes und der Länder (March 2023). *Bruttoinlandsprodukt (BIP) je Einwohner nach Bundesländern im Jahr 2022*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/73061/umfrage/bundeslaender-im-vergleich-bruttoinlandsprodukt/>

Statistisches Bundesamt (April 2023). *Anzahl der baugenehmigungen für Wohnungen in Deutschland von Januar 2021 bis Februar 2023*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/1340287/umfrage/baugenehmigungen-fuer-wohnungen-in-deutschland/>

Statistisches Bundesamt (August 2022). *Anzahl der Baugenehmigungen für Wohnungen in Deutschland in den Jahren 2003 bis 2021 (in 1.000)*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/73432/umfrage/baugenehmigungen-fuer-wohnungen-in-deutschland/>

Statistisches Bundesamt. (February 2023). *Bruttoinlandsprodukt (BIP) je Einwohner in Deutschland von 1991 bis 2022*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/1252/umfrage/entwicklung-des-bruttoinlandsprodukts-je-einwohner-seit-1991/>

Statistisches Bundesamt (March 2023). *Erwerbslosenstatistik nach dem ILO-Konzept Tabelle: 13231-0001*. <https://www-genesis.destatis.de/genesis/online?operation=previous&levelindex=3&levelid=1682438867569&levelid=1682438784906&step=2#abreadcrumb>

Statistisches Bundesamt (September 2022). *Bevölkerungsdichte in Deutschland nach Bundesländern zum 31. Dezember 2021 (Einwohner je km2)*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/1242/umfrage/bevoelkerungsdichte-in-deutschland-nach-bundeslaendern/>

Stepstone. (February 2023). *Durchschnittliche Bruttojahresgehälter von Vollzeitarbeitenden in Deutschland nach Bundesländern im Jahr 2022 (Median)*. <https://de.statista.com.login.bibproxy.whu.edu/statistik/daten/studie/603596/umfrage/bruttojahresgehaelter-in-deutschland-nach-bundeslaendern/>

Sutton, G. D. (2002). Explaining changes in house prices. *BIS quarterly review*, *32*(1), 46-60.

Voigtländer, M. (2014). The stability of the German housing market. *Journal of Housing and the Built Environment*, *29*, 583-594. <https://doi.org/10.1007/s10901-013-9366-1>

Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: impact of housing characteristics, accessibility and neighbouring apartments–a case study of Dortmund, Germany. *Urban, Planning and Transport Research*, *8*(1), 44-70. <https://doi.org/10.1080/21650020.2019.1704429>

# Affidavit

First Name: Philipp

Last Name: von Seydlitz-Kurzbach

I hereby declare that I have written this thesis on my own and with no other help than the literature and other supportive material listed in the appendix. Citations of sentences and parts of sentences are declared as such, while other imitations are clearly marked and linked to original sources with regard to the extent and intention of the statements made. This thesis has never been handed in to any examination authority before and it is also not yet published.

………………………………….. ……………………………………

Place, Date Signature

# AI Declaration