External Effects on Demand in Empirical Urban Economics

Universität Hamburg

Fakultät für Wirtschafts- und Sozialwissenschaften

(kumulative) Dissertation

Zur Erlangung der Würde des Doktors der Wirtschafts- und Sozialwissenschaften

(gemäß der Promotionsordnung vom 18. Januar 2017)

vorgelegt von

Waldemar Beimer

aus Alma-Ata

Hamburg, 2022

Vorsitzender: Prof. Dr. Thomas Straubhaar

Erstgutachter: Prof. Dr. Wolfgang Maennig

Zweitgutachter: Prof. Dr. Gabriel Ahlfeldt

The disputation was held on June 21, 2022

Alice: “Would you tell me, please, which way I ought to go from here?”

The Cheshire Cat: “That depends a good deal on where you want to get to.”

Alice: “I don't much care where.”

The Cheshire Cat: “Then it doesn't much matter which way you go.”

Alice: “...So long as I get somewhere.”

The Cheshire Cat: “Oh, you're sure to do that if only you walk long enough.”

Lewis Carroll, Alice in Wonderland

Acknowledgements

I am deeply indebted to Wolfgang Maennig, who is a great advisor and coauthor. He did outstanding work in supporting my academic success. I also thank my colleagues Felix Richter, Malte Steenbeck, Viktoria Schumann, Steffen Müller and Franziska Kruse. This dissertation would not have been possible without our monthly jour fixe and the priceless feedback thereof. Thankfulness also goes to Gesche Wegener, Edgar Baake, Silke Sturm and Giulia Thomsen for their support. Primarily Gesche Wegener provided invaluable help by discovering the StadtRAD dataset. I owe my friend Freya Diederich my gratitude for her help and encouragement throughout my Master and Ph.D. And, most important, I thank my parents, my sister, and my wife, Sandra.

CONTENTS

List of Tables

List of Figures

1. Introduction

Urban economics can be regarded as a particular field of spatial economics that mainly deals with questions regarding cities. Spatial economics deals with the fact that geographical distribution affects economic behavior. The idea behind spatial economics is very similar to the phenomenon in physics that the strength of the gravitational force depends on the mass and distance of the entities involved (Newton 1687). Even using the concept of gravity Isard (1954) has shown that that country size and distance between countries are excellent estimators to predict the trade flow between those countries. One of the first analyses in the area of spatial economics was done by Garrison (1959) who tried to answer the core question of “What determines the spatial arrangement (structure, pattern, or location) of economic activity?” (Garrison 1959, p. 232). However, this analysis was from a rather geographic instead of an economic perspective. In economics, this spatial connection was cleverly framed as: "everything is related to everything else, but near things are more related than distant things." (Tobler 1970, p. 236). Like gravitation, it seems reasonable that economic activity somehow depends on other economic activity. This dependency should be higher the smaller the distance between both activities is. Distance in this context can mean economic or geographical distance. Following the reasoning, economic activities connected and intertwined with each other should influence each other more. (LeSage 2008)

These kinds of spatial interconnections between entities are usually subsumed under spatial correlation in the economic literature. As the name suggests, the field of spatial economics tries to estimate and incorporate these effects between subjects. The field of spatial economics is grounded in findings of the combination of geography and statistics. However, more specialized fields of study at least try to incorporate these interactions in their models. This is because most economic relations seem to have a spatial component. These components range from the placement of real estate goods in a city to the travel patterns of the population. Most obviously, transportation costs of goods (Thisse 2010). One could argue that these effects become less relevant in the information age. Since a lot of economic activity is done in the form of information and not physical goods. However, even in the stock market, location matters. For example, high-frequency traders pay high markups to be as close as possible to the physical location of a stock exchange (Gomber and Haferkorn 2015). Most fields incorporate specific models and methods to control for these effects.

Empirical approaches to estimating spatial effects, particularly spatial correlation, had many improvements over time. Most research on this topic is summarized under the term spatial econometrics. Spatial econometrics applies spatial methods to real-world data to prove spatial correlation and determine the spatial effects' magnitude. Since spatial econometrics models as well as data and computational power were more limited in the early years, most models only tried to estimate the spatial dimension of the effects. Anselin (1988) gives a good impression about the state of research in this field before 1990. Including all major models and ideas up to this point. However, it also shows how limited the field was in this period caused by the lack of available data and computational power.

The following years led to more sophisticated models incorporating a temporal dimension. In addition, there was an increase in computational power and available data to test different hypotheses. Elhorst (2012) summarized these improvements in model structure as well as methods to tackle the problems at hand. They adequately show there is no Goldilocks solution to the spatial problem. Even so, there we many improvements regarding available answers to spatial correlation. Every researcher has to analyze the data and questions to find the best method for their problem. Anselin (2010) also made a good summary of the 30-year development in the field of spatial econometrics between before 2010. He describes three distinctive phases through his career in particular "preconditions, take off and maturity". While giving an outlook on what possible challenges in the future. Mainly he was right on the whole field now called big data analysis and how it warped some methods and assumptions. For example, the notion of a variable being significant in a single model. Since using enough data points nearly every variable tends to be significant. For a more real estate specific overview of applied spatial models at the time, the work of Wilhelmsson (2002) is advisable. The most recent overview of spatial economics in general, was done by Redding and Rossi-Hansberg (2017). In Particular they shed light on the macroeconomic side of spatial economics and how empirically analyzing the theoretical frameworks has improves our understanding of the spatial interactions on the production side. This thesis uses these newer spatial-temporal models to advance the literature.

This thesis combines all mentioned improvements made in recent years. Using data sources with large amounts of observations over prolonged periods of time and modern computational technologies, the analysis can be made for disaggregated domains in one urban area. This allows us for example to control for the inherent effects some areas have on real estate prices without taking into account each of these effects individually.

We are confident that our results especially concerning real estate economics, hold true in general. For example, it seems unlikely that transportation noise has no adverse effects in rural areas. However, cities are our main point of interest. That leads to somewhat narrowing down our research to the field of urban economics.

Urban economics deals with the areas where most people globally live. In 2018, urban areas inhibited 55% of the human population. These numbers are far higher for high-income countries (like our case studies). To give some indications, in 1950 the urbanization already reached 59% in these countries. In 2018, it was 81% and is estimated to reach 88% by 2050. Following these numbers, 4 out of 5 people in high-income countries can directly benefit from our research. It is important to mention that the UN does not use a consistent definition of urban area. Instead, each country defines for themselves what an urban area is and the UN uses that definitions (United Nations 2019). A report for the EU in 2016 defined urban areas identical for all nations by fixed principles dependent on population density and minimum population. It finds that globally urbanization was already at 85% in 2015 (Pesaresi Martino, Melchiorri Michele, Siragusa Alice, Kemper Thomas 2016).

Our research focusses on existing cities and the improvement of the aggregation of goods in these cities. A strong focus lies in the amenities and disamenities these cities provide. In particular, these are measured by the demand distributed around the city. Another area of interest of urban economies lies in the productivity site of cities and the corresponding wage structure. It is crucial to use data from different cities to estimate these effects in an urban economics context. Our focus is on the consumption site of cities. In particular, how amenities (as a way to spend leisure time for example, or reduce commuting through better transports) can increase the quality of living. The quality of a location can somewhat be attributed to the sum of its amenities. Measuring the quality of life differences between different points in a city is a primary goal of this thesis.

In our opinion, the main goal of Urban Economics should not be to just find and explain existing effects. It should help policymakers and other economic entities make the right decisions in an urban environment. Modern city development is somewhat “unnatural” since policymakers decide which areas can develop which kind of amenities and infrastructure. So at least the broader development of an urban environment is primarily policy-driven. This has also a direct influence on neighboring regions that are more rural.

These policies can have different goals depending on the reasons behind them. Most often, they either aim to increase the well-being of a specific demographic to assure votes, or try to maximize some form of welfare (Neumark and Simpson 2015; Small 2013). This welfare can be measured in economic terms or the sociodemographic composition (improvement). Often these policies contradict each other. To better understand the context at hand, it is essential to explore some possible policies and their implications. Since our case studies are in Berlin and Hamburg, we will use policies applied at these cities. Both cities are growing and have policies in place to increase the amount of available living space. Situationally both cities cap the market rent for living space, possibly decreasing the overall supply of living space. In addition, both cities try to protect specific areas from gentrification. Naturally, this leads to an inefficient overall distribution of living space measured in for example total rent (Olsen 1972; Diamond et al. 2019). However, it could protect the area's character in the long run. One could argue that without these policies, Hamburg and Berlin would have much higher rents in the city centers. These high rents would highly increase the incentives for property owners to increase the overall supply of living space. Lastly, especially Hamburg started some essential programs to increase the supply of public transportation and discourage the usage of private cars in the city center.

All policies can lead to unexpected and often even undesirable results (Oliveira et al. 2017). Spatial policies are no exception to this general rule (Henderson et al. 2013). They could be even more prone to this kind of side effect since more dimensions and interactions are involved. Some policies even fail their initial goal but lead to entirely different outcomes changing the status quo (Wells et al. 2009). For example Englander (2021) shows how a spatial policy intended to decrease the overall fishing of juvenile fish in Peru actually increased the fished quantity of this fish type. These effects can be triggered by not including external factors of the policy and the targeted good.

We define *external factors* as everything that is not an inherent property of the evaluated good but affects the demand. In contrast, the internal factors are the inherent properties of a good. External factors play the same role concerning the productivity side of economics (Anselin et al. 2004). For example, apparent external effects for real estate prices are all amenities a city offers and the distance between these amenities and the house in question. Some amenities can simultaneously produce positive and negative external factors. Airports in general have a positive effect on the economic development and demand for living space of adjacent areas (Cidell 2015). This can be measured easily through the correlation between house prices and distance to the nearest airport. However, airports or more specifically airplanes produce noise, decreasing the quality of living and thereby influence real estate prices in affected areas. Accordingly, while taking into account a single amenity, we need to check if this amenity produces more than one relevant external effect.

This thesis aims to give policymakers indicators of which dependencies between spatial entities exist. These dependencies can show how interventions in existing systems can change them. The focus is on the effects of amenities in the broadest sense on demand for goods with relatively fixed supply and somewhat fixed location. We consider the positive and negative internal and external effects of amenities. For example, we include the positive effects of public transport through an availability measure to the station as well as negative effects like noise produced by public transportation.

This thesis comprises three self-contained chapters on urban economics. Each chapter uses spatial econometric methods combined with real-world data to answer specific questions. This real-world data consists of cross-sectional and panel observations. Housing and transportation are two of the major challenges of most industrial nations especially, in urban environments (Addams 1902; Gurran and Phibbs 2015; Hartman and Robinson 2003; Cruz and Sarmento 2020). We focus on housing in chapters 2 and 3 and on transportation in chapter 4.

Chapter 2 of the dissertation contains an article titled “Noise effects and real estate prices: A simultaneous analysis of different noise sources”. Noise in general has a negative effect on the quality of living (Seidman and Standring 2010). Since it reduces the quality of living at specific locations, we should be able to measure a negative effect on the price of houses in that area. The three most common noise sources affecting the value of houses are road, train and air traffic. The effects of these single noise sources are well analyzed. However, since the study designs and areas are different, it is hard to compare the strength of the effects to each other. Incorporating the interactions between noise sources is only possible while using a single dataset. We use a dataset consisting of all real estate transactions of single family houses in berlin for a period of 20 years. This rich dataset enables us to control for time variables. Our results regarding single noise sources are similar to the literature. In addition, we can show that the effects of train and road noise do not significantly differ from each other. Air traffic noise has a higher effect per decibel (dB) than the other two. We also show that the effects are intertwined. Simply estimating the effects of single noise sources and adding up these effects would lead to biased results if a house is affected by more than one source.

Chapter 3 of this dissertation consists of an article called “On the Price Gap between Single Family Houses and Apartments”. The price differences between different forms of living arrangements are rarely analyzed in a broader setting. Real estate economics often wants to specify the effects on the price from the change in one variable like the number of rooms or lot size. Comparing different forms of living arrangements gives the possibility to analyze preference changes between amenities over time. This is impossible if only one form of living is observed since there would be no control group. Our work decomposes the price gap between single family houses and apartments in Berlin from 1990 until 2015. These are the primary forms of living in our dataset. For this paper, we extended the previous mentioned dataset twofold. The timeframe now includes the years until 2015 leading to a dataset reaching over 25 years. Additionally adding all transactions of apartments leads to a total of more than 300.000 observations to analyze. Using standard microeconomic determinants applied in most hedonic real estate models, we find rather exciting results regarding the drivers behind preference changes in the form of living. Most of the existing price gap between both forms of living arrangements can be explained by their average individual characteristics. The unexplained price gap somewhat follows a *u*-shaped over the time perdiod. Showing a higher preference for single family houses in the first half of the 1990s. This preference is negated until 2006. From 2007 ongoing this positive effect somewhat returns. In the analyses, we show precisely which preference changes drive those effects. The two most important factors are floor space as well as proximity to the central business district.

The final Chapter 4 is called “Impact of Social Structure on Bike-Sharing Demand: Evidence from StadtRAD users in Hamburg”. There was a massive rise in Bike-Sharing systems worldwide in the last decade. The reasons for this are mainly GPS and online billing systems. These systems produce massive amounts of data, which can be used to analyze the demand functions. Although Germany has many Bike-Sharing systems, there is no empirical study using this data for the German market. We analyze every trip made with the biggest Bike-Sharing system in 2015 and 2016, to show which variables effect bike sharing demand in Hamburg in those years. It covers the impact of similar amenities and socioeconomic effects like Chapters 2 and 3. It adds variables special to Bike-Sharing demand like the weather, time of day and lagged demand. We are able to confirm the general results found in the literature. Mainly that good weather increases the demand and that the demand on business days is higher compared to weekends. In addition, we introduce a method to calculate the actual bike availability at the station level. Our results expand the literature by revealing that the elasticities for most variables are much higher on weekends compared to weekdays. Finally, yet importantly, we are able to prove that the socioeconomic composition of the population has an impact on the Bike-Sharing demand.

Real estate markets help us to estimate the real value attributed to a location. These values aggregate all attributes contributing to the quality of living a location has. As mentioned before, these attributes include costs like noise or pollution as well as advantages like higher wages or cultural amenities like public parks. However, since house prices as a whole can differ strongly over time, it is seldom helpful to compare absolute values. At least in the context of urban economics. In chapters 2 and 3, our goal is to estimate the factors responsible for price differences between locations. In our case in particular between locations in the same city. By controlling for the structural factors of a house or apartment, we can estimate the values of the externalities. The primary forms of living in Germany are apartments and single-family houses. We do not expect to see significant shake-ups in the general form of living in the future and concentrate on analyzing these main supplies of living space.

Transportation on the other hand underwent some major shifts in the past, especially in cities. Cars, buses, city trains, metros and more recently car- and bike-sharing are some examples of transportation. For comparison Hamburg has a modal split of 22% public transport, 27% walking, 36% car and 15% bicycle. The modal split is measured per trip not per travel distance. 19% of the inhabitants have used a bike-sharing system at least once (BWI 2021). In the fourth chapter, we follow Hörcher and Tirachini (2021, p. 1) for classical public transport and define it “as high-capacity vehicle sharing with fixed routes and schedules…”. The bike-sharing system does not strictly fall into this category. However, since the stations are fixed, there are – to some extent - natural routes between stations, and the vehicles are shared.

Although the sharing economy is only a minor part of the transportation mode in cities, it is a growing market, with insufficient research to guide policy decisions. Bike-sharing can play an important role in increasing the well-being of the population. In addition, as it uses fewer resources compared to cars, it helps the environment. These effects are crucial in cities where congestion needs to be reduced and air quality increases by not producing CO2 or fine particles. We aim to make these areas more efficient and to find economic connections and incentives.

Like in most sciences, models are used to describe reality. Most models start as theoretical models. Economists demonstrate possible connections between economic entities through axioms and assumptions. Depending on the special case considered, the models are often expanded. Often Theoretical models get replaced by new models explaining the reality more closely. The empirical approach (leading to empirical models) is eager to verify the theoretical predictions. Therefore, instead of predicting a positive interaction between two entities, we can use real-world data to verify the assumed effects and estimate their specific strength. (Kennedy 2002; Lau 1986; Phillips 1996; Hansen 2005)

Empirical models most often use some form of regression. Regressions enable us to explain the variation found in the data. By explaining the variation, we can find unbiased estimators for our parameters of interest. If done right, these parameters of interest show the causal effect between one entity and another. In a best-case scenario, this effect shows the strength and direction of the interaction. Ensuring the unbiasedness of the estimation is of most importance (Ioannidis et al. 2017).

Results can be biased through various means. Some possibilities are spatial clustering, uncontrolled spatial heterogeneity, omitted variables and just measurement errors (Plümper and Neumayer 2010). Biased estimates can be somewhat helpful if the direction of the bias is known. Therefore, the estimates could give a lower or upper bound for the actual effects. An example of this is the Nickel downwards bias leading the estimates of the spatial lag to be smaller for small sample sizes (Nickell 1981).

There is some criticism specific to spatial econometric modeling. For example, Corrado and Fingleton (2012) argue that spatial econometrics focusses too much on data-analysis and using model results to choose the proper configuration. This criticism was addressed in the publications where applicable. Here we want to give a broader understanding of how this criticism influences the work done and the limitations of interpretation. In addition, a general criticism of spatial economics was made by Gibbons and Overman (2012). They argue that spatial econometrics often does not show causal effects but instead models best describing the data at hand. Partridge et al. (2012) directly compare classical urban economic approaches and the theoretical difficulties of using spatial econometric methods.

As stated before, many spatial models use some form of spatial lag. These spatial lag models control for the spatial correlation through these lag variables. In the most basic form this leads to an spatial autoregressive model (SAR(1)) model (Pineda-Ríos et al. 2019). Although we do not use this standard model in our analysis, the fundamental problems still apply. Using only possible spatial dependencies between the dependent variables and ignoring dependencies between lagged independent variables could lead to endogeneity. This kind of model assumes a linear dependency that we address by adjusting our weighting matrixes accordingly. Nonetheless, the whole approach of defining the complete spatial dependency in a single coefficient is far too simple to approach reality (Pinkse and Slade 2010). This is why we refrain in all papers from interpreting the spatial lag term beyond it being significant and direction. Even considering the limitations, spatial lag models are rather helpful in explaining the data and reducing omitted variable bias.

The results found in the respective models have to be interpreted with regard to the specifications of the models. In addition, we must check the results for robustness. We do this by using different model specifications as well as different parameter compositions. After being sufficiently confident of the robustness of the results, the best specification is chosen to discuss the causal effects of interest. In particular, the significance, direction and strength of the estimates. To finalize the validity of our results, they are compared with theoretical predictions as well as results found by other empirical studies in the literature.

Taking into account the spatial dimension has become state of the art in many economic research fields (Proost and Thisse 2019; Ye and Dang 2013). This obviously includes areas like urban economics, regional economics and transportation economics. It also includes the fields of labor and health economics by taking into account the locations of living for example. Looking at the broader branches of microeconomics and macroeconomics. Modern quantitative spatial models are more often used in the area of microeconomics. The reason behind this is the higher degree of spatial segregation and the sheer number of economic entities observed. However, Proost and Thisse (2019) show impressively how spatial economics does improve more macroeconomics specific areas as well by stating that: “It is our contention that spatial economics has reached a sufficiently mature level to trigger cross-fertilization.” (Proost and Thisse 2019, p. 635). In the more microeconomic influenced areas, especially urban economics, transportation economics, real estate economics, geographic economics and obviously spatial economics, the spatial dimension is of interest for our models and questions at hand. Although these fields often overlap strongly. In our context, quantitative spatial economics has become more of a toolbox for the specific research fields to find tools to help answer respective questions. These toolboxes have become rather extensive. It seems unreasonable to assume that some form of unification will occur since most datasets are distinct and need their own specifications. However, the impact of these tools in the last decades of research is undeniable. Through the increase of computational power and the broader availability of (big) data, including spatial coordinates not including spatial controls is basically lousy practice. Consequently, in the future, it is less about if research should control for spatial effects but instead how it should do so. This process is similar to the inclusion of time controls in business cycles. This thesis builds on the tools provided and shows methods for improving them. Reading it should help researchers make better decisions on which tools to use and how to tackle their own data.

2. Noise effects and real estate prices: A simultaneous analysis of different noise sources[1]

Abstract: We simultaneously analyze the effects of alternative noise sources in order to isolate the relative harms of alternative noise sources. This research adds to the literature, which has only analyzed one noise source or has aggregated the noise levels of different sources. Flight noise had the most negative effect on housing prices, and road and train noises had similar, but smaller effects.

Keywords: traffic noise, real estate prices, noise depreciation index, Berlin

2.1 Introduction

The effects of noise on housings prices have been analyzed since long (for an overview see Navrud (2002) and Nelson (2008)). However, to our knowledge, all publications concerning this topic concentrate on one noise source, such as road, train, or air traffic noises. No study simultaneously analyzes the effects of different noise sources. In addition, studies that include several noise sources either aggregate the noise levels (Baranzini and Ramirez 2005) or use dummy variables for noise levels, but they do not differentiate the noise sources (Theebe 2004). This study aimed to close this gap and isolate the potentially differing effects of alternative noise sources on housing prices.

2.2 Data

The study analyzes all single-family home transactions in Berlin, the capital of Germany, reported in the Kaufpreissammlung (statistics of real estate transactions provided by the Committee of Valuation Experts) between 1990 and 2012. The statistics also included characteristics, such as floor space, surface area, age of the building, amount of stories, (good and bad) condition, etc. (for a detailed description of the data see Ahlfeldt and Maennig (2015)). The statistical dataset included approximately 27,000 transactions.

Noise data were taken from the Department for Urban Development and the Environment. This data included information about the noise levels for every source of transportation noise on the basis of 10 x 10-m grid cells. The noise levels were quantified using a long-term sound pressure index (Lden). Although Lden has the same scale as the log-decibel scale (dB), Lden accepts that humans may perceive night noise c.p. to be more annoying than day noise. Lden

[1] Coauthored with Wolfgang Maennig (University of Hamburg). We thank the Berlin Committee of Valuation Experts and the Senate Department for Urban Development and the Environment for data provision. Published as Beimer and Maennig (2017) in Transportation Research Part D: Transport and Environment.

adds 5 dB for the time period between 6 pm and 10 pm and 10 dB for the time period between 10 pm and 6 am. Following the WHO recommendation, a Lden score over 55 is a serious annoyance (den Boer and Schroten 2007). Therefore, we set noise values below 55 Lden at zero and subtracted 55 from all other actual noise levels (Ahlfeldt and Maennig 2015).

The selection of variables of socioeconomic and locational characteristics refer to one of the most referred noise studies Sirmans et al. (2005), to Chang and Kim (2013) who analyze an area similar to Berlin, to Swoboda et al. (2015) who provide one of the most recent publications on noise effects, and to Ahlfeldt et al. (2017) who provide the most recent hedonic model on real estate prices in Berlin (Campbell and Cocco 2005).

The locational GIS control variables included the distance to: the Central Business District (CBD), the nearest green space, the nearest body of water, the nearest metro rail station, the nearest playground, the nearest main street, and the nearest landmark. Sociodemographic characteristics included: population density, population share over 55 years old, cars per capita, non-German individuals (for 15,937 blocks), and the unemployment rate (for 338 traffic cells). These sociodemographic characteristics were provided by the statistical office in Berlin. Data on income and income distribution were available for the Bezirk (district) area; as a proxy we use the percentage of the high income population (>3200€/month), the most relevant income class in the real estate market. All data were provided by the Statistisches Landesamt Berlin (Statistical Office Berlin). We used a set of dummy variables for the years (Hiller 2015). Due to the historic particularity of Berlin, we also added a dummy variable for properties located in the former East Berlin area.

2.3 Methods

The current study followed most studies that tested the noise effects (Sirmans et al. 2005) and used a semi-log hedonic model, which allows to directly interpret the coefficients of the noise variables in terms of the Noise Depreciation Index (NDI):

$$\ln(p) = \alpha_i * Year_i + \beta_j * NOISE_j + \gamma * SpatLag + \delta_k * X_k + \theta_l * Y_l \tag{2.1}$$

Where p is the price, $Year_i$ is the year dummies, $NOISE_j$ is the noise sources, X_k is the house's characteristics, and Y_l is the socioeconomic and locational characteristics. Following Can and Megbolugbe (1997), the blocks surrounding the property at a range of up to 3,000 meters were included. The spatial lag variable, SpatLag, was generated using weighted matrixes $(W_{i,j,t})$ that account for the inverse distance between the block centroids using the following equation:

$$W_{i,j,t} = \frac{\frac{1}{d_{i,j}}}{\sum_i \left(\frac{1}{d_{i,j}}\right)} \tag{2.2}$$

where $d_{i,j}$ is the value of the Euclidian distance between two weighted centroids in meters. The block centroids were defined by the observations within each block, so the distributional information within each block could be used. Using all observations would make a weighted matrix too big to compute. Our design avoids effects of future observations on present prices and prevents restrictions on the amount of observations that can be used in the study. The actual SpatLag variable was calculated by multiplying every weighting matrix by the logarithm of the mean prices of the blocks used for the matrix. To test for robustness, the model was re-estimated without the SpatLag variable; in addition a more general model was run, using a Box-Cox-Transformation (Mok et al. 1995).

2.4 **Results**

As a point of reference, Models A, B, and C in Table 4.1 depict the estimates if only one noise source is analyzed at a time, following the majority of publications on this topic. Model A focused on road noise exclusively and found an NDI of 0.611, which is in line with the results of Swoboda et al. (2015), whose study found an NDI range of 0.25 to 0.5 for road noise. Model B analyzed air traffic noise exclusively and found an NDI of 1.27, which is in line with results from Nelson (2008), whose meta-study reported an average of 0.92 for air traffic noise. Model C analyzed train noise exclusively and found an NDI of 0.648, which is in line with Andersson et al. (2015), whose study found an NDI of 0.681 for train noise.

The coefficients for the control variables generally confirmed the findings in the referential studies on housing price determinants. The current study found that distance to the next green space in models A to C had an insignificant effect on housing price, which may be due to the fact that the vast majority of single family houses in Berlin are located in green areas. The study found a significantly positive coefficient for the share of foreigners, which may be because internationals who can afford to live in areas of single family houses (SFH) may have different impacts on real estate prices than internationals who (are forced to) live in more dense areas of Berlin. The SpatLag coefficient proposed by Can and Megbolugbe (1997) is significant. It shows that the price of a house depends positively on the prices of houses in close proximity. We avoid an endogeneity problem by only taking into account transactions for the spatial lag parameter of an observation, which took place before the transaction.

Model D simultaneously analyzed all noise sources. The coefficients for all noise sources were (insignificantly) lower for every noise source compared to the estimates in Models A, B, and C. Air traffic noise had the greatest negative effect. Comparing AIC values model D has the best fit compared to model A, B and C. The coefficients in model D can be translated into a reduced willingness to pay of -3118,3€ (±402,7) per dB for flight noise, -1530,6€ (±130,4) per dB for road noise, and -1494,2€ (±288,4) per dB for train noise .

Several explanations are at hand: Noise has several direct and indirect health impacts on humans, ranging from annoyance to hypertension. An overview over the evidence based impact is provided by Passchier-Vermeer and Passchier (2000). Miedema and Vos (1998) show that at the same exposure level measured in dB the effect of air noise is most annoying followed by road noise and train noise supporting our results.[2] The different effects on health may be due to the unsteady nature of air traffic noise; the different effects on the willingness to pay may be connected to the fact that it is difficult for humans to block this type of noise (Nelson 2008).

Model E extends Model D by using interactions terms between the noise variables. The main possible critique regarding interaction terms is a possible multicollinearity between the terms. However, we checked for this and found no linear correlation higher than 0,5. There is no significant change neither in the coefficients of the noise variables nor in the order of the impacts. We conclude that for houses only effected by one noise source the results are unbiased (Brambor et al. 2006). However, three of the four interactives are significant, implying that model D will be biased in cases of houses affected by more than one noise source. The mostly positive coefficients indicate that noise may have a decreasing disutility; the reduction of the willingness to pay for a house affected by more than one noise source can not be estimated by simply adding the effects from the individual noise sources.

As a robustness check, model F does not include the SpatLag variable. Air traffic and rail noise had (insignificantly) larger negative impacts in this model, according to the AIC and adj R^2 this model has the worst fit of the data. Model G uses a Box-Cox model without the SpatLag variable, implying that a need to transform the left handside variable exclusively, with no assumptions needed for the transformation of the mean block prices.[3] The order of the relative noise impacts remained stable with this model, as well. Flight noise again had the greatest negative impact on housing prices (per Lden), followed by road noise and train noise, which had similar, but smaller impacts. The results of this study are thus robust to alternative

[2] Basner et al. (2011) focus on the effects of noise on sleep also support differences in the effect of the three noise sources, but find road noise to be the most annoying.

[3] Cassel and Mendelsohn (1985) show that the original intent of the Box-Cox transformation is to get rid of the need for interaction terms.

specifications with and without spatial lag variable inclusion or estimations with a semi-log model or a Box-Cox model.

Table 2.1 Results of Models A-G for the effects of noise on real estate prices

	Model A	Model B	Model C	Model D	Model E	Model F	Model G
SpatLag	0.195*** (0.0118)	0.196*** (0.0117)	0.194*** (0.0118)	0.191*** (0.0117)	0.190*** (0.0117)		
Road Noise	-0.00611*** (0.000500)			-0.00589*** (0.000502)	-0.00652*** (0.000526)	-0.00732*** (0.000531)	-0.0126 (.)
Air Traffic Noise			-0.0127*** (0.00154)	-0.0120*** (0.00155)	-0.0120*** (0.00162)	-0.0166*** (0.00173)	-0.0319 (.)
Train Noise		-0.00648*** (0.00111)		-0.00575*** (0.00111)	-0.00594*** (0.00130)	-0.00701*** (0.00136)	-0.0144 (.)
Interact all Noise					0.000589** (0.000183)	0.000839*** (0.000197)	
Interact Train & Road Noise					0.000747*** (0.000207)	0.000758*** (0.000221)	
Interact Train & Flight Noise					-0.00423*** (0.00112)	-0.00675*** (0.00119)	
Interact Flight & Road Noise					0.000513 (0.000504)	0.000635 (0.000533)	
Floor Space	0.00306*** (0.000298)	0.00306*** (0.000297)	0.00306*** (0.000297)	0.00306*** (0.000297)	0.00306*** (0.000297)	0.00343*** (0.000320)	0.00657 (.)
Storeys	0.0754*** (0.0103)	0.0776*** (0.0103)	0.0774*** (0.0103)	0.0776*** (0.0103)	0.0769*** (0.0103)	0.0777*** (0.0113)	0.146 (.)
Attic Apartment	0.0555*** (0.00845)	0.0561*** (0.00846)	0.0563*** (0.00845)	0.0565*** (0.00845)	0.0563*** (0.00845)	0.0523*** (0.00909)	0.0957 (.)
Building Age	-0.00540*** (0.000508)	-0.00521*** (0.000515)	-0.00514*** (0.000513)	-0.00532*** (0.000508)	-0.00532*** (0.000507)	-0.00632*** (0.000543)	-0.0116 (.)
(Building Age)^2	0.0000261*** (0.00000554)	0.0000245*** (0.00000562)	0.0000239*** (0.00000559)	0.0000254*** (0.00000553)	0.0000256*** (0.00000553)	0.0000316*** (0.00000597)	0.0000581 (.)
Good Condition	0.169*** (0.00710)	0.172*** (0.00712)	0.172*** (0.00712)	0.171*** (0.00710)	0.171*** (0.00711)	0.178*** (0.00770)	0.340 (.)
Bad Condition	-0.277*** (0.0106)	-0.276*** (0.0107)	-0.276*** (0.0107)	-0.274*** (0.0106)	-0.274*** (0.0106)	-0.274*** (0.0109)	-0.505 (.)
Basement	0.0715*** (0.00644)	0.0738*** (0.00644)	0.0713*** (0.00649)	0.0681*** (0.00646)	0.0673*** (0.00646)	0.0696*** (0.00662)	0.131 (.)
East Berlin	-0.120*** (0.00682)	-0.122*** (0.00680)	-0.130*** (0.00689)	-0.126*** (0.00688)	-0.128*** (0.00697)	-0.184*** (0.00818)	-0.337 (.)
Floor Space Index	-0.608*** (0.0264)	-0.618*** (0.0266)	-0.615*** (0.0264)	-0.613*** (0.0266)	-0.614*** (0.0266)	-0.677*** (0.0290)	-1.277 (.)

Table 2.1 Results of Models A-G for the effects of noise on real estate prices

Dist. to CBD	-0.0171*** (0.000889)	-0.0165*** (0.000889)	-0.0170*** (0.000890)	-0.0174*** (0.000889)	-0.0174*** (0.000889)	-0.0214*** (0.00102)	-0.0412 (.)
Dist. to Greenspace	-0.000674 (0.00177)	-0.000833 (0.00179)	-0.00282 (0.00181)	-0.00382* (0.00182)	-0.00349 (0.00182)	-0.00717*** (0.00197)	-0.0146 (.)
Dist. to Street	0.0296** (0.0101)	0.0709*** (0.00960)	0.0803*** (0.00977)	0.0446*** (0.0102)	0.0497*** (0.0102)	0.0481*** (0.0105)	0.0809 (.)
Dist. to Monument	-0.0731*** (0.00733)	-0.0782*** (0.00734)	-0.0826*** (0.00739)	-0.0779*** (0.00735)	-0.0806*** (0.00736)	-0.0914*** (0.00757)	-0.168 (.)
Dist. to Playground	0.0253*** (0.00704)	0.0239*** (0.00706)	0.0233*** (0.00705)	0.0261*** (0.00702)	0.0237*** (0.00705)	0.0298*** (0.00739)	0.0650 (.)
Dist. to Station	-0.0243*** (0.00240)	-0.0257*** (0.00240)	-0.0234*** (0.00241)	-0.0242*** (0.00240)	-0.0227*** (0.00244)	-0.0243*** (0.00249)	-0.0501 (.)
Unemployed Share	-0.0203*** (0.000724)	-0.0202*** (0.000727)	-0.0189*** (0.000744)	-0.0189*** (0.000744)	-0.0190*** (0.000745)	-0.0231*** (0.000869)	-0.0440 (.)
Income Share >3200€	1.240*** (0.0640)	1.246*** (0.0642)	1.195*** (0.0646)	1.180*** (0.0645)	1.181*** (0.0645)	1.555*** (0.0682)	2.935 (.)
Population Density	-5.250*** -1.023	-4.955*** -1.031	-4.873*** -1.027	-5.622*** -1.027	-5.680*** -1.027	-6.336*** (1.125)	-12.07 (.)
Foreigners Share	0.442*** (0.0482)	0.417*** (0.0480)	0.445*** (0.0484)	0.463*** (0.0485)	0.467*** (0.0484)	0.674*** (0.0588)	1.312 (.)
Population Share >55years	0.0856*** (0.0169)	0.0833*** (0.0170)	0.0976*** (0.0171)	0.0852*** (0.0171)	0.0825*** (0.0172)	0.0966*** (0.0176)	0.205 (.)
Cars Per Capita	-0.0413** (0.0146)	-0.0411** (0.0145)	-0.0467** (0.0146)	-0.0408** (0.0146)	-0.0439** (0.0148)	-0.0277 (0.0143)	-0.0457 (.)
Constant	10.07*** (0.130)	10.03*** (0.130)	10.05*** (0.130)	10.12*** (0.129)	10.14*** (0.129)	12.51*** (0.0483)	17.10 (.)
Theta Constant							0.0513** (0.00555)
Sigma Constant							0.665 (.)
Observations	27756	27754	27756	27754	27754	27754	27754
Adjusted R^2	0.609	0.608	0.608	0.610	0.611	0.586	
AIC	19074.7	19177.6	19147.9	18988.7	18958.5	20710.4	703787.5

t statistics in parentheses
* p < 0.05, ** p < 0.01, *** p < 0.001

2.5 Summary

We simultaneously analyze the effects of alternative noise sources on the prices of single family homes in Berlin, Germany. We analyse more than 25.000 observations, controlling for price determinants which have been found to be significant in earlier studies on house values, including socio economic data and spatial amenities, GIS denominated data. We find that the effect of flight noise, compared to road or train noise has a significantly higher negative impact per dB. The effect of an increase of one dB of flight noise decreases the price of a house for around 3,000€. The price effect of road or train noise is at some 1,500€ per dB.

2.6 Acknowledgments

We thank the Berlin Committee of Valuation Experts and the Senate Department for Urban Development and the Environment for providing the data. In addition we like to thank two anonymous Referees for their inspiring and helpful comments. This research did not receive any specific grant funding from any public, commercial, or not-for-profit sector agencies.

3. On the Price Gap between Single Family Houses and Apartments[4]

Abstract: We provide a first decomposition of the price gap between prices for single family houses and apartments into their microeconomic determinants. Using individual data from the collection of real estate transactions data for Berlin, Germany, we find a positive price gap for houses in Berlin in 1990 which is decreasing up to 1998. When controlling for structural characteristics, no positive price gap can be identified between 1996 and 2006. The application of Oaxaca decomposition reveals a growing preference for size (favoring houses), which is – since 2006 - offset by a growing preference for a more central location.

Keywords: Apartments, Single Family Houses, preference changes, Berlin, Blinder-Oaxaca decomposition

3.1 Introduction

Relative price developments of apartments and single family houses may be due to macroeconomic as well as meso/microeconomic reasons. For simplicity single family houses will be called houses from now on. In a nation-wide macroeconomic analysis of the US market, Chinloy et al. (2014) argue that apartments and houses are substitutes in demand (consumption) and in supply (production). They find that construction volumes of houses and apartments exhibit different fluctuations, especially due to the variation of interest rates, with the construction of houses being more interest-rate elastic. In the US, houses are mostly owner occupied, but there is an equity market for apartments. As a result, the relative prices of apartments are increasing in times of decreasing interest rates.

On a microlevel, changing preferences in forms of living may be relevant (e.g., a growing preference for living in centrally located apartments rather than in houses). Chen and Harding (2016) find a growing preference for living space between 1985 and 2011 for the area of Chicago. Such changing preferences may be due to changes in lifestyles, such as those related to “urbanization” (Bögel et al. 2014), but may also be due to changes in demographics (Bogin et al. 2018). The growth of single-person households may affect the structure of

[4] Coauthored with Wolfgang Maennig (University of Hamburg). We thank conference participants of the 2016 and 2017 Homes-uP conferences and the 2018 ERES conference. Published as Beimer and Maennig (2020) in the Journal of Housing Economics.

demand for housing (Dieleman and Schouw 1989). In many western cities, more than 50 percent of dwellings are occupied by single households (Eurostat 2015).

We contribute to the analysis of the price gap by using the collection of real estate transactions ("Kaufpreissammlung") for Berlin, Germany, as well as by estimating the relative price developments of houses and apartments by controlling for the usual determinants of real estate prices. The analysis of the differences of submarkets in the housing sector and the analysis of the valuation of buildings of different types over time are well-established fields of real estate economics (Watkins 1999; Keskin and Watkins 2016; Dermisi and McDonald 2010). However, to our knowledge, there has been no decomposition of price differences into their microeconomic determinants, which could disclose potential dynamics in preferences. We add to a study by Chinloy et al. (2014) by analyzing individual transaction data and by controlling for spatial determinants as well as object characteristics rather than aggregated macro data.

We also offer a regional diversion to the literature on price gaps by analyzing the real estate market in Berlin, Germany. Our research relates to the literature on the urban economics of Berlin following reunification occurring in 1990 that uses a well-established set of data, determinants and model structures (Ahlfeldt and Maennig 2010a, 2010b, 2010c, 2013; Beimer and Maennig 2017), evaluates the impact of policy interventions (Alberts 2009 ; Ahlfeldt et al. 2017), and controls for the impact of local interest groups (Molnar 2010; Ahlfeldt 2011; Marquardt et al. 2013; Ahlfeldt and Maennig 2015) and for external shocks such as tourism (Schafer and Hirsch 2017). Our study also relates to Thomschke (2015), who decomposes the rent price boom, applies spatial explanatory variables such as amenities and socioeconomic structures, and differentiates between the specific impacts of each variable. Following Sah et al. (2016) we make sure to use robust hedonic controls. Comparing different spatial controls assures robust results.

As a second step, we provide a better understanding of the dynamics and most relevant driving forces of the price gap between houses and apartments by using Oaxaca-Binder decomposition (OD). OD was developed to decompose the gender wage gap by Blinder (1973) and Oaxaca (1973) and has led to the identification of a gap of approximately 30% that can be attributed in large part to different characteristics of the two gender groups. Since its first use, OD has been used to analyze the income gap between black and white populations (Melly 2005; Moral-Arce et al. 2012; Ospino et al. 2010), variations in the FDI indices of countries (Wei 2005), smoking behavior (Bauer et al. 2007), infant mortality rates (Demombynes and Trommlerová 2016), birthweights (Lhila and Long 2012) and hunting lease rates (Munn and Hussain 2010). In housing economics and real estate economics, Coulson and Dalton (2010), Newman et al. (2018) and Zorlu et al. (2014) use OD in order to identify potential effects of ethnicity on household formation and home ownership. LeSage and Charles (2008) use OD to explain the gap between newer and older houses, but do not consider amenities or socioeconomics variables. Another difference from our study relates to the time horizon used; their study is cross-sectional and does not analyze potential dynamics of preferences. Most recently Zhang and Yi (2018) using OD have shown for Beijing that price increases of housing units between 2012 and 2015 can mostly be attributed to changes in preferences in contrast to changes in the housing characteristics. Mathä et al. (2017) use OD and find that homeownership and house price dynamics have a strong effect on the net wealth of a person.

3.2 Data, Empirical Strategy, and Results

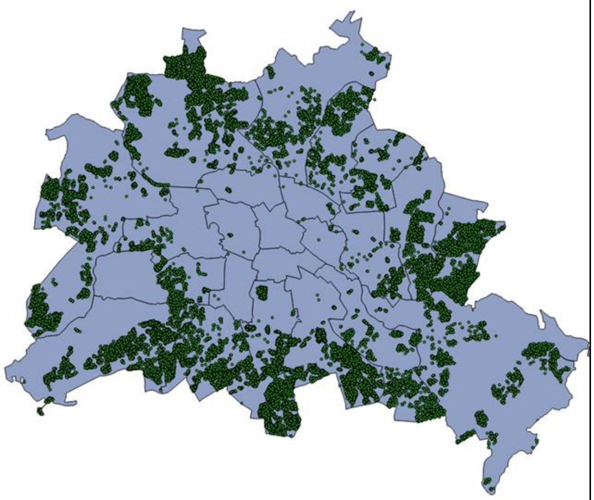
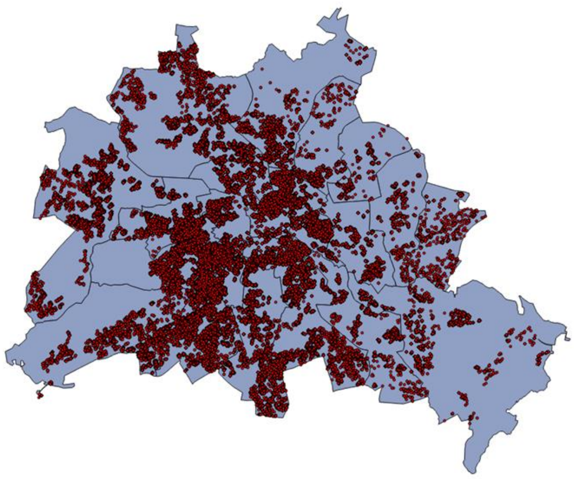
We use the collection of real estate transactions ("Kaufpreissammlung") provided by the "Gutachterausschuss für Grundstückswerte in Berlin" including all transactions made in Berlin between 1990 and 2015. The data collected includes transaction dates and a full set of the most relevant data on housing characteristics, such as spatial positions, building ages, floor areas, plot sizes, and the number of stories. We analyzed 40,868 transactions for houses. The number of transactions per year started at 814 in 1990 and rose to approximately 2,000 between 2008 and 2013; for 2014 and 2015, the number was approximately 1,650. Neither age nor floorspace followed a trend over time.

In total, 282,754 transactions were observed for apartments. The number of transactions per year started at 4,441 in 1990, jumped to a peak of 21,186 in 2009, and fluctuated between 14,040 and 17,680 from 2010 to 2016. The average floorspace increased from 69 m² on average in the 1990s to 74 m² from 2000 onward. The average age of the apartments traded also increased over time from 48 years in 1990 to 71 years in 2015, with some fluctuations. Following Schafer and Hirsch (2017), we merged this dataset with socioeconomic data available for the living environment areas (LEA, in German: "lebensweltlich orientierte Räume", LOR) level (Senatsverwaltung für Stadtentwicklung und Umwelt, Monitoring Soziale Stadtentwicklung).

We base our data selection on an international meta-study conducted by Sirmans et al. (2005), who determined the variables most often used in real estate analyses, and on the most recent empirical analysis of real estate in Berlin by Ahlfeldt et al. (2017). We consider age and age squared. The living area is included as floorspace measured in m². The variable "number of rooms" is not available in our dataset, but floorspace should serve as a good proxy. Figure 3.1 illustrates that apartments are located almost everywhere in Berlin, but houses are less often located in the more central districts. There is strong evidence that the housing stock of East Berlin had a different composition than that of West Berlin at the start of our study period and underwent more redevelopment across our observed timeframe (Molnar 2010; Ahlfeldt et al.

2017). We thus included a time-flexible dummy = 1 for real estate for East Berlin. We include distances to the closest central business district (CBD) area defined by Ahlfeldt (2011) and a control for amenities by measuring Euclidean distances (Senatsverwaltung für Stadtentwicklung und Wohnen) between objects and nearest amenities such as greenspaces, landmarks, schools, subway stations, and city train stations. Since the mixture of positive and negative externalities of public transport stations may lead to a nonlinear effect of the distance to stations (Mohammad et al. 2013), we include the distance to stations as well as the squared distance.

Figure 3.1 Spatial Distribution of Houses and Apartments in Berlin

Houses	Apartments
	

The maps show the location of every house and apartment transaction occurring in Berlin from 1990-2015. Data source: Gutachterausschuss für Grundstückswerte in Berlin, collection of real estate transactions (“Kaufpreissammlung”)

Detailed information on sociodemographic characteristics, including unemployment rates, shares of single parents, elderly poverty percentages, minorities and minority unemployment, and the percentage of people living in the area for more than 5 years were available from the statistical office in Berlin. The information is aggregated at 447 living world-orientated areas. We use the average yearly income per capita in € for each zip code from the “Gesellschaft für Konsumforschung.” All data refer to 2014.

Table 3.1 displays descriptive statistics for the variables. Over the full length of the panel, houses are on average 20 years younger than apartments; they are 74 m² larger in floorspace, and they cost approximately 143k € more.

Table 3.1 Descriptive Statistics for Houses

Variable	Unit	Mean	SD	Min	Max
Price	1,000€	276.133	213.544	15.000	5112.919
Floorspace	m²	147.596	65.508	19.000	3284.000
Age	years	43.795	29.356	0.000	238.000
Distance to closest:					
School	km	0.517	0.339	0.011	3.727
Station	km	1.408	1.180	0.029	9.310
Greenspace	km	1.866	1.464	0.006	6.368
Landmark	km	0.362	0.334	0.000	2.960
CBD	km	12.604	3.453	2.348	26.063
unemployment	percent	5.857	5.192	0.810	45.270
Single parent	percent	24.879	6.714	9.910	54.040
Elderly poverty	percent	1.894	1.696	0.000	17.080
Minorities	percent	7.782	4.991	0.890	40.990
unemployed Minorities	percent	17.285	12.407	0.000	69.110
Public housing	percent	6.074	11.009	0.000	93.830
Tenure >5 years	percent	68.718	5.618	28.090	82.180
Income per capita	1,000€	23.359	3.202	14.616	29.005

40,730 Observations for each variable. In the extended $SpatLag$ model, the number of observations is reduced to 39,916. However, the descriptive statistics do not change significantly. Statistics are available from the authors upon request. All socioeconomic variables were available at the LEA level.

Table 3.2 Descriptive Statistics for Apartments

Variable	Unit	Mean	SD	Min	Max
Price	1,000€	133.486	99.885	11.000	945.750
Floorspace	m²	73.278	29.543	28.100	194.840
Age	years	64.061	37.764	0.000	236.000
Distance to closest:					
School	km	0.300	0.195	0.003	3.199
Station	km	0.666	0.632	0.006	7.072
Greenspace	km	2.282	1.116	0.016	6.362
Landmark	km	0.149	0.174	0.001	2.337
CBD	km	7.008	3.763	0.176	25.365
unemployment	percent	11.137	6.845	0.810	45.270
Single parent	percent	30.551	5.823	9.000	54.040
Elderly poverty	percent	6.410	4.649	0.000	30.340
Minorities	percent	17.911	8.758	0.890	72.780
unemployed Minorities	percent	21.678	12.672	0.000	69.110
Public housing	percent	9.647	13.127	0.000	94.130
Tenure >5 years	percent	59.604	7.518	21.200	82.180
Income per capita	1,000€	19.393	2.884	13.572	29.005

Notes: 281,698 Observations for each variable. In the extended $SpatLag$ model, the number of observations is reduced to 276,304. However, the descriptive statistics do not change significantly. The statistics are available from the authors upon request. All socioeconomic variables were available at the LEA level.

In a first set of estimations we use hedonic semilog models including the explanatory variables discussed thus far, which are standard in the literature on real estate pricing (Ahlfeldt and Maennig 2015; Anderson and West 2006), but we also add an interaction term between the house dummy and the period dummies to capture the change in the gap over time (Fik et al. 2003; Costanigro et al. 2007). We also add period dummies for real estate transactions made in East Berlin and a time-varying interaction for CBD distance, which should control for effects of changing preferences to live close to the city center. Potential macro effects that may have an impact on the price level (i.e., tax rate, interest rate, and migration) are controlled for by yearly dummies. We use 1990 as a baseline, and all interaction terms show the total effect for the respective year following the interpretation of Balli and Sørensen (2013). Model 1.1 controls for spatial effects by including distances to amenities as follows:

$$log\, P = \beta_0 + \gamma_{jt} T_{jt} * Z + \alpha_t * T_t + \beta_k * X_k + \varepsilon \tag{3.1}$$

where vector X_k includes the explanatory variables described above; β_0 is a constant; and T_t is a vector for period dummies starting at t=1991. Z is a vector including the house dummy (HOUSE = 1), the East dummy (East = 1), and the control for centrality measured as the distance in km to a CBD. The error term is given by ε and is assumed to be iid. The set of variables we estimate is represented by $\gamma_{jt}, \alpha_t, \beta_k$.

Results are displayed in Table 3.3, column 1. As was expected, floorspace has a highly significant positive impact on the price. The coefficients for age and age² are highly significant with the expected signs. The positive impact of increasing distance to greenspace is counterintuitive. The "wrong" sign found may be due to an omitted variable bias since this model does not control for spatial fixed effects: apartments and houses surrounded by green space are on average positioned far away from the city center, and the distance to a CBD variable may not be able to fully compensate for this effect. For the socioeconomic variables, most impacts are as expected.

Model 1.2 controls for spatial dependencies by including fixed effects on the postal code level μ_n (Jun 2016) where observations from 186 postal code areas are included. The income variable drops for this model and all other models that include fixed effects since income is measured at the postal code level:

$$log\, P = \beta_0 + \gamma_{jt} T_{jt} * Z + \alpha_t * T_t + \beta_k * X_k + \lambda_n * \mu_n + \varepsilon \tag{3.2}$$

Here, and later on in equation 1.3, λ_n represents our estimates for each zip code fix effect. Our results (Table 3.3, column 2) confirm the findings of Herath and Maier (2013), who found such a specification to be superior relative to the simple OLS model. The results take the usual signs and values for variables such as floorspace and age structures. Some differences from model 1.1 are observed, i.e., distances to green spaces now have the expected negative impact, and the percentage of public housing in the area no longer has a significant impact.

Table 3.3 Regression Results of Hedonic Semi-Log Models

Variable/ Model	Model 1.1	Model 1.2	Model 1.3	Model 1.4	Model 1.5
Model explanation	Basic Hedonic Model using interaction terms for the yearly HOUSE gap	Model 1 + Spatial Fixed Effects	Model 2 + a Spatial Lag Parameter	Model 2 + Interaction Terms between all explanatory variables and the HOUSE dummy	Model 4 but only taking into account ZIP codes that include Apartments and Houses
Floorspace	0.010***	0.010***	0.010***	0.015***	0.015***
Age	-0.015***	-0.015***	-0.011***	-0.013***	-0.013***
Age^2	0.000***	0.000***	0.000***	0.000***	0.000***
Distance to closest:					
School	0.040***	0.031***	0.010*	0.044***	0.017***
Station	0.017***	0.023***	0.006	0.007	-0.014**
Station sqr	-0.004***	-0.004***	-0.000	0.003**	0.004***
Greenspace	0.013***	-0.031***	-0.022***	-0.024***	-0.007***
Landmark	-0.124***	-0.067***	-0.029***	-0.105***	-0.073***
unemployment	-0.013***	-0.000	0.002*	-0.000	0.000
Single parent	0.002***	-0.002***	0.000	-0.002***	-0.005***
Elderly poverty	0.001*	-0.001*	-0.002**	-0.001	-0.015***
Minorities	-0.001**	-0.004***	-0.002***	-0.007***	-0.010***
unemployed Minorities	-0.004***	-0.004***	-0.003***	-0.003***	-0.001*
Public housing	0.001***	-0.000	-0.000**	0.000	0.001***
Tenure >5 years	-0.006***	-0.003***	-0.001**	-0.005***	-0.007***
Income per capita	0.018***	0.000	0.000	0.000	0.000
SpatLag			0.295***		
Const.	11.135***	11.456***	11.113***	11.225***	11.366***
Yrs Dummy	YES	YES	YES	YES	YES

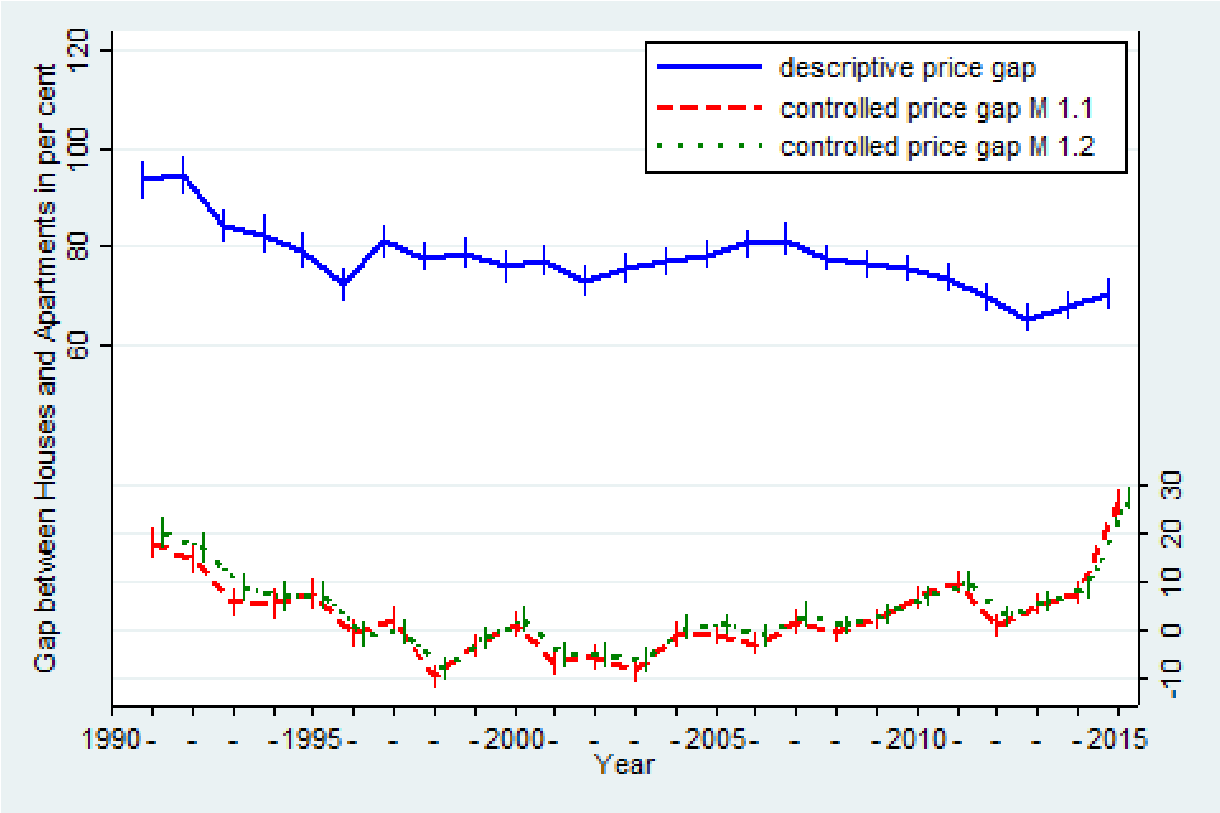
Table 3.3 Regression Results of Hedonic Semi-Log Models

Yrs *HOUSE Dummy	YES	YES	YES	YES	YES
Y*East Dummy	YES	YES	YES	YES	YES
Y*CBD Dummy	YES	YES	YES	YES	YES
ZIP Code FE	NO	YES	YES	YES	YES
Using Subgroup	NO	NO	YES	NO	YES
Interactions between HOUSE and explanatory variables	NO	NO	NO	YES	YES
N	322428	322428	295333	322428	161195
Adj. R^2	0.647	0.666	0.694	0.743	0.765
AIC	3.56e+05	3.38e+05	2.88e+05	2.54e+05	1.16e+05

Notes: *p< 0.1,**p< 0.05,***p< 0.01. We use a semilog model with Log(Price) as the dependent variable. All variables have the expected sign with respect to the meta-study by (Sirmans et al. 2005). The “Yrs Dummy*HOUSE Dummy” results of Models 1.1 and 1.2 are displayed in Figure 2.

Figure 3.2 illustrates the development of yearly house price gaps. It compares interaction terms for houses and year dummies for models 1.1 and 1.2 (dashed and dotted line), which are equivalent to the price gap while taking into account respective controls. As a reference, it also shows the descriptive gap (solid line) which is equivalent to the difference in mean log prices between houses and apartments. The graph illustrates a decreasing descriptive gap and shows that much of the descriptive gap disappears or even becomes negative when controlling for the variables included in models 1.1. and 1.2. Changes in the gap occurring between 1998 and 2014 are minor. At the end of this process, there are no more systematic significant differences in preferences for houses and apartments of similar sizes and locations.

Figure 3.2 House Price Gaps Over Time Estimated using Hedonic Models 1.1 and 1.2



The descriptive gap shows the relative average price difference between houses and apartmens for each year. The controlled gaps display interaction terms between houses and year dummies of Models 1.1. and 1.2. The controlled price gap decreases to 1998 and fluctuates at approximately 0 most of the time.

As part of an alternative strategy to control for spatial dependency, we estimate a spatial auto regression (SAR) model (model 1.3). For the estimation of the spatial lag (τ), we use a weighting matrix that takes into account the temporal dimension (Thanos et al. 2016) using exclusively earlier transactions. We also run SAR model 1.3 by considering spatial fixed effects to control for omitted variables (Gibbons and Overman 2012) as follows:

$$log\ P = \beta_0 + \gamma_{jt}T_{jt} * Z + \alpha_t * T_t + \beta_k * X_k + \lambda_n * \mu_n + \tau * SpatLag + \varepsilon \tag{3.3}$$

$SpatLag$ shows a positive sign as expected. Most coefficients do not change their signs in comparison to model 1.2. However, some are insignificant; such variables may have a similar effect on neighboring houses, implying that their effect may be absorbed by τ.

One may question how much the yearly gaps observed depend on explanatory variables for houses and apartments. For example, an extra m² of floorspace may have different effects on houses and apartments. Model 1.4 thus expands model 1.2 by including interaction terms between the houses dummy and every explanatory variable to control for potential time-dependent impacts of explanatory variables. The estimation equation is as follows:

$$log\, P = \beta_0 + \gamma_{jt} T_{jt} * Z + \alpha_t * T_t + \beta_k * X_k + \omega_k * X_k * \mathrm{HOUSE} + \lambda * \mu_n + \varepsilon \quad (3.4)$$

We find that most interaction terms do make significant estimates. In general, the sign and significance of the other explanatory variables remain unchanged.

Model (1.5) re-estimates model (1.4) with a reduced data set by limiting itself to observations for zip codes with houses and apartments in the same year. This is done to measure whether differences in the betas and spatial distributions are significant and whether the results are robust. Again, the results do not change in a significant way.

The results of the hedonic pricing estimates provide a limited insight into the extent to which certain variables contribute to the decreasing price gap. The respective contributions of each variable can be found via OD which is used as a means to identify the most important determinants of a gap between two groups - in our case between houses and apartments.

We add to the OD real estate studies of LeSage and Charles (2008), Mathä et al. (2017) and Zhang and Yi (2018) to compare two different structural forms of housing, potentially proxying different life styles (i.e. houses and apartments) using transaction data. Our dataset allows the decomposition for each year separately to find possible preference changes over time.

We start from the model design developed by LeSage and Charles (2008) and estimate the prices of houses (SFH) and apartments (AP) for each year as follows:

$$log\, P_t^{SFH} = \beta_{t,0}^{SFH} + \beta_{t,k}^{SFH} * X_k + \lambda_{t,n}^{SFH} * \mu_n + \varepsilon_t^{SFH} \quad (3.2.1a)$$

$$log\, P_t^{AP} = \beta_{t,0}^{AP} + \beta_{t,k}^{AP} * X_k + \lambda_{t,n}^{AP} * \mu_n + \varepsilon_t^{AP} \quad (3.2.1b)$$

The estimation of Equation (2.1a) uses the same variables as those of model (1.1), excludingtime dummies and the house dummy since we estimate each year separately for apartments and houses.1 The variables of models 1.4 and 1.5 cannot be used for an OD analysis since we need to use models without interactions between the houses and X_k.

Equation (2.1b) does the same for apartments. The yearly estimations of Equation (2.1a) for houses result in adj. R^2s in the range of 0.574 to 0.689. For Equation (2.1b), values of the adj. R^2s range between 0.691 and 0.810. The estimated results for each regression are in line with the literature and with our previous findings. (We also estimated equation 2.1a by including spatial lags as in (1.3) and found no significant changes in the main results. (The results are available from the authors on request.)

The gap between the mean log price estimates of houses $\bar{Y}^{SFH}$ and apartments $\bar{Y}^{APP}$, which are from this point on referred to as the descriptive gap, can be decomposed into an explained Q and an unexplained U; the latter can be attributed to discrimination (Jann 2008) as follows:

$$\bar{Y}^{SFH} - \bar{Y}^{AP} = Q + U \tag{3.2.2}$$

The decomposition of the explained part reads as follows:

$$Q = \left\{E\left(X_i^{SFH}\right) - E\left(X_i^{AP}\right)\right\}' \beta^* = (\bar{X}_1^{SFH} - \bar{X}_1^{AP})\beta_1^* + (\bar{X}_2^{SFH} - \bar{X}_2^{AP})\beta_2^* + \cdots \tag{3.2.3}$$

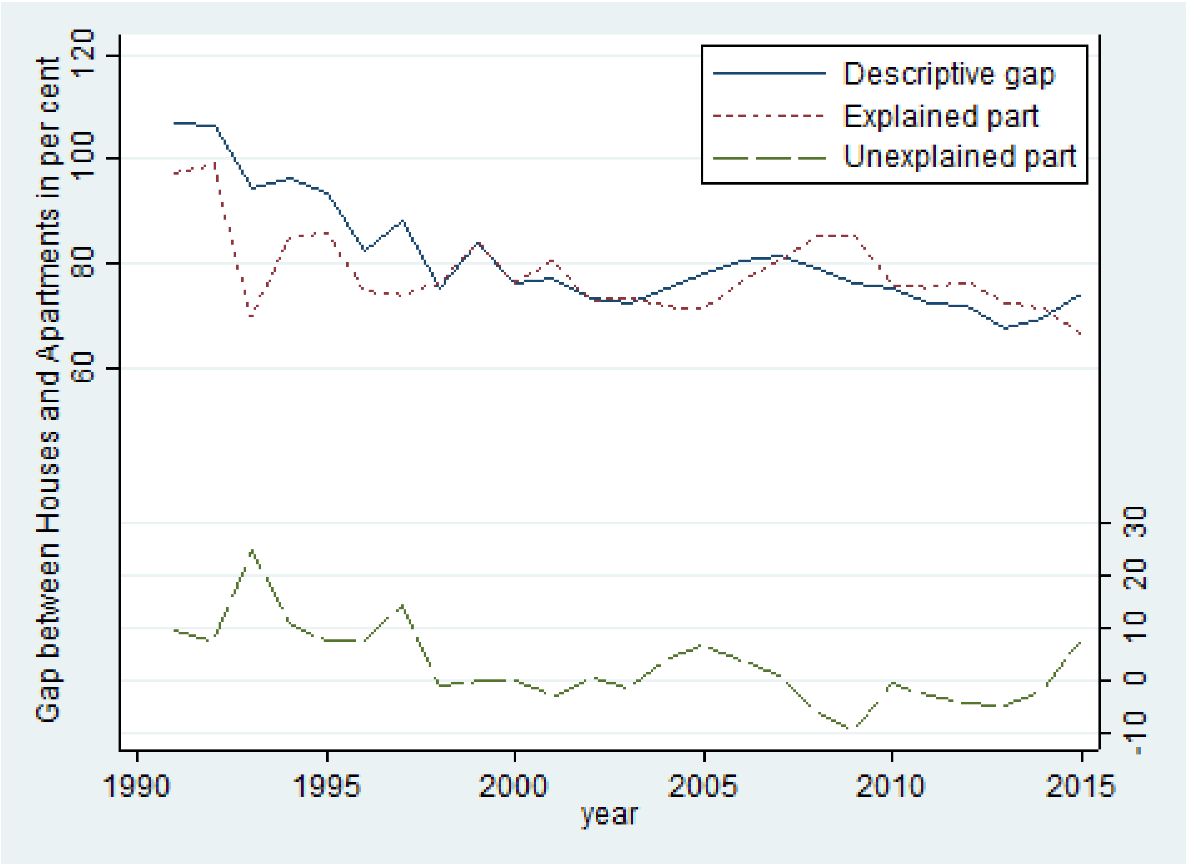
We estimate parameter β^* of Equation (2.3) in a pooled regression excluding a group dummy (Neumark 1988). (The results do not change significantly when we use a specification that includes the dummy.) Note that Q may be larger than the descriptive gap. Additionally, the impacts of individual determinants may be greater than those of Q when other impacts have a negative effect on the gap. The estimation equation reads as follows:

$$log\, P_t = \beta_{t,0}^* + \beta_{t,k}^* * X_{t,k} + \lambda_{t,n}^* * \mu_n + \varepsilon_t \tag{3.2.4}$$

Figure 3.3 shows the results of the OD test for the observed gaps. It replicates a decreasing trend in the descriptive gap as a percentage and shows the share of the gap that can be explained by diverging characteristics of apartments and houses. On average, 94% of the gap can be attributed to housing characteristics.

Figure 3.3 also illustrates the unexplained part of the gap. In labor economics, the unexplained part is usually considered to result from wage discrimination (Moral-Arce et al. 2012). In our study, we interpret the unexplained part as preference for houses. From 2008, preference in favor of houses has disappeared and even reverted to preference in favor of apartments (with the exception of 2015).

Figure 3.3 House Price Gaps Over Time Estimated via Oaxaca Decomposition



The graph shows the results for every year using formula 2.2. The descriptive gap $(\bar{Y}_t^{SFH} - \bar{Y}_t^{AP})$ follows the same trend observed in the hedonic model. From 1998 onward, there is no systematic unexplained part (U). The results are consistent with those based on hedonic regressions.

We now use OD to decompose the gap into respective effects of each explanatory variable. Figure 3.4 displays the results of this OD decomposition. To a large and even growing extent, the descriptive gap can be explained by the different sizes of houses compared to apartments. Figure 3.4 illustrates that from 2008, up to 120% of the descriptive gap can be explained by this difference in size, implying a growing preference for size.

The diverging age structure of houses and apartments and socioeconomic characteristics of neighborhoods have hardly any systematic explanatory power for the descriptive gap.

However, from 2008, a negative and absolutely growing part of the gap is explained by the different distances of apartments and houses to amenities. Distances to amenities offset up to 20% of the descriptive price gap, implying growing preferences for amenity structures typically related to apartments.

Figure 3.4 Oaxaca Decomposition of the Explanatory Variables

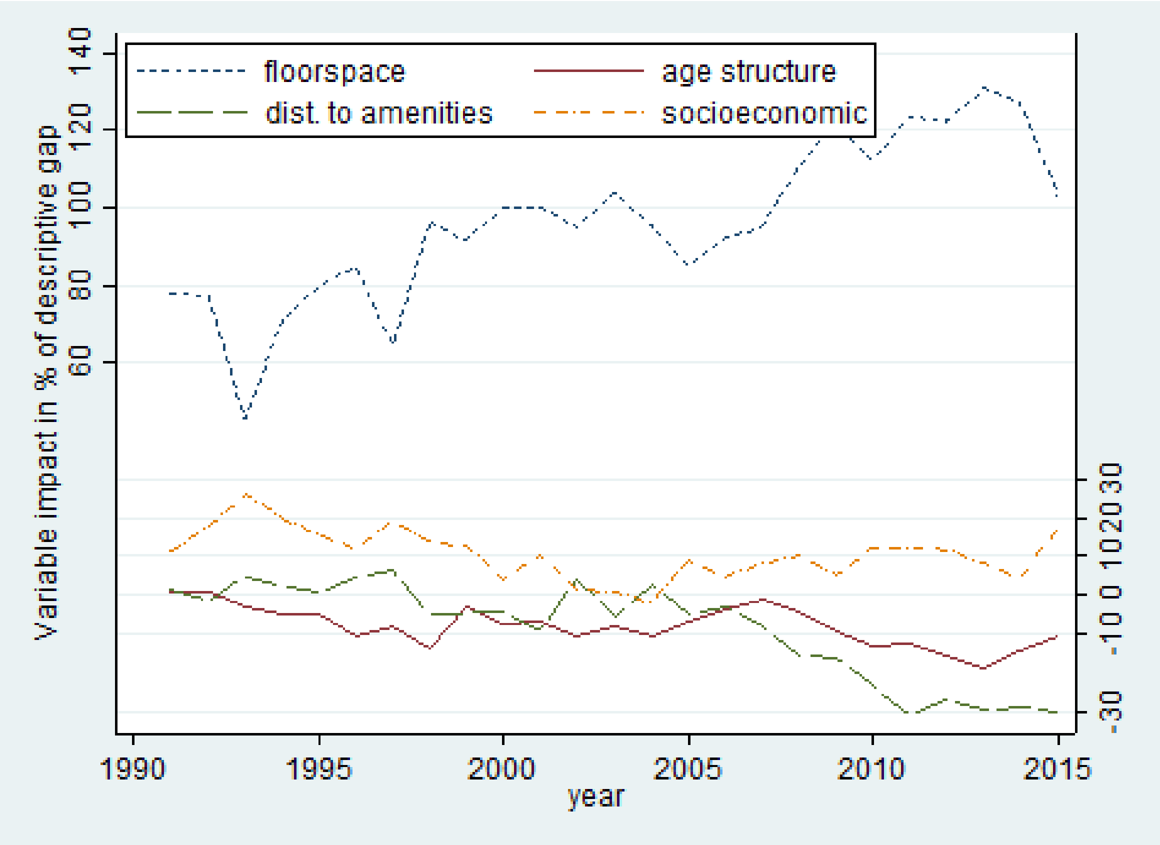
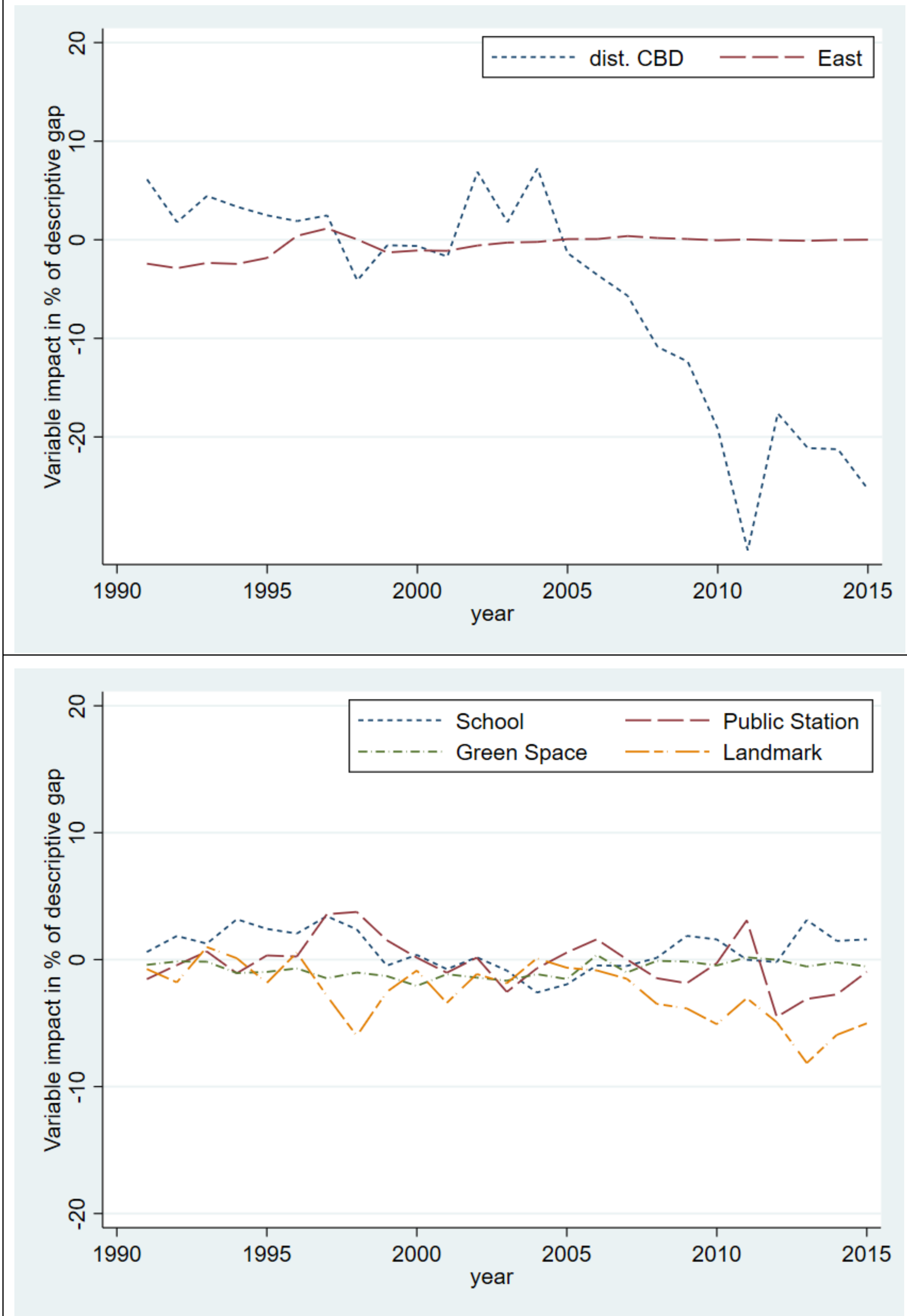


Figure 3.4 shows the decomposition of the explained part (Q) as shown in formula 2.3 for every year. Floorspace has the strongest effect in explaining the descriptive gap between houses and apartments, explaining up to 130% of the gap and indicating a growing preference for size. The effect of floorspace is offset (up 30%) by the effect of the distance to amenities, implying a growing preference for centrality provided by apartments. The age structure has no major impact. The effects of various amenities are differentiated in Figure 5. Results for socioeconomic variables vary considerably and lead to mostly insignificant results.

Figure 3.5 decomposes the distance effects of different amenities. It illustrates that the largest proportion of amenity preferences is attributed to the distance to a CBD, implying a growing preference for more central locations. This preference is a more recent development; until 2005, if anything, a greater distance from a CBD was preferred.

Figure 3.5 Oaxaca Decomposition of Amenities



Distance to a CBD is the main source of the negative amenity effect observed in Figure 3.3. There is a strong positive trend regarding preferences for centrality from 2005 onward.

3.3 Summary

We find an initially positive but declining nominal price gap between houses and apartments for Berlin between 1990-2015. Using Oaxaca decomposition, we find two major determinants explaining the overwhelming proportion of the nominal gap, namely, size and location. There is a growing preference for size that favors houses. From 2006 onward this trend should lead to positive price signals for houses. However this effect is offset by a growing preference for housing in more central locations since 2006. When controlling for the two determinants, only a minor price preference in favor of houses remains in the early nineties. Since 1996, preference in favor of houses has disappeared, and in some years preference has even reverted in favor of apartments. However since 2007 there seems to be a reviving preference for houses.

4. Impact of Social Structure on Bike-Sharing Demand: Evidence from StadtRAD users in Hamburg[5]

Abstract: This paper analyses the effects between the social structure of the general population surrounding a bike-sharing station and the demand for that station. Compared to other European countries, Germany has a rather big bike-sharing market with systems in more than 40 cities. The most extensive German bicycle sharing system (StadtRAD) is placed in Hamburg and used for this case study. The demand is estimated using daily trip starts for every station. Our empirical approach consists of different configurations of semi-log models. The models control for standard explanatory variables shown in the literature to affect bike-sharing usage, i.e. weather, season, lagged demand, travel distance and infrastructure. The models extend the literature by including real-time bicycle availability and socioeconomic structure variables like age and gender. Our results reinforce findings in the literature for baseline variables. In addition, most notably, a higher share of young adults and females positively influences demand. Bicycle availability also has a positive impact.

Keywords: Bicycle Sharing Systems, Public Transportation, StadtRAD, Hamburg

4.1 Introduction

Bike-sharing (BS) systems have been around for more than half a century (DeMaio 2009). However, with the implementation of modern Radio and Global Positioning System (GPS) technology, there has been a substantial uptick in the last two decades (Parkes et al. 2013). These technologies make it easy to use and track the bikes and have an efficient billing mechanism. BS systems are mainly stationed in Europe, North America and foremost Asia (here mainly China). BS is often introduced with assumptions regarding its positive influence on the urban environment. Some examples are shorter travel time, better public transport connections, decreased transportation noise and air pollution, higher public health, better road safety and promotion of cycling (Fishman et al. 2013). In a quantitative study, Otero et al. (2018) show that BS in Europe has economic and health benefits, although the results differ between cities.

BS is a mixture between traditional cycling and public transport. It is part of the shared economy and thus probably shares user types with other products provided by the shared economy (Standing et al. 2019). Since cycling is required to use a BS station, the same effects influence BS and bicycle usage. Heinen et al. (2010) made the most recent literature overview

[5] I would like to thank the participants of the PhD Seminar at the University of Hamburg for their feedback and helpful suggestions.

of influencing factors for bicycle usage. Their principal findings include the negative effect of travel distance, an adverse effect of bad weather and car ownership. Most socioeconomic variables show inconclusive results. The results primarily depend on the country of interest. For example, women seem to cycle less except of areas where cycling, in general, is more common (Belgium, Netherlands). In addition, Tin Tin et al. (2012) provide an in-depth analysis of the effects of weather and seasonal effects on cycle volume. Their results are in line with the general findings of Heinen et al..

Moreover, it is unclear how BS as part of the shared economy, and traditional public transport interact with each other. There are some indications that BS is complementary to traditional public transport regarding the last and first mile. Shaheen and Chan (2016) discuss these effects concerning all forms of shared mobility and traditional public transport. The effects between BS and other forms of shared mobility are also still not well understood. Ceccato and Diana (2018) show that at least between car-sharing and bike-sharing, there are strong complementary effects. This supports the results regarding public transport that BS is often used for first and last mile travel.

This study aims to empirically analyze the impact of the social structure on bike-sharing demand using Hamburg in Germany as a case study. Similar to Belgium and the Netherlands, Germany has a relatively high cycling rate (Quality of transport 2014). BS systems are relatively common in German cities. Around 40 German cities have at least one BS vendor; some like Berlin have many more (Wikipedia 2020). The biggest provider in Germany is the Deutsche Bahn (German Railway Company) with their “Call a Bike” system (www.callabike-interaktiv.de). Our case study is the most extensive German BS system StadtRAD, which is located in Hamburg and part of the Deutsche Bahn system. Even though cycling in general and BS in particular are well established in Germany there is no quantitative research using German data. This poses a substantial lack of public information, since BS demand and user behavior could be different compared to other regions, and validation for existing results is needed.

StadtRAD was implemented in 2009 as a joint venture between the city of Hamburg and the Deutsche Bahn. It has over 200 stations around the city of Hamburg. Most stations are located in close proximity to public transport or points of interest (POI). These POI include for example tourist attractions, larger companies, shopping centers and parks. The highest station density can be found in the central business district. Since nearly all major public transport stations are equipped with a StadtRAD docking station users can use it for the first and last mile of their journey (Liu et al. 2012). Around 2100 bicycles are available in the system. After registration, bicycles can be rented by using a debit card at the station, calling a phone number, or using an application (APP) for mobile phones. Similar to many other BS systems, the first 30 minutes of every trip are free of charge.

From a technological point of view, StadtRAD is in between traditional BS systems using exclusively docking stations like BIXI (Faghih-Imani and Eluru 2016) and dockless BS systems widely used for example in China (Chang et al. 2018). StadtRAD users can park bicycles at any BS station without needing a free docking slot. They just need to be close to the station. Therefore, stations have more than 100 bicycles parked simultaneously on some occasions. This can happen especially if a significant event takes place in close proximity to a single BS station. The StadtRAD APP shows the number of available bicycles for every station. For users not using the APP, this information is additionally available on the StadtRAD website. Therefore, users rarely have a bad experience by going to empty stations. However, this information is not stored and thus not available in our historic dataset.

Before 2010 research analyzing bike-sharing demand was rare. A search in Google Scholar for “bike-sharing” shows around 18.500 results in total, but only 446 were published before 2010. The main reasons for this effect are probably a small number of existing systems and in addition deficient availability of usage data for these systems. The old systems did neither save nor collect renting data. However, since the mentioned changes in technology took place, far more systems have been implemented. These new systems produce a far more utilizable form of data. This increase in available data has led to a surge in studies analyzing different aspects of BS since 2010. There are recent studies by Ricci (2015), Fishman (2016) and (Eren and Uz 2020) reviewing this literature. Obviously, these studies have some overlap in examined literature. Their focus points regarding effects and region of interest are different and complement each other concerning actual results. We will analyze the impact of the social structure on BS demand taking the studies‘ main findings into account. Weather variables have been shown to have an impact, which is higher for leisure usage (Ashqar et al. 2019; Kim 2018; V E and Cho 2020). Litman (2004) finds similar results regarding the price elasticity of public transportation. They show an elasticity increase of around 100% for leisure trips compared to commuting trips. The availability of bicycles has shown to have an effect. However, this effect was mostly measured through proxy by looking at station capacity (Zhang and Pan 2018).

There are connections between public transportation and BS; however, the results differ strongly between the form of transport and between case studies (Ulrich Leth et al. 2017). Most recently, Radzimski and Dzięcielski (2021) have analyzed the reason behind these inconsistencies. They find that there is a positive effect between BS trips up to 3km and the frequency of public transport. Therefore, BS and public transport act as complements for example for the last or first mile of a trip but there is no interaction for longer BS trips. The socioeconomic structure of BS users differs from the socioeconomic structure of the general population. It also seems to differ between the cycling population and BS users. Compared to an average cyclist, BS users are more often female, younger, and have a lower income level. Differences in the impact of the explanatory variable on BS usage between age cohorts were analyzed by Wang et al. (2018). Wang et al. find no essential differences between age groups regarding temporal and weather effects. Similar to studies using survey data, they find differences in general BS demand between the cohort groups. The effects of the socioeconomic structure of the general population surrounding a BS station on BS demand are mostly ignored in the literature. Clark and Curl (2016) look at the social structure surrounding BS stations and compare it with the average structure of the city. However, they do not analyze how these differences affect the usage of said stations. We aim to close this gap in the literature especially. We use semi-log models estimating the daily trip starts per station for our analysis. While controlling for variables shown in the literature to affect bike-sharing demand i.e. weather, season, lagged demand and infrastructure, we find evidence for socioeconomic structure influencing BS demand.

4.2 Data

We generated our unique dataset by merging and aggregating four different data sources. Consisting of the StadtRAD dataset (see 4.2.1 and 4.2.4), weather data (see 4.2.3) and data provided by the Statistical Office for Hamburg and Schleswig-Holstein (see 4.2.5) as well as the geoportal-hamburg.de (see. 4.2.2). Regarding the baseline variables like time effects, weather, infrastructure and trip distance, we follow the meta-studies of Ricci (2015), Fishman (2016) and Eren and Uz (2020). Studies regarding general cycling are primarily survey-based and do not control for variables not included in BS literature. Since the results regarding variables important to general cycling and BS are identical, we stick to the BS literature (Damant-Sirois and El-Geneidy 2015). The data types as well as their processing are explained in the following sections.

4.2.1 StadtRAD

Our data source for the StadtRAD dataset is the Deutsche Bahn AG (2017). The data is available under a Creative Commons Attribution 4.0 license. It is openly available for further research like the rest of our data. We strongly encourage using the data to tackle more questions regarding bike-sharing. The raw StadtRAD dataset consists of every trip made with a StadtRAD bicycle in 2015 (approximately 2.5 million) and 2016 (approximately 3 million). The data is not available for the years prior to or after. Detailed information is available for every trip. This information includes start and end station, unique bike ID, unique User ID as well as the time of trip start and end. We obtained the GPS coordinates for every station.

However, the amount of docking stations per StadtRAD station is unavailable. This variable is sometimes used in the literature to proxy the availability of bicycles. It also often indicates the possibility to end a trip at a particular station, since in some BS systems it is not possible to end a trip without a free docking station. A higher amount of docking stations indicates a higher station usage in the studies using it as a variable. There could be an endogeneity problem with using the number of docking stations for bike availability. Stations with more docking places are likely placed in areas assumed to have a higher demand like central stations. This could lead to biased estimates.

Additionally, systems requiring free docking stations could in some cases lead to misleading results since users cannot use their preferred stations to terminate a trip. This can happen if there are no free docking stations available. Therefore, travel distance as well as renting time would be increased in comparison. However, this reasoning does not apply to our study for the following reasons. The real-time bicycle availability is on hand for the users in the StadtRAD APP. Therefore, they do not need docking stations as a proxy. However, bicycle availability is not included in the dataset. Therefore, we calculated this variable ex-post through the data at hand. To end a trip in the StadtRad BS system it is unnecessary to have a free docking station. If all docking stations are occupied it is sufficient to park in close proximity to the station. Not including the number of docking stations seems unproblematic for these reasons. The exact time of departure and arrival are included. By definition, the trip duration is the difference between both. Every bicycle in the system as well as every user has a unique identifier in the dataset. All bicycles are identical except possible wear through usage. We could not obtain additional data regarding the user characteristics like age or gender.

Our approach using historic real trip data is comparable to other studies (El-Assi et al. 2017; Médard de Chardon et al. 2017; Caulfield et al. 2017; Tran et al. 2015; Faghih-Imani et al. 2017; Faghih-Imani and Eluru 2016) analyzing BS system usage. Our main improvement regarding the actual data usage is a method to calculate accurate bicycle availability. This method determines the number of bicycles available at each station for every moment in time. Since no additional information in the raw data is required, future studies could easily improve their results compared to existing literature, using our method. Since these improvements strictly rely on data improvement, they are independent of the empirical approach or the tested theory. The trip dataset was extended to include this availability value.

The approach to calculating the available bicycles is as follows. Having a dataset only including information about the trips we need to take several steps to acquire the variable capturing bicycle availability. The first information needed is the initial bicycle distribution between the stations. This means the starting position of every bicycle at a fixed point in time aggregated to the initial endowment of every station. We have no initial information in the raw data for bicycles leaving/entering the system over time. The base endowment for every station is calculated anew every $\mathrm{Month_j}$ separately. This controls for possible changes in the bicycle endowment of the BS system as a whole. For every station, we aggregate the initial starts of any bicycle in the system. Considering every bicycle has only one initial starting station per month, this leads to the starting distribution. Depending on the research question, one could use this calculation for initial endowment more frequently. For example, if analyzing shocks. Having the initial endowment, we need to keep the information up to date over time. For this we keep track of each bicycle movement i.e. trip made.

When a trip starts at a station the number of available bicycles is reduced by one from this moment in time onward. In the same way, if a trip ends at a station the bicycle availability increases by one. Trips that do not terminate at a station are excluded from the dataset. This can happen if users lose the bike, or the bikes are damaged heavily while used and left behind on the street. This method's primary concern is bicycle redistribution between stations through the BS system operator. Redistribution is not included in the raw data, which only includes trips made by BS users. We expect other BS data sources to have a similar challenge.

Nevertheless, through every bicycle having a unique ID, we can control for redistribution by observing the moving pattern. If a bicycle terminates a trip at station B but starts the next trip at station A, it can be assumed that it was redistributed through an operator. For these instances, the availability at station B is not increased after the termination. Instead, the availability at station A for the next trip made with the bicycle is increased. Using this method we also consider bicycles that are tagged by their users as damaged after returning them to a station. These Bicycles cannot be rented until the operators repair them. Normally repairing them requires the bicycles to be collected and repaired in a workshop. After that the repaired bicycles are redistributed. This fulfills the same requirements as normal distribution so it is considered in our estimated availability. Bicycles can be parked independently of the docking capacity at every station. In some cases of huge events, this leads to over 100 bicycles for particular stations.

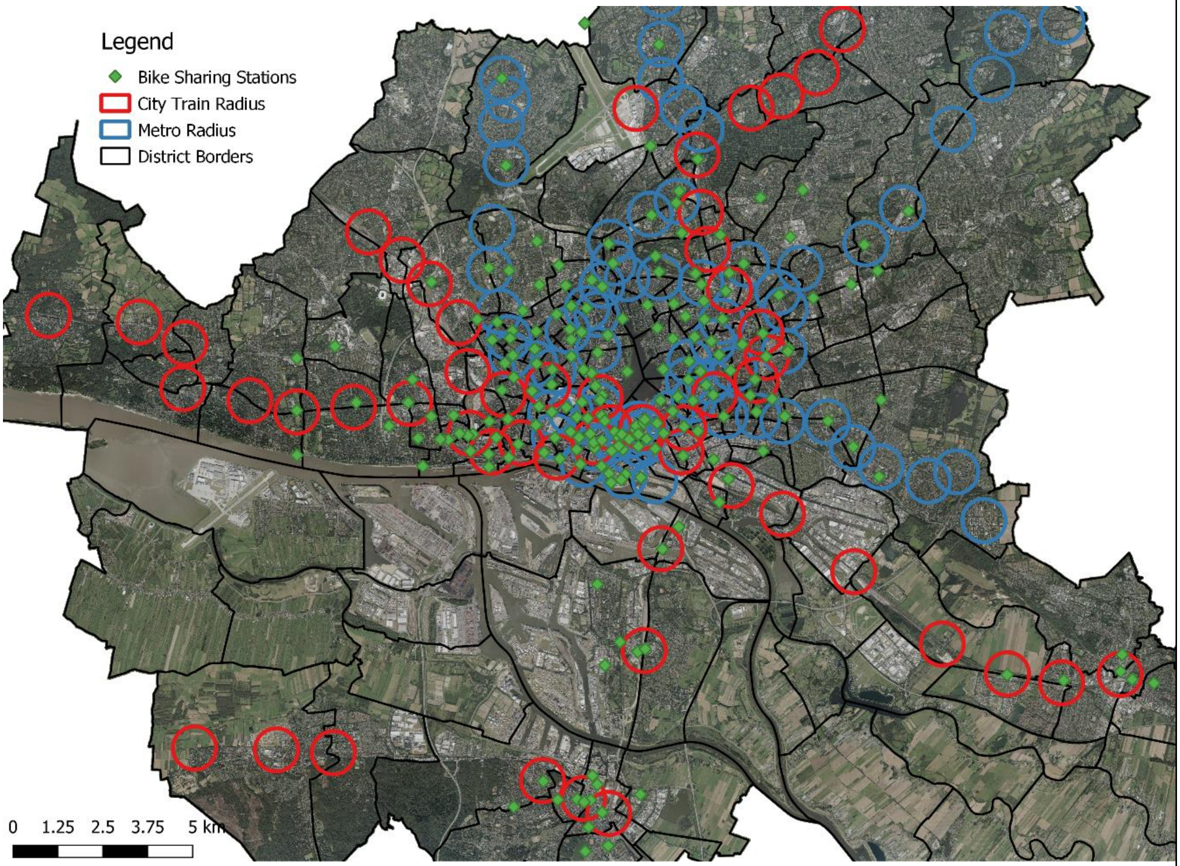
Actual travel distance cannot be tracked directly in the dataset. Following El-Assi et al. (2017), we used the Google API to obtain real-world optimal travel distances in meters for cyclists between all station pairs. The pathways used by Google are optimized for travel time.

4.2.2 Public Transportation:

Like many forms of shared mobility, bike-sharing has a lot in common with traditional public transportation. For example, it has fixed stations operating as start and finish points. In contrast to regular public transport, there is no fixed schedule between the stations. Depending on the circumstances, bike-sharing can substitute or complement these traditional forms of public transport. In Hamburg, public transport consists of buses, city trains and a metro. Buses are used to travel relatively short distances compared to the city train. Most bike-sharing trips are also rather short compared to distances traveled by a city train. It is unclear how the users interact between the different kinds of public transport. Inner-city public transport tickets are independent of the form of public transport. Even though StadtRAD like the city train system belongs to the Deutsche Bahn there is no direct connection between tickets bought for one system on the price of the other system. Hamburg has a rather well-established bus system so that nearly every BS station is in walking proximity to at least one bus station. Former studies have shown inconclusive results on how public transportation influences bike-sharing and vice versa. Martin and Shaheen (2014) have shown that the interaction between public transportation on rails and bike-sharing depends on the spatial position. Areas in the city center had a negative interaction while areas further away showed positive interactions. As seen in Figure 4.1 the Metro system in Hamburg is stronger in the inner-city area while the city train also connects areas that are more rural. Since there are over 1300 bus stations, they are not included in the Figure. We use a catchment radius for city train and Metro stations of 600 meters. Bus and bike-sharing stations have a shorter catchment radius of 200 meters. The public transportation operator defined the catchment radius for bus and rail stations (Geoportal-Hamburg.de 2018). We assume that BS and bus stations have a comparable catchment radius.

An obvious interesting economic connection to analyze would be the interactions between bike-sharing pricing and public transport pricing. However, we faced insurmountable econometric challenges. We observe zero price increases for the bike-sharing systems in our timeframe and in addition, approximately 90 percent of the trips are below the pay-to-use time of 30 minutes. There was only one price increase in the public transport prices between 2015 and 2016 of 2.6 percent (Hamburger Verkehrsverbund 2021). With little variance and no information on how the unique StadtRAD user IDs use public transport, directly estimating causal effects is impossible. We observe an increase in total usage of the BS system that would be counterintuitive if the price increase had a significant negativ impact.

Figure 4.1 Map of Hamburg



The map shows the main part of Hamburg, including every Bike-sharing station. Bike-sharing stations are shown as green dots that are mostly allocated in the city center and around public transport lines. For public transport, we distinguish between City train and Metro Stations. Both types of public transport are illustrated with their respective entry radius. For clarity reasons, only the circuit is shown. Districts are separated by black lines and give an overview of the spatial distribution.

4.2.3 Weather

Many studies show that weather conditions impact bicycle and BS usage (Dill and Carr 2003; Nankervis 1999). We use official historical weather data recorded at the closest weather station run by the German Meteorological Service. The data is reported hourly. The variables are averages or aggregates depending on the measured outcome. For example, the temperature is average and precipitation is the whole precipitation for the given hour. Following the literature, we expect some variables to indicate "good" and some to indicate "bad" bicycle weather. The "good" variables should have a positive impact on BS and the "bad" ones a negative one (Connolly 2008; Smith 1993).

Cools et al. (2010) show that the effect strength differs depending on the trip purpose. Possible purposes are for example leisure time or commuting to work. Leisure time usage includes using the BS as a leisure activity itself as well as traveling to indoor or outdoor leisure activities. If the weather is bad the chances for BS usage as an actual leisure activity or commute to outdoor leisure activities are relatively low. Even traveling to indoor leisure activities with BS seems less likely. Using BS for work commute if the weather is bad is mostly still possible but it seems reasonable that people would prefer other ways of transport to stay warm and dry. However, commuting to work has to be made independent of the weather. We expect lousy weather variables to have a lower elasticity on weekdays compared to weekends. The variables indicating good weather[6] are the number of sunshine minutes per hour as well as the temperature in C°. The variables perceived as bad for BS usage are rainy weather measured in precipitation in mm, humidity measured in percent and wind strength measured in the standard Beaufort scale. In contrast to other studies (El-Assi et al. 2017), we did not use the apparent air temperature. A formula is used to calculate the apparent air temperature, which includes other weather variables. Since we include those variables in the regression, the estimates would be highly correlated. (DWD Climate Data Center 2018)

4.2.4 Temporal Controls

Figure 4.2 shows the average number of trips undertaken around a day, week and month in 2015 and 2016. It shows that StadtRAD usage is highly heterogeneous over time. These temporal distributions for all time units are similar to other studies (Ricci (2015) and Fishman (2016)). This makes it evident that we need to consider the temporal dimension of the data. Otherwise, some estimates could be biased, or we might have a general omitted variable bias. Seasons do have an impact on the road conditions for example, as well as holiday seasons. Estimates that are observed more frequently like the weather would likely see higher bias than variables that are available on a yearly basis. In a first step we compared the trip distributions for an average day, week and over the years 2015 and 2016 (see Figure 4.2). For every temporal dimension, we checked for significant differences between BS usage in 2015 and 2016. 2016 had more trips than 2015 such that we compared the distribution in percent to make them comparable.

For the intraday as well as the intraweek distributions, there are no significant[7] differences between 2015 and 2016. That is why we did not differentiate our empirical approach between

[6] A „good“ variable indicates good weather if their value increases and vice versa a “bad” variable indicates bad weather if their value increases.

[7] We inspected the estimates for 2015 and 2016 to check for not overlapping confidence intervals. As long as the confidence intervals overlap the estimates are not significantly different from each other. For better visualization the confidence intervals are not included in the graph.

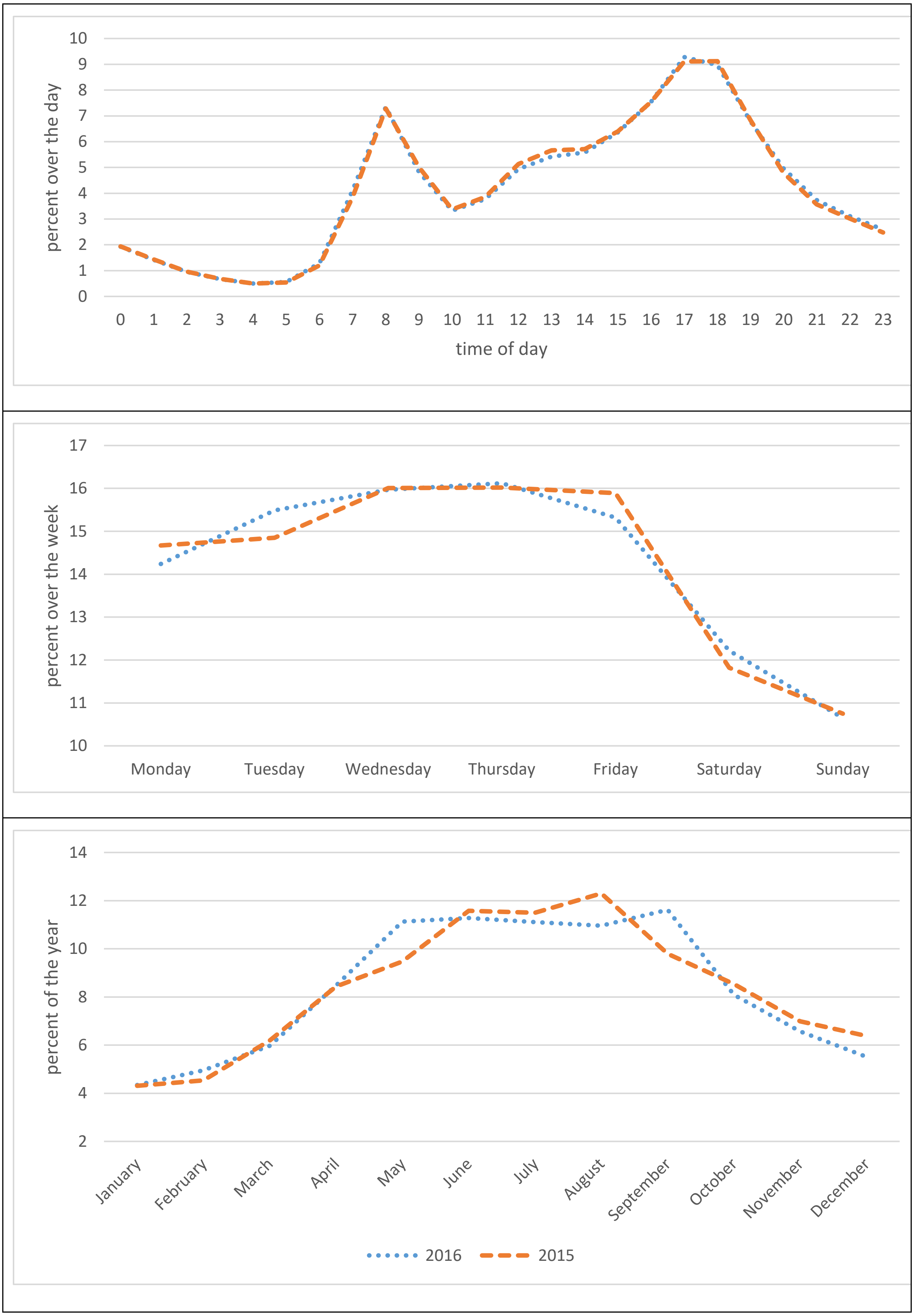
the years for these two time units. For the intrayear distributions in 2015 the demand in May was significantly higher and in September significantly lower compared to 2016. Our outcome of interest is the daily trip starts per station. To obtain unbiased estimates, we need to control for these observed variances in demand. We use the following control mechanisms for the different time dimensions. The intraday distribution is in general irrelevant for the aggregated daily demand. However, we need to scale them accordingly to obtain unbiased estimates for the explanatory weather variables. This is achieved through weighting the trip distribution. This process is further explained in chapter 4.2.6.

The intraweek demand from Monday to Friday is relatively consistent (business days); the same is true for Saturdays and Sundays (weekend days). However, the demand on business days is on average higher compared to weekend days. To control for the intraweek distribution, we differentiate between weekend and business days. The main reason behind this gap in BS usage between business days and weekends should be the trip purpose. BS demand on weekends is mostly leisure driven while on business days work commute is an additional trip purpose. Weekends are coded as a dummy variable. The weekend dummy also includes public holidays to take into account the decrease in work commute for these days.

Lastly, we need to control for the intrayear distributions[8] as well as the in general higher demand in 2016. We can control for both facts by using a set of dummies $\mathrm{Month_j}$. $\mathrm{Month_j}$ has a unique dummy for every month in our dataset. Using separate dummies for every month in both years allows the estimation to be more precise. It takes care of the discrepancy in the two mentioned months. It also controls for the fact that the total usage in 2016 is higher compared to 2015, and all other effects can be interpreted as changes in percent.

[8] Which as shown slightly differs between 2015 and 2016.

Figure 4.2 Temporal distribution of trip data



4.2.5 Socioeconomic Structure

Most BS studies using trip data fall short on taking into account the effects of the socioeconomic structure. So far, studies using trip datasets comparable to ours only used population (-density) and employment (-density) on district/zonal level (Caulfield et al. 2017; El-Assi et al. 2017; Tran et al. 2015; Faghih-Imani et al. 2014). Both variables indicate a positive effect on the BS demand. Wang et al. (2018) differentiate between age cohorts in their study. Similar to studies using survey data, they use the age distribution of the actual users and not of the general population. They also do not compare the effects of age on BS demand but instead the differences of the other explanatory variables between the age cohorts. However, studies using survey data have shown that other socioeconomic variables (so far ignored in our kind of setting) play a role. All survey studies control at least for age and gender of the participants (Chen 2016; Fuller et al. 2013; González et al. 2016).

Like most studies using historic trip data, we have no direct information about the user characteristics. On the district (Stadtteil) level, we have precise information on the socioeconomic structure. Hamburg consists of 104 districts, as displayed in Figure 4.1. StadtRAD is available in 53 of these districts. The official statistical bureau provided information about the socioeconomic distribution in the districts. All socioeconomic variables are aggregated at the district level. Clark and Curl (2016) have shown that the population living in proximity to BS stations differs from the general population. This indicates that there should be differences between the general BS usage of a city and the usage of a specific station depending on the population structure.

Following other trip data studies, we include the population density. We have no data on employment density. As stated before, age structure and gender distribution should have an effect. We are the first study using trip data estimating the effects of these variables on the district level. The age structure is available as the average age in the district as well as age group distribution. Since age has shown no linear effect in the literature, age groups should lead to more precise estimates (Eren and Uz 2020). The study period is not sufficient to distinguish between age and cohort effects. This could mean that we actually measure cohort effects. However, this problem cannot be solved with one study and needs to be approached in future meta-studies. Gender distribution is included as female share.

We include the foreigner rate as an additional control variable. Another variable shown to have an impact on public transport usage in general and BS in particular is income. We include the average district income per employed capita. In addition to other forms of public transportation, district residents could use their own car as a form of transportation. Especially since a mode shift away from cars is a desired effect of bike-sharing, one should check if car ownership/availability has an effect. To control for it we include cars per capita. Leyden (2003) has shown that there is a connection between building environment and social capital measured as participation in public voting. In addition, strong public participation can increase the quality of planning and infrastructure (Tang and Waters 2005). We include public participation as voting share in the last Federal election. Using regional election data leads to comparable results. (Statistisches Amt für Hamburg und Schleswig-Holstein)

4.2.6 Sample Selection and Merging

Our initial StadtRad dataset consists of around 5.5 million trip observations. However, not all trips observed in this set are valid or include every information needed for our research. As a first step, we need to exclude observations departing or arriving at a temporal station. Temporal stations are sometimes set in place if a regular station is under repair or major infrastructure work is done blocking this station. Temporal stations have similar usability as regular stations. However, our set does not include the spatial positions for these temporal stations. We would be unable to merge or generate our explanatory variables for trips starting or ending at these stations. Our variable determining the number of bikes at a given station for every point in time was generated before these observations were dropped. This way, the timeframe for the actual availability is most precise. We set a minimum timeframe of 3 minutes for a trip, explained in detail later. We did not set an upper bound on trip length. The maximum cost to rent a bike for the whole day is capped at 12€. Using a bike for the whole day making leisure trips around the town or to the suburbs seems reasonable. Obviously, there can be cases where a bike is rented but not returned. These trips are not included since they lack an ending station for the trip.

Like other BS systems, most journeys (~90%) are below the ‘pay to use’ threshold. This threshold is 30 minutes per rent and resets for every new trip. Users have an incentive to split longer trips, so each individual rent stays below that threshold. Since the goal is to estimate real demand for the BS stations, these “intermediate stops” would lead to an estimation bias. Our selection process to what is an intermedia stop works by using the unique user id. Intermediate stops are defined as having a timeframe between arrival and departure at the same station below 3 minutes. Stops under 3 minutes seem unreasonable for actual leisure or work activities. Another possible explanation for this kind of intermediate stop is recognizing that the bicycle is somehow damaged (broken light, bad brake) and needs to be replaced by a better one. Approximately 10% of all observations fall under this definition of intermedia stops and are excluded from the dataset. These stops do not indicate real demand at the intermedia station and are taken care of through our method. Even so, many BS systems have similar forms of free usage. We are unaware of any other study considering this problem. The last kind of dropped observations are trips below 3 minutes, starting and ending at the same station. This kind of trip seems unreasonable to have an actual function. Our approach to very short trips is similar to El-Assi et al. (2017), who disregarded all trips below 30 seconds as invalid trips. However, unlike El-Assi et al. we included trips below the time threshold, if the starting and ending stations differed. For these trips, a real journey took place.

Using this cleaned dataset, we have 4.6 million usable trip observations. We use a stepwise approach to aggregate the information to the correct spatial and temporal dimensions. The first step is to aggregate the trip data on an hourly station level. We sum the trip starts per station for every hour to get the hourly trip starts for every station. This step includes calculating the average amount of available bikes for this hour at the station. It also determines the average trip length made in that hour for each target station. The next step is to merge this hourly station level data with the weather data. The last step is aggregating this hourly data to our final level of aggregation the daily trip starts. As mentioned before, we want to control for the intraday variation of trip starts by controlling for the weather variables. We expect the weather to have a higher impact on the overall demand through times of the day like the rush hour in the morning or evening compared to the night where demand is relatively low[9]. We use a temporal weighting matrix to consider this higher importance while aggregating the hourly to daily demand.

$$W_{h,t,\mathrm{k}} = \frac{\sum_{k \neq t} trips\ in\ h}{\sum_{k \neq t} all\ trips}$$

$W_{i,k,t}$ is a unique weighting matrix for every day t. k is an index for the day of the month. Our weights are the sum of all trips made in an hour h in one month divided by all trips made in this month. We do not include the trips of day t for which we design the weight for calculating the weighting matrix. We are aware that this approach in general could lead to endogeneity problems. In general, we weigh independent variables using the dependent variable. Alternatively, to frame it differently, the trip demand is dependent on the independent variables, which we want to weigh. We avoid this endogeneity problem by using the average daily trip distribution of the corresponding month not including the actual day. After multiplying our explanatory variables (weather, number of available bicycles and average travel distance) by $W_{t,h}$ and aggregating on daily trips, we have our variable of interest i.e. the daily starts per station with the respectively weighted independent variables.

The last step is to include the time-invariant infrastructure and socioeconomic variables. The infrastructure variables are obtained by using QGIS to check for the number of respective

[9] See Figure 4.2

public transport stations in proximity of the BS station. Similarly, the socioeconomic variables are merged by matching every BS station with the respective city district and extending the observations with socioeconomic variables representing the population of the corresponding city district. Our final dataset consists of 127,037 daily station observations.[10]

Table 4.1 shows the descriptive statistics for all variables. Some interesting outliers are the maximum bicycle availability, which corresponds to significant cultural events close to stations with few other public modes of transportation. Furthermore, some of our socioeconomic variables show a relatively high spectrum. For example, the average income per capita lies between 15.800€ and 120.000€ between the districts. In addition, the age-group discrepancies between the districts are higher than we expected. Showing a minimum difference for the age group 50-64 of 9.7 percent and a maximum difference for the age group of 65 and above of 22.5 percent.

Table 4.1 Descriptive statistics

Variable	Units	Mean	SD	Min	Max
Logtrips	Log of trip starts	3.16	1.04	0	5.71
Distance	km	2.47	1.13	0	15.64
Weekend	Dummy	0.30	0.45	0	1
Avail. Bikes	# of bikes at station	9.91	8.13	0	138.92
Wind strengh	m/s	4.38	1.83	0	16.7
Rainfall	mm per day	0.068	0.23	0	16.47
Air temperature	C°	11.24	6.91	-8.1	33.22
Sunshine	Minutes per hour	14.96	15.27	0	60
Humidity	percent	76.44	13.28	26.62	100
City train	# of st. in proximity	0.48	0.64	0	2
Metro	# of st. in proximity	1.18	1.35	0	6
Bike-sharing	# of st. in proximity	0.13	0.35	0	2
Bus	# of st. in proximity	6.85	3.73	0	18
Foreigner rate	Percent	18.63	7.80	8.9	44
Pop. density	Inhabitants per km²	6,727.40	3,926.83	167.27	18,191.33

[10] Stations that do not produce trips for any given day are excluded from the dataset for that day.

Income p.c.	1,000 €	41.71	20.28	15.83	120.71
Cars	Per capita	0.28	0.058	0.12	0.50
Female share	Percent	49.77	3.35	37.96	54.72
Age <18	Percent	13.16	2.90	6.2	22.8
Age 18-24	Percent	8.03	2.82	4.8	21.7
Age 25-29	Percent	10.84	2.86	2.8	21.8
Age 30-49	Percent	36.52	4.52	25.6	44.6
Age 50-64	Percent	16.94	1.63	11.8	21.5
Age >65	Percent	14.48	3.67	4	26.5

Notes: Our observation count is 127,037. Observations are defined are defined on the daily station level. The columns of the table show the variable names, the measurement units, the mean, the standard deviation (SD), the minimum outcome (Min) and the maximum outcome (Max). Mean and SD are defined on the observation level.

4.3 Empirical approach

As shown before, estimating BS demand with econometrical methods is a relatively new field of study. We can look at the broader context of urban economics and more precisely to real estate economics for estimation models. BS as well as real estate models estimate the factors affecting the demand of a stationary good. Real estate economics uses hedonic models to estimate the price of a building. It uses structural variables as well as neighborhood and environmental attributes (Bolitzer and Netusil 2000). Our socioeconomic variables are similar to a real estate study on Berlin by Ahlfeldt and Maennig (2015). For Hamburg, a comparable set of variables was used by Brandt and Maennig (2011). We utilize the knowledge acquired in this field and apply it to our estimation problem. We need to estimate the effects of neighborhood and environmental attributes, which are very similar to the ones in real estate economics since they somewhat indicate the quality of living in an area. Our outcome variable is the daily trip starts per station and consists per definition of only non-negative observations. Using a standard semi-log model, we can control for possible outliers and ensure that the outcome variable has a suitable distributional form. Another advantage is that the estimates are interpretable directly as demand elasticities (Rasmussen and Zuehlke 1990; Bolitzer and Netusil 2000; Ahlfeldt et al. 2017). We estimate and compare four different empirical approaches. The models differ with regard to the interpretation of the effects as well as the temporal and spatial controls.

$$log\, Y_{i,t} = \beta_0 + \beta_k * X_{ki} + \gamma * V_{gi} + \alpha * WE + \tau_k * WE * X_{ki} + \omega_j * \mathrm{Month_j} + \varepsilon \tag{4.1}$$

$Y_{i,t}$ indicates the number of trips generated per day at station i at day t; β_0 is a constant; X_k is a vector including the weather variables, bike availability and travel distance; V is a vector including socioeconomic and infrastructure variables; WE stands for the weekend dummy taking the value 1 if the day is Saturday, Sunday or a public holiday; $WE * X_{ki}$ gives us interaction terms to take into account preference differences between business day and weekend usage; the month vector ($\mathrm{Month_j}$) includes dummies for each month for both years; the error term is given by ε and is assumed to be independent and identically distributed (iid).

The first model (Model 1, equation 4.1) uses a basic random effects approach commonly utilized for panel data. All observations are pooled but treated as independent of each other. In the field of bike-sharing this approach is used for example by Wang et al. (2018). Depending on the question and data at hand it can be useful. The estimated values in this specification are interpreted as the effect of an explanatory variable on the overall usage of the BS system. The model controls for all available explanatory variables. We control for the possibility of differences in the elasticities between weekends and business days by including interaction terms between the weekend dummy and all weather variables as well as the bike availability. We interpret interaction terms as the difference between the primary term and the interaction case. In our case, it is the difference between the effect on business days and weekends. In sum, the overall effect of the variable for weekends consists of the sum of the primary estimate of the variable plus the estimate of the interaction term (Berrington de González and Cox 2007). However, the assumption that the outcomes of any given station i at two connected moments in time are independent of each other seems highly unreasonable. Similar to any dataset having the properties of a panel some kind of group control should occur. Since our goal is to give a comparison concerning the models used for this kind of data we will use it as our most straightforward to replicate baseline. The following models consider the intergroup correlation. Following this reasoning it seems reasonable to assume that model 1 has the highest possibility of an omitted variable bias compared to model 4.2-4.4 (Bell and Jones 2015).

$$log\, Y_{i,t} = \beta_0 + \beta_k * X_{ki} + \alpha * WE + \tau_k * WE * X_{ki} + FE_i + \omega_j * \mathrm{Month_j} + \varepsilon \tag{4.2}$$

$Y_{i,t}$ indicates the number of trips generated per day at station i at day t; β_0 is a constant; X_k is a vector including the weather variables and bike availability; WE stands for the weekend dummy ; $WE * X_{ki}$ gives us interaction terms; the month vector is $\mathrm{Month_j}$; FE_i stands for fixed effects on the station level; the error term is given by ε and is assumed to be iid.

The most rigid way to control for the high inner group correlation of observations for the same station is using fixed effects (FE) on the station level (Model 2, equation 4.2). Using a FE panel model has major implications. Eliminating the between group variation most likely highly increases the degree of explanation of the whole model measured in adj. R^2. To phrase it differently, it estimates the within group variation. It also makes sure that all possible time invariant omitted variables concerning the station's spatial position are dealt with. On the other side it forces us to exclude those time invariant variables i.e. the socioeconomic structure and infrastructure in which we are interested (Hill et al. 2020; Mummolo and Peterson 2018). We obtain the most robust results regarding weather effects, seasonal effects, bike availability, and travel distance using this model. The estimated effects using FE are interpreted as the average effect within a single station over time. We are unaware of any study using this approach. The disadvantage of excluding all time invariant variables seems too major for most research questions. However, the model does an excellent job as a control for the estimates of time variant variables. If the time variant estimates do not differ systematically between the models, we can assume two main implications. The first one is that we do not have a relevant omitted variable bias in the models not using FE on station levels. The second one is that at least the results of the time variant variables are unbiased in general.

$$log\, Y_{i,t} = \beta_0 + \beta_k * X_{ki} + \gamma * V_{gi} + \alpha * WE + \tau_k * WE * X_{ki} + \rho * log\, Y_{i,\mathrm{t-lag}} + \varepsilon \tag{4.3}$$

$Y_{i,t}$ indicates the number of trips ; $Y_{i,t-lag}$ indicates the lagged demand with *lag* being the difference in time ; ρ shows the impact of the lagged bike demand; β_0 is a constant; X_k is a vector including the weather variables and bike availability; V is a vector including socioeconomic and infrastructure variables; WE stands for the weekend dummy; $WE * X_{ki}$ gives us interaction terms; the error term is given by ε and is assumed to be iid.

The final approach (Models 3 and 4, equation 4.3 with different lags) is based on the work of El-Assi et al. (2017). The basic idea is to include the lagged demand for every station. These models assume that there is some form of autocorrelation. This seems reasonable since stations mainly used for commuting are used every weekday similarly for this purpose. This model is mainly equivalent to a dynamic panel model with exogenous regressors (Sarafidis et al. 2009). The estimates are interpreted as the change in demand compared to the lagged demand caused by the change in the explanatory variables. Since we use daily data over two years our T is rather big and we can safely ignore the Nickel-Bias (Phillips and Sul 2007; Beck et al. 2014). It is unsuitable to consider the seasonal effects while including the lagged effects. Lagged effects already take the changing level in demand over time into account. Since our data structure differs from El-assi et al. we consider two different forms of lagged demand. The first one is more closely to El-Assi et al. the other is more intuitive for our data structure. El-Assi et al. did separate regressions for weekends and business days. Their approach to model the lagged demand is to use the day before so $Y_{i,\mathrm{t}-1}$ in their data structure (Model 3, equation 4.3 (with lag=1)). Their specification differentiates between business days and weekends. Lagged demand for business days is defined as trip starts at the same station the business day before. For example, it is Thursday for a Friday and for a Monday the last Friday. For weekends, the method is equivalent. Since our data structure follows more closely Wang et al. we adopted an alternative approach to the lagged demand. Our alternative approach uses trip starts at the last equivalent day of the week as lagged demand. So lagged demand can simply be defined as $Y_{i,\mathrm{t}-7}$ (Model 4 equation 4.3 (with lag=7)). Figure 4.2 shows that no matter the distribution of demand around the week, our approach should lead to more intuitive results.

4.4 Results

We use four distinct empirical approaches to make sure the results are robust. To keep the comparison of the results consistent, we will separate our discussion into thematic fields. It is important to remember that all models use a semi-log approach, and the coefficients can generally be interpreted as elasticities. We will start our analysis by looking at the temporal effects. To be more precise the seasonal effects measured in models 1 and 2 as well as the general effect of the weekend dummy in all models. In the next step, we will discuss the autocorrelation encountered in models 3 and 4. Following are the time variant effects. This part includes the effects of the weather variables. Since these variables are commonly used in the literature and are consistent overall studies, they are good indicators of the general result quality. Should our results in this area not be consistent with the literature, we could have a flawed data basis or inadequate empirical approaches. The next block includes our particular field of interest for this study the time invariant variables. In particular, the socioeconomics variables for the general population in the city districts surrounding the BS stations. We are not particularly interested in the absolute strength of the effects and will not go into detail about most coefficients. The last part compares the general usefulness and explanatory power of the different models at hand.

Table 4.2 Regression results

	Model 1	Model 2	Model 3	Model 4
Preday Lag			0.702***	
Weekday Lag				0.654***
Weekend dummy	-0.328***	-0.402***	-0.110**	0.0699
Travel Distance	-0.0577***	-0.0357***	-0.0196***	-0.0225***
Available bikes	0.0300***	0.0153***	0.0122***	0.0134***
Wind strengh	-0.0354***	-0.0336***	-0.0261***	-0.0436***
Rainfall	-0.146***	-0.121***	-0.141***	-0.120***
Air Temperature	0.0200***	0.0219***	0.00851***	0.0226***
Sunshine	0.00309***	0.00326***	0.00192***	0.00255***
Humidity	-0.00517***	-0.00680***	-0.00519***	-0.00715***
Interaction term between Weekend dummy and:				
Travel Distance	0.0550***	0.0453***	0.0116***	0.0190***
Available bikes	0.00210***	0.00330***	0.00204***	0.00165***
Wind strengh	-0.0330***	-0.0319***	-0.00358*	-0.0238***
Rainfall	-0.277***	-0.270***	-0.303***	-0.327***
Air Temperature	0.0125***	0.0132***	0.00423***	0.00467***
Sunshine	-0.000730	0.000172	0.00229***	0.000688*
Humidity	-0.00315***	-0.00254***	-0.00152***	-0.00340***
Metro	-0.0401***		-0.0110***	-0.0128***
City train	0.0368***		0.00999***	0.0117***
BS station	-0.107***		-0.0421***	-0.0501***
Bus	0.00956***		0.00236***	0.00301***
Average Income	0.00505***		0.00135***	0.00152***

Table 4.2 Regression results

cars per capita	-5.948***		-1.676***	-1.981***
Population density	0.0000124***		0.00000505***	0.00000588***
Foreigner rate	0.0722***		0.0213***	0.0244***
Voting rate	0.0618***		0.0180***	0.0208***
female share	0.109***		0.0334***	0.0391***
Age <18	-0.0358***		-0.0110***	-0.0118***
Age 18-24	-0.188***		-0.0539***	-0.0627***
Age 25-29	0.224***		0.0629***	0.0745***
Age 50-64	0.0742***		0.0204***	0.0249***
Age >=65	0.00968***		0.00102	0.00153
Monthly Dummies	YES	YES	NO	NO
Station FE	NO	YES	NO	NO
Constant	-7.548***	3.363***	-1.797***	-1.941***
N	127035	127035	126623	125593
adj. R-sqr	0.574	0.816	0.804	0.780
AIC	263222.3	156608.1	163531.8	176082.1

Notes: Our depend variable are the logged daily trips per station.
The results show highly significant and consistent effects over all model. Confirming our general assumptions as well as previous findings in the literature.
* p < 0.05, ** p < 0.01, *** p < 0.001

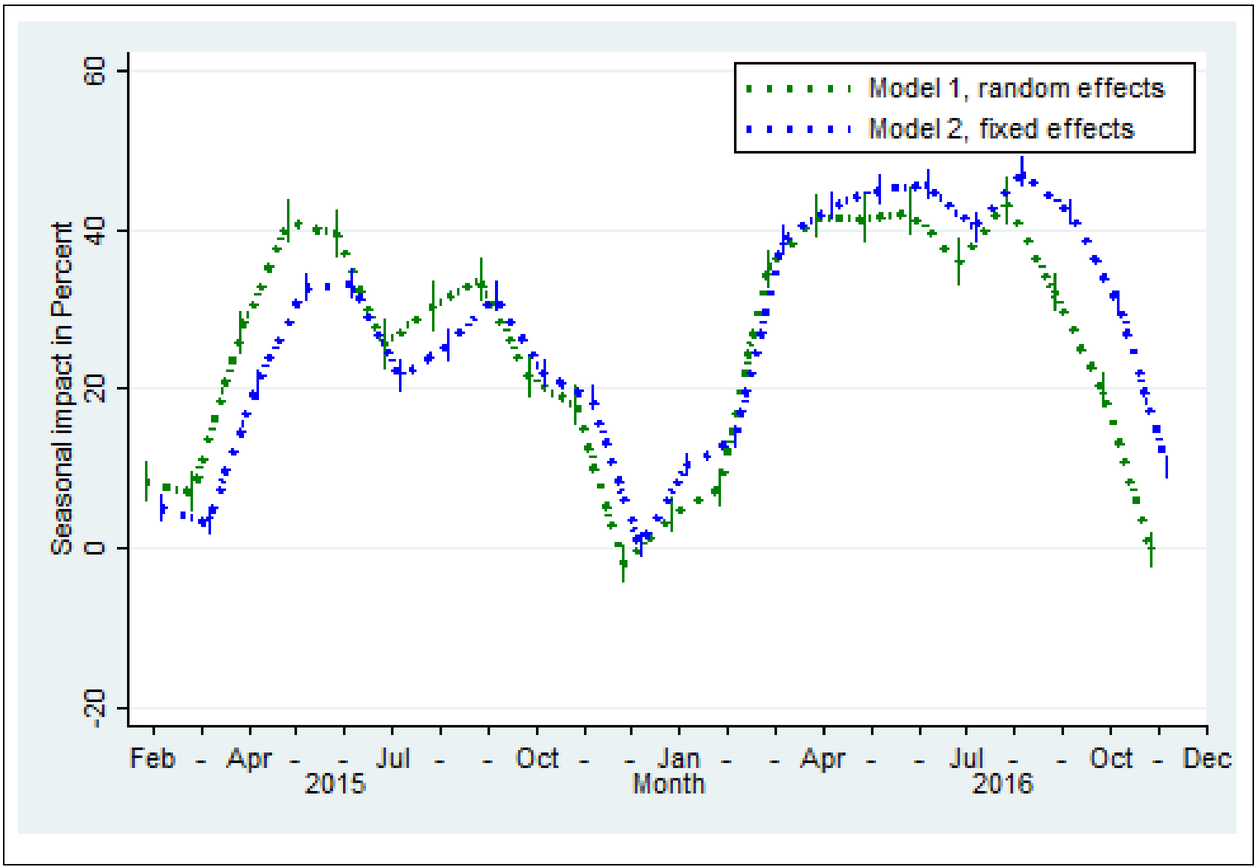
4.4.1 Temporal effects

In a first step, we will discuss the seasonal effects measured as monthly dummies in Model 1 and 2. Fournier et al. (2017) have shown that demand follows a sin curve over the year for several BS systems. Both models show that behavior (Figure 4.3[11]), which means that BS usage is highest in summer and decreases until it is at its lowest in January. There is an apparent dip in demand around July. At this time, Hamburg has summer vacation time for schools as well as no course time at the university, leading to an overall reduced travel demand in the city. Summer is also the primary vacation time for most Germans. It shows that if the impact of weather is accounted for, which should be the main difference between seasons, there is still a considerable seasonal effect. Differences in tourism as well as differences in public transport behavior could trigger this effect. One of the most essential factors in the summer month are the summer school holidays. Kashfi et al. (2015) have shown that public transport usage, in general, is highly negatively impacted by holiday seasons. The usage reduction is established through reduced commuting demand. This indicates that any model not using lagged demand as an explanatory variable should control for seasonal effects regardless of other explanatory variables.

Model 1, 2 and 3 find a significant negative effect of the weekend dummy on the estimated demand, as shown in Table 4.2. This effect is expected due to the descriptive results seen in Figure 4.2. Like mentioned earlier, these effects are consistent with our expectations. There is less commuting to work on weekends and most of the demand on weekends is leisure driven. The coefficient is not significant in Model 4. This seems reasonable since the model includes the lagged demand from t-7. It would be unusual if the change in demand would be different in percent from one week to the next because it is a weekend day compared to a weekday. In other words, the decreased demand on weekends is already captured. The following sections will discuss the differences between the estimates of the other explanatory variables and their corresponding interaction terms on weekends.

[11] January of 2015 is our baseline month. It was omitted from the regression and respectably not included in the Graph.

Figure 4.3 Estimated monthly effects for Model 1 and 2



4.4.2 Autocorrelation

Models 3 and 4 include an autoregressive term ρ that captures the lagged demand's effects on demand. El-Assi et al. have estimated a similar term as Model 3 separately for a weekend and a business day model. They find an autocorrelation of 0.43 for business days and 0.418 for weekends. We find a much higher autocorrelation of 0.702 for Model 3, which uses the pre day demand similar to El-Assi et al. Our Model 4 using the demand of the last equivalent day of the week shows a slightly lower autocorrelation of 0.654. Still, both models show that our dataset has a much higher BS station usage consistency than other studies. This high degree of autocorrelation indicates that not controlling for either the past demand of a station or a general FE term on the station level would lead to biased estimates.

4.4.3 Time variant effects

All time variant variables are included in every model. As can be seen in table 4.2, the estimated effects are consistent over all empirical specifications. To be precise, all estimates are highly significant and have the same direction independent of the model. The interaction terms are also mostly consistent. The estimates for sunshine on weekends are not significant in Models 1 and 2. The rest of the interaction term estimates are consistent. Exemplary we will interpret the effects of Model 2 to give an idea about the estimated effect strength of the explanatory variables. Model 2 is chosen since it should have the most unbiased estimates and has the overall highest explanatory power of all models.

Stations with a higher average travel distance show less frequent usage. This is consistent with results found in the literature that distance has a negative impact on commuting by bike (Heinen et al. 2010). For every additional km of average travel distance, the trip generation decreases for 3.5 percent. Since the interaction shows an opposite effect it is hard to give a precise effect of travel distance on demand on weekends. In Model 2 the interaction term in absolute terms is bigger than the initial effect indication a small positive effect of travel distance on demand on weekends. However, for Models 1, 3 and 4 the interaction term is smaller in absolute terms. On average there seems to be very little absolute effect of average travel distance on station demand on weekends. This result seems intuitive since the share of trips as actual leisure activity should be higher and so trips should be longer on average on weekends compared to weekdays.

As expected, the real number of available bicycles has a positive impact on station usage. For every available bicycle the usage increases by 1,5 percent. So far, studies only used the docking capacity to proxy the possibility to get a free bicycle at a station. Our results close this gap in the literature and confirm the results obtained using proxies. The effects of the proxies in the literature are also positive so our results with regard to the measured entity are consistent. This effect is by a small margin higher on weekends. Although, we confirm the general results in the literature, the estimates should not be compared lightly with future studies. These estimates were calculated for a BS system with real time user information about bicycle availability and could differ for other types of systems. However, even in systems without real time information for the users through apps or websites, the real availability should be a better indicator than the amount of docking stations.

Similar to older studies (Rudloff and Lackner 2014; Ashqar et al. 2019), variables indicating weather conditions are highly significant. Factors associated with bad weather have a negative impact, e.g., rain, humidity and wind strength. Sunshine per hour and air temperature have an expected positive impact. For weekends, which should indicate mostly leisure demand, we find higher elasticities. We find that in particular rainfall has a high negative effect on the demand. Between the models the effects are rather similar in a range between 1.2 and 1.41 percent less demand per mm precipitation. Since the interaction terms are at least twice the size of the base effect of rain, the overall effect in weekends is three times as high as on weekdays.

4.4.4 Time invariant effects

Finally, we are going to discuss the explanatory variables included in the vector V_g. Vector V_g consists of the public transport as well as the socioeconomic variables. Model 2 as explained does not contain the vector V_g. The results of the variables included in vector V_g are highly consistent. The only difference regarding significance is the age group of 65 years and older. This variable is not significant in models 3 and 4.

Since the literature is inconclusive, we have no clear expectations for the effects of different public transportation forms. Metro stations show a negative effect on BS, while city train stations show a positive effect. A reason for this difference could be the different spatial distribution of both station types. The average distance between metro station stops is smaller than the distance between city train stops. In addition, city trains are available in more extensive parts of the city[12]. Bus stations have a positive effect. However, the effects are relatively small in all models, and since the absolute amount of stations in Hamburg is rather considerable, it is hard to argue that the BS station placement should be influenced by bus station availability. On the other side, the variable should not be ignored in future studies since it has a significant effect and would lead to an omitted variable bias in the whole model if dropped. In addition, we do not control for the number of bus lines that hold at a given station. Other BS stations in close proximity have a negative effect on the demand of a BS station. This result seems unreasonable at first. We showed that travel distance negatively affects demand, so close stations should increase it. However, since all BS stations captured in this variable are in very close proximity, it seems unreasonable to take a bicycle for this distance. On the other hand, other BS stations could eat up parts of the demand. Indicating that areas with many BS stations are somewhat oversaturated. The whole BS usage in the city could most likely be increased if station clusters are avoided.

Most of our assumptions regarding socioeconomic variables are found in the literature results of survey studies. We expect that these results also hold for our German case study. Most of the general results regarding socioeconomic variables are consistent with the literature and our expectations. As expected from the meta-study results by Eren and Uz (2020), we find a positive impact of the population's average income on BS and a negative impact between the car ownership rate and BS. Population density is included in most studies using historic data, and similar to those studies, we find a positive effect of population density on BS. This seems very reasonable since areas with more potential users should have a higher usage rate. Like bus availability, it is less interesting for policymakers and more important for the quality of the model. One could argue that an extremely high population rate can lead to a transportation

[12] The whole southern part of Hamburg is not equipped with metro stations.

collapse. However, we are only comparing relative densities in Hamburg and cannot make any informed prediction regarding the effects of absolute population densities.

We find positive effects of the voting rate on BS usage. We find positive effects of the voting rate on BS usage. Voting rates are an essential indicator for social participation and the interest in the structure of the neighborhood (DiPasquale and Glaeser 1999; Ahlfeldt and Maennig 2015). In their meta-study Fishman (2016) finds that BS users' ethnicity strongly differs from the general population in many case studies. In most cases, white people are highly overrepresented. However, they state that “Many BSPs [Bike-Sharing Providers] do not cover the full residential area of the city, and this may offer an explanation for the demographic biases of bikeshare users.” (Fishman 2016, p. 100). Our study implicitly controls for precisely this fact. Even though we cannot know the exact foreigner rate using BS in Hamburg is we can conclude that districts with higher foreigner rates have a higher demand for BS than districts with lower foreigner rates.

For the effect of social structure, our main interest lies in the effects of gender and age distribution. Since voter groups are often focused in these categories, they are often essential target factors for policymakers (Holman et al. 2015; Iyer et al. 2017). We find contradictory results for the effects of gender compared to most studies since the female share in a district increases BS usage. However, most BS studies have been conducted in countries with lower cycling rates. As shown before, countries with relatively high cycling rates show a higher demand by females compared to males. To confirm these results, similar BS studies in for example the Netherlands, would be helpful.

For age, we find strong non-linear effects for BS usage. The biggest age group of 30-49 years old is our baseline, so all other groups are in comparison. The only groups having a negative effect are children (below 18 years) and 18-24 years olds. For our observation years, the age groups have a high overlap with the cohorts defined by Wang et al. (2018). The group of 18-24 olds is equivalent to “younger Millennials”. Children as well as “younger Millennials” often receive high subventions for public transport and/or possess their own bicycles. Most pupils and students in Germany have a public transport pass at a reduced rate. Looking at “mid Millennials”, that is, the age group of 25-29 year olds shows the highest positive impact on BS usage. The effects for older people (i.e., aged 50-64 years) are significant and positive in all models, however relatively small. The group of retirees (65 years and older) shows inconsistent but always non-negative effects.

These results show that the responding share of age groups in a district is essential. The results are mostly consistent with studies using survey data. Should the goal of policymakers be to maximize, the BS usage following remarks could be of importance. When considering the age distribution in a district, policymakers should not target very young users who have a relatively low car per capita rate. Instead, the main target could be the female share of the young workforce. This group seems to have a high demand for very cheap public transport in exchange for physical activity. Especially aiming at this group could further decrease car and traditional public transport demand in cities if the BS station availability is sufficient.

4.4.5 Model fitting and robustness

We ensure that our general results are robust using different models that include a comprehensive set of spatial and temporal control variables. We can show that nearly all variables have consistent effects in different specifications even though the size of the estimates differs between the models. This is partly due to differences in the interpretation of the estimates. No variables have contradictory results in different models. With a high degree of certainty, we can say that our results are robust independent of the model specification. Compared to other studies, our models show a remarkable degree of explanatory power. The lowest explanatory power is given in the baseline random effects model with an adj. R^2 of 57 percent. Models 2, 3 and 4 can explain between 78 and 82 percent of the variance in the model. In comparison, the study, which is most similar to model 3 (El-Assi et al. 2017), can only explain 65 percent of the manifold variance.

4.5 Conclusion

Our main novelty is being the first study to calculate the effects of the socioeconomic structure surrounding a BS station on BS demand. The effects are mostly consistent with other studies using survey or usage data. As well as the effects found regarding bicycle usage in general (Eren and Uz 2020; Rudloff and Lackner 2014). However, it also shows that the results strongly differ among regions since, for example, only countries with a relatively high bicycle rate show more female than male cyclists.

In addition, we developed a method to calculate the real-time bicycle availability at the station level. This is a vast improvement to the established method of using the total amount of docking places at the station level (El-Assi et al. 2017). Using this additional information, future research could estimate excess demand. This would lead to a better understanding of the actual demand at the station level. At the moment, this excess demand could be a concern regarding an omitted variable bias. Using FE at the station level does not eliminate this bias since it is time variant.

In summary, our study adds to the literature in manifold ways. We are the first empirical BS study using a German dataset. Our results are consistent with the effects found in the literature for the basic control variables. These variables include the effects of weather, lagged bicycle demand, seasonal effects, travel distance and infrastructure (Fishman 2016).

All data used is openly available and proved to be very useful concerning standard analyses regarding bike-sharing. Building on these findings, we encourage other researchers to build on our methods and the data at hand.

Our research has some limitations, mainly through unavailable information. We are unable to control for education, even so Santos et al. (2013) have shown that it has an impact on the modal share. In addition, we lack the number of docking places per station so we cannot directly compare the results from other studies to our dataset. It would be most interesting to estimate docking stations as a proxy in our model since they are not needed to park a bike even though they exist. Most critically, the timeframe makes it impossible to find price elasticities concerning other forms of transportation. The price structure of StadtRAD did not change in the timeframe of the study.

5. Bibliography

Addams, Jane (1902): The Housing Problem in Chicago. In *The ANNALS of the American Academy of Political and Social Science* 20 (1), pp. 99–107. DOI: 10.1177/000271620202000108.

Ahlfeldt, Gabriel M. (2011): Blessing or curse? Appreciation, amenities and resistance to urban renewal. In *Regional Science and Urban Economics* 41 (1), pp. 32–45.

Ahlfeldt, Gabriel M.; Maennig, Wolfgang (2010a): Impact of sports arenas on land values: evidence from Berlin. In *Ann Reg Sci* 44 (2), pp. 205–227. DOI: 10.1007/s00168-008-0249-4.

Ahlfeldt, Gabriel M.; Maennig, Wolfgang (2010b): Stadium Architecture and Urban Development from the Perspective of Urban Economics. In *Int J Urban & Regional Res* 34 (3), pp. 629–646. DOI: 10.1111/j.1468-2427.2010.00908.x.

Ahlfeldt, Gabriel M.; Maennig, Wolfgang (2010c): Substitutability and Complementarity of Urban Amenities: External Effects of Built Heritage in Berlin. In *Real Estate Econ* 38 (2), pp. 285–323. DOI: 10.1111/j.1540-6229.2010.00268.x.

Ahlfeldt, Gabriel M.; Maennig, Wolfgang (2013): External Productivity and Utility Effects of City Airports. In *Regional Studies* 47 (4), pp. 508–529. DOI: 10.1080/00343404.2011.581652.

Ahlfeldt, Gabriel M.; Maennig, Wolfgang (2015): Homevoters vs. leasevoters. A spatial analysis of airport effects. In *Journal of Urban Economics* 87, pp. 85–99. DOI: 10.1016/j.jue.2015.03.002.

Ahlfeldt, Gabriel M.; Maennig, Wolfgang; Richter, Felix J. (2017): Urban renewal after the Berlin Wall. A place-based policy evaluation. In *J Econ Geogr* (17), pp. 129–156. DOI: 10.1093/jeg/lbw003.

Alberts, Heike C. (2009): Berlin's Failed Bid to Host the 2000 Summer Olympic Games: Urban Development and the Improvement of Sports Facilities. In *International Journal of Urban and Regional Research* 33 (2), pp. 502–516. DOI: 10.1111/j.1468-2427.2009.00847.x.

Anderson, Soren T.; West, Sarah E. (2006): Open space, residential property values, and spatial context. In *Regional science & urban economics* 36 (6), pp. 773–789. DOI: 10.1016/j.regsciurbeco.2006.03.007.

Andersson, Henrik; Swärdh, Jan-Erik; Ögren, Mikael (2015): Traffic noise effects of property prices: Hedonic estimates based on multiple noise indicators.

Anselin, Luc (1988): Spatial Econometrics: Methods and Models. Dordrecht: Springer (Studies in Operational Regional Science, 4).

Anselin, Luc (2010): Thirty years of spatial econometrics. In *Papers in Regional Science* 89 (1), pp. 3–25. DOI: 10.1111/j.1435-5957.2010.00279.x.

Anselin, Luc; Florax, Raymond J. G. M.; Rey, Sergio J. (Eds.) (2004): Advances in Spatial Econometrics. Berlin, Heidelberg: Springer Berlin Heidelberg (Advances in Spatial Science).

Ashqar, Huthaifa I.; Elhenawy, Mohammed; Rakha, Hesham A. (2019): Modeling bike counts in a bike-sharing system considering the effect of weather conditions. In *Case Studies on Transport Policy* 7 (2), pp. 261–268. DOI: 10.1016/j.cstp.2019.02.011.

Balli, Hatice Ozer; Sørensen, Bent E. (2013): Interaction effects in econometrics. In *Empir Econ* 45 (1), pp. 583–603. DOI: 10.1007/s00181-012-0604-2.

Baranzini, Andrea; Ramirez, Jose (2005): Paying for quietness: The impact of noise on Geneva rents. In *Urban Studies (Routledge)* 42 (4), pp. 633–646.

Basner, Mathias; Mueller, Uwe; Elmenhorst, Eva-Maria (2011): Single and Combined Effects of Air, Road, and Rail Traffic Noise on Sleep and Recuperation. In *SLEEP* 34 (1), pp. 11–23.

Bauer, Thomas; Göhlmann, Silja; Sinning, Mathias (2007): Gender differences in smoking behavior. In *Health economics* 16 (9), pp. 895–909. DOI: 10.1002/hec.1259.

Beck, Nathaniel L.; Katz, Jonathan N.; Mignozzetti, Umberto G. (2014): Of Nickell Bias and its Cures: Comment on Gaibulloev, Sandler, and Sul. In *Polit. anal.* 22 (02), pp. 274–278. DOI: 10.1093/pan/mpu004.

Beimer, Waldemar; Maennig, Wolfgang (2017): Noise effects and real estate prices. A simultaneous analysis of different noise sources. In *Transportation Research Part D: Transport and Environment* 54, pp. 282–286. DOI: 10.1016/j.trd.2017.05.010.

Beimer, Waldemar; Maennig, Wolfgang (2020): On the Price Gap between Single Family Houses and Apartments. In *Journal of Housing Economics* 49 (1), p. 101686. DOI: 10.1016/j.jhe.2020.101686.

Bell, Andrew; Jones, Kelvyn (2015): Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. In *PSRM* 3 (1), pp. 133–153. DOI: 10.1017/psrm.2014.7.

Berrington de González, Amy; Cox, D. R. (2007): Interpretation of interaction: A review. In *Ann. Appl. Stat.* 1 (2), pp. 371–385. DOI: 10.1214/07-AOAS124.

Blinder, Alan S. (1973): Wage discrimination. Reduced form and structural estimates. In *Journal of human resources : JHR* 8 (4), pp. 436–455.

Bögel, Jan; Mesecke, Fabian; Schult, Clemens; Becker, Ole (2014): Gebäude- und Wohnungsbestand in Deutschland. Erste Ergebnisse der Gebäude- und Wohnungszählung 2011. Edited by Statistische Ämter des Bundes und der Länder.

Bogin, Alexander; Doerner, William; Larson, William (2018): Local House Price Dynamics. New Indices and Stylized Facts. In *Real Estate Econ* 58 (304), p. 933. DOI: 10.1111/1540-6229.12233.

Bolitzer, B.; Netusil, N.R (2000): The impact of open spaces on property values in Portland, Oregon. In *Journal of Environmental Management* 59 (3), pp. 185–193. DOI: 10.1006/jema.2000.0351.

Brambor, Thomas; Clark, William Roberts; Golder, Matt (2006): Understanding Interaction Models. Improving Empirical Analyses. In *Polit. anal.* 14 (01), pp. 63–82. DOI: 10.1093/pan/mpi014.

Brandt, S.; Maennig, Wolfgang (2011): Road noise exposure and residential property prices. Evidence from Hamburg. In *Transportation research / D* 16 (1), pp. 23–30.

BWI (2021): Mobilität in Hamburg. Mobilitäts-Quartett. Behörde für Wirtschaft und Innovation.

Campbell, John; Cocco, João (2005): How Do House Prices Affect Consumption? Evidence From Micro Data. Cambridge, MA: National Bureau of Economic Research.

Can, Ayse; Megbolugbe, Isaac F. (1997): Spatial dependence and house price index construction. In *The Journal of Real Estate Finance and Economics* 14.

Cassel, Eric; Mendelsohn, Robert (1985): The choice of functional forms for hedonic price equations. Comment. In *Journal of Urban Economics* 18 (2), pp. 135–142. DOI: 10.1016/0094-1190(85)90012-9.

Caulfield, Brian; O'Mahony, Margaret; Brazil, William; Weldon, Peter (2017): Examining usage patterns of a bike-sharing scheme in a medium sized city. In *Transportation Research Part A: Policy and Practice* 100, pp. 152–161. DOI: 10.1016/j.tra.2017.04.023.

Ceccato, Riccardo; Diana, Marco (2018): Substitution and complementarity patterns between traditional transport means and car sharing: a person and trip level analysis. In *Transportation* 1927, p. 207. DOI: 10.1007/s11116-018-9901-8.

Chang, Justin S.; Kim, Dae-Jin (2013): Hedonic estimates of rail noise in Seoul. In *Transportation Research Part D: Transport and Environment* 19, pp. 1–4. DOI: 10.1016/j.trd.2012.11.002.

Chang, Shan; Song, Rui; He, Shiwei; Qiu, Guo (2018): Innovative Bike-Sharing in China: Solving Faulty Bike-Sharing Recycling Problem. In *Journal of Advanced Transportation* 2018 (1), pp. 1–10. DOI: 10.1155/2018/4941029.

Chen, Shang-Yu (2016): Using the sustainable modified TAM and TPB to analyze the effects of perceived green value on loyalty to a public bike system. In *Transportation Research Part A: Policy and Practice* 88, pp. 58–72. DOI: 10.1016/j.tra.2016.03.008.

Chen, Tao; Harding, John P. (2016): Changing Tastes: Estimating Changing Attribute Prices in Hedonic and Repeat Sales Models. In *J Real Estate Finan Econ* 52 (2), pp. 141–175. DOI: 10.1007/s11146-015-9497-0.

Chinloy, Peter; Das, Prashant K.; Wiley, Jonathan A. (2014): Houses and apartments. Similar assets, different financials. In *Journal of Real Estate Research* 36 (4), pp. 409–433.

Cidell, Julie (2015): The role of major infrastructure in subregional economic development: an empirical study of airports and cities. In *J Econ Geogr* 15 (6), pp. 1125–1144. DOI: 10.1093/jeg/lbu029.

Clark, Julie; Curl, Angela (2016): Bicycle and Car Share Schemes as Inclusive Modes of Travel? A Socio-Spatial Analysis in Glasgow, UK. In *SI* 4 (3), p. 83. DOI: 10.17645/si.v4i3.510.

Connolly, Marie (2008): Here Comes the Rain Again: Weather and the Intertemporal Substitution of Leisure. In *Journal of labor economics* 26 (1), pp. 73–100. DOI: 10.1086/522067.

Cools, Mario; Moons, Elke; Creemers, Lieve; Wets, Geert (2010): Changes in Travel Behavior in Response to Weather Conditions. In *Transportation Research Record* 2157 (1), pp. 22–28. DOI: 10.3141/2157-03.

Corrado, Luisa; Fingleton, Bernard (2012): WHERE IS THE ECONOMICS IN SPATIAL ECONOMETRICS?*. In *Journal of Regional Science* 52 (2), pp. 210–239. DOI: 10.1111/j.1467-9787.2011.00726.x.

Costanigro, Marco; McCluskey, Jill J.; Mittelhammer, Ron C. (2007): Segmenting the Wine Market Based on Price. Hedonic Regression when Different Prices mean Different Products. In *J Agricultural Economics* 58 (3), pp. 454–466. DOI: 10.1111/j.1477-9552.2007.00118.x.

Coulson, N. Edward; Dalton, Maurice (2010): Temporal and ethnic decompositions of homeownership rates: Synthetic cohorts across five censuses. In *Journal of Housing Economics* 19 (3), pp. 155–166. DOI: 10.1016/j.jhe.2010.03.001.

Cruz, Carlos Oliveira; Sarmento, Joaquim Miranda (2020): Traffic forecast inaccuracy in transportation: a literature review of roads and railways projects. In *Transportation* 47 (4), pp. 1571–1606. DOI: 10.1007/s11116-019-09972-y.

Damant-Sirois, Gabriel; El-Geneidy, Ahmed M. (2015): Who cycles more? Determining cycling frequency through a segmentation approach in Montreal, Canada. In *Transportation Research Part A: Policy and Practice* 77, pp. 113–125. DOI: 10.1016/j.tra.2015.03.028.

DeMaio, Paul (2009): Bike-sharing: History, Impacts, Models of Provision, and Future. In *JPT* 12 (4), pp. 41–56. DOI: 10.5038/2375-0901.12.4.3.

Demombynes, Gabriel; Trommlerová, Sofia Karina (2016): What has driven the decline of infant mortality in Kenya in the 2000s? In *Economics and human biology* 21, pp. 17–32. DOI: 10.1016/j.ehb.2015.11.004.

den Boer, Eelco; Schroten, Arno (2007): Traffic noise reductionin Europe. Health effects, social costs andtechnical and policy options toreduce road and rail traffic noise. Edited by CE Delft. Solutions for environment, economy and technology.

Dermisi, Sofia V.; McDonald, John F. (2010): Selling Prices/Sq. Ft. of Office Buildings in Downtown Chicago—How Much Is It Worth to Be an Old But Class A Building? In *Journal of Real Estate Research* 32 (1), pp. 1–22.

Deutsche Bahn AG (2017): Call A Bike. Available online at http://data.deutschebahn.com/dataset/data-call-a-bike.

Diamond, Rebecca; McQuade, Tim; Qian, Franklin (2019): The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco. In *American Economic Review* 109 (9), pp. 3365–3394. DOI: 10.1257/aer.20181289.

Dieleman, F. M.; Schouw, R. J. (1989): Divorce, mobility and housing demand. In *Eur J Population* 5 (3), pp. 235–252. DOI: 10.1007/BF01796818.

Dill, Jennifer; Carr, Theresa (2003): Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them. In *Transportation Research Record: Journal of the Transportation Research Board* 1828, pp. 116–123. DOI: 10.3141/1828-14.

DiPasquale, Denise; Glaeser, Edward L. (1999): Incentives and Social Capital: Are Homeowners Better Citizens? In *Journal of Urban Economics* 45 (2), pp. 354–384. DOI: 10.1006/juec.1998.2098.

DWD Climate Data Center (2018): Historische stündliche Stationsmessungen für Deutschland. Deutscher Wetter Dienst. Available online at https://www.dwd.de.

El-Assi, Wafic; Salah Mahmoud, Mohamed; Nurul Habib, Khandker (2017): Effects of built environment and weather on bike sharing demand. A station level analysis of commercial bike sharing in Toronto. In *Transportation* 44 (3), pp. 589–613. DOI: 10.1007/s11116-015-9669-z.

Elhorst, J. Paul (2012): Dynamic spatial panels: models, methods, and inferences. In *J Geograph Syst* 14 (1), pp. 5–28. DOI: 10.1007/s10109-011-0158-4.

Englander, Gabriel (2021): Information and Spillovers from Targeting Policy in Peru's Anchoveta Fishery. In *SSRN Journal*. DOI: 10.2139/ssrn.3807560.

Eren, Ezgi; Uz, Volkan Emre (2020): A review on bike-sharing: The factors affecting bike-sharing demand. In *Sustainable Cities and Society* 54, p. 101882. DOI: 10.1016/j.scs.2019.101882.

Eurostat (2015): People in the EU. Who are we and how do we live? 2015 edition. Luxembourg: EUR-OP (Statistical books).

Faghih-Imani, Ahmadreza; Eluru, Naveen (2016): Determining the role of bicycle sharing system infrastructure installation decision on usage. Case study of montreal BIXI system. In *Transportation Research Part A: Policy and Practice* 94, pp. 685–698. DOI: 10.1016/j.tra.2016.10.024.

Faghih-Imani, Ahmadreza; Eluru, Naveen; El-Geneidy, Ahmed M.; Rabbat, Michael; Haq, Usama (2014): How land-use and urban form impact bicycle flows. Evidence from the bicycle-sharing system (BIXI) in Montreal. In *Journal of Transport Geography* 41, pp. 306–314. DOI: 10.1016/j.jtrangeo.2014.01.013.

Faghih-Imani, Ahmadreza; Hampshire, Robert; Marla, Lavanya; Eluru, Naveen (2017): An empirical analysis of bike sharing usage and rebalancing. Evidence from Barcelona and Seville. In *Transportation Research Part A: Policy and Practice* 97, pp. 177–191.

Fik, Timothy J.; Ling, David C.; Mulligan, Gordon F. (2003): Modeling Spatial Variation in Housing Prices. A Variable Interaction Approach. In *Real Estate Econ* 31 (4), pp. 623–646. DOI: 10.1046/j.1080-8620.2003.00079.x.

Fishman, Elliot (2016): Bikeshare. A Review of Recent Literature. In *Transport Reviews* 36 (1), pp. 92–113. DOI: 10.1080/01441647.2015.1033036.

Fishman, Elliot; Washington, Simon; Haworth, Narelle (2013): Bike Share. A Synthesis of the Literature. In *Transport Reviews* 33 (2), pp. 148–165. DOI: 10.1080/01441647.2013.775612.

Fournier, Nicholas; Christofa, Eleni; Knodler, Michael A. (2017): A sinusoidal model for seasonal bicycle demand estimation. In *Transportation Research Part D: Transport and Environment* 50, pp. 154–169. DOI: 10.1016/j.trd.2016.10.021.

Fuller, Daniel; Gauvin, Lise; Kestens, Yan; Daniel, Mark; Fournier, Michel; Morency, Patrick; Drouin, Louis (2013): Impact evaluation of a public bicycle share program on cycling. A case example of BIXI in Montreal, Quebec. In *American journal of public health* 103 (3), e85-92. DOI: 10.2105/AJPH.2012.300917.

Garrison, William L. (1959): SPATIAL STRUCTURE OF THE ECONOMY: I. In *Annals of the Association of American Geographers* 49 (2), pp. 232–239. DOI: 10.1111/j.1467-8306.1959.tb01610.x.

Geoportal-Hamburg.de (2018): Einzugsbereich von HVV Haltestellen. Edited by Landesbetrieb Geoinformation und Vermessung. Available online at https://geoportal-hamburg.de/Geoportal/geo-online/#.

Gibbons, Steve; Overman, Henry G. (2012): Mostly pointless spatial econometrics? In *Journal of Regional Science* 52 (2), pp. 172–191.

Gomber, Peter; Haferkorn, Martin (2015): High Frequency Trading. In D.B.A.Mehdi Khosrow-Pour (Ed.): Encyclopedia of Information Science and Technology, Third Edition: IGI Global, pp. 1–9.

González, Felipe; Melo-Riquelme, Carlos; Grange, Louis de (2016): A combined destination and route choice model for a bicycle sharing system. In *Transportation* 43 (3), pp. 407–423. DOI: 10.1007/s11116-015-9581-6.

Gurran, Nicole; Phibbs, Peter (2015): Are Governments Really Interested in Fixing the Housing Problem? Policy Capture and Busy Work in Australia. In *Housing Studies* 30 (5), pp. 711–729. DOI: 10.1080/02673037.2015.1044948.

Hamburger Verkehrsverbund (2021): hvv-Preiserhoehung. HVV. Available online at https://www.hamburg.de/hvv/4096344/hvv-preiserhoehung/, checked on 9/22/2021.

Hansen, Bruce E. (2005): Challenges for econometric model selection (21).

Hartman, Chester; Robinson, David (2003): Evictions: The hidden housing problem. In *Housing Policy Debate* 14 (4), pp. 461–501. DOI: 10.1080/10511482.2003.9521483.

Heinen, Eva; van Wee, Bert; Maat, Kees (2010): Commuting by Bicycle: An Overview of the Literature. In *Transport Reviews* 30 (1), pp. 59–96. DOI: 10.1080/01441640903187001.

Henderson, Kirsten A.; Anand, Madhur; Bauch, Chris T. (2013): Carrot or stick? Modelling how landowner behavioural responses can cause incentive-based forest governance to backfire. In *PloS one* 8 (10), e77735. DOI: 10.1371/journal.pone.0077735.

Herath, Shanaka; Maier, Gunther (2013): Local particularities or distance gradient. In *J of Eur Real Est Research* 6 (2), pp. 163–185. DOI: 10.1108/JERER-10-2011-0022.

Hill, Terrence D.; Davis, Andrew P.; Roos, J. Micah; French, Michael T. (2020): Limitations of Fixed-Effects Models for Panel Data. In *Sociological Perspectives* 63 (3), pp. 357–369. DOI: 10.1177/0731121419863785.

Hiller, Norbert (2015): The relative importance of structural and locational effects on apartment price variations in a medium-sized German city. In *Review of regional research* 35 (1), pp. 73–102.

Holman, Mirya R.; Schneider, Monica C.; Pondel, Kristin (2015): Gender Targeting in Political Advertisements. In *Political Research Quarterly* 68 (4), pp. 816–829. DOI: 10.1177/1065912915605182.

Hörcher, Daniel; Tirachini, Alejandro (2021): A review of public transport economics. In *Economics of Transportation* 25, p. 100196. DOI: 10.1016/j.ecotra.2021.100196.

Ioannidis, John P. A.; Stanley, T. D.; Doucouliagos, Hristos (2017): The Power of Bias in Economics Research. In *Economic Journal* 127 (605), F236-F265. DOI: 10.1111/ecoj.12461.

Isard, Walter (1954): Location Theory and Trade Theory: Short-Run Analysis. In *The Quarterly Journal of Economics* 68 (2), p. 305. DOI: 10.2307/1884452.

Iyer, Pramod; Yazdanparast, Atefeh; Strutton, David (2017): Examining the effectiveness of WOM/eWOM communications across age-based cohorts: implications for political marketers. In *JCM* 34 (7), pp. 646–663. DOI: 10.1108/JCM-11-2015-1605.

Jann, Ben (2008): The Blinder-Oaxaca decomposition for linear regression models. In *STATA JOURNAL* 8 (4), pp. 453–479.

Jun, Hee-Jung (2016): The spatial dynamics of neighborhood change. Exploring spatial dependence in neighborhood housing value change. In *Housing Studies* 32 (6), pp. 717–741. DOI: 10.1080/02673037.2016.1228852.

Kashfi, Syeed Anta; Bunker, Jonathan M.; Yigitcanlar, Tan (2015): Understanding the effects of complex seasonality on suburban daily transit ridership. In *Journal of Transport Geography* 46, pp. 67–80. DOI: 10.1016/j.jtrangeo.2015.05.008.

Kennedy, Peter E. (2002): Sinning in the Basement: What are the Rules? The Ten Commandments of Applied Econometrics. In *J Economic Surveys* 16 (4), pp. 569–589. DOI: 10.1111/1467-6419.00179.

Keskin, Berna; Watkins, Craig (2016): Defining spatial housing submarkets. Exploring the case for expert delineated boundaries. In *Urban Studies* 54 (6), pp. 1446–1462. DOI: 10.1177/0042098015620351.

Kim, Kyoungok (2018): Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. In *Journal of Transport Geography* 66, pp. 309–320. DOI: 10.1016/j.jtrangeo.2018.01.001.

Lau, Lawrence J. (1986): Chapter 26 Functional forms in econometric model building. In, vol. 3: Elsevier (Handbook of Econometrics), pp. 1515–1566.

LeSage, James P. (2008): An Introduction to Spatial Econometrics. In *rei* (123), pp. 19–44. DOI: 10.4000/rei.3887.

LeSage, James P.; Charles, Joni S. (2008): Using home buyers' revealed preferences to define the urban–rural fringe. In *J Geograph Syst* 10 (1), pp. 1–21. DOI: 10.1007/s10109-007-0055-z.

Leyden, Kevin M. (2003): Social Capital and the Built Environment: The Importance of Walkable Neighborhoods. In *American journal of public health* 93 (9), pp. 1546–1551. DOI: 10.2105/AJPH.93.9.1546.

Lhila, Aparna; Long, Sharon (2012): What is driving the black-white difference in low birthweight in the US? In *Health economics* 21 (3), pp. 301–315. DOI: 10.1002/hec.1715.

Litman, Todd (2004): Transit Price Elasticities and Cross - Elasticities. In *JPT* 7 (2), pp. 37–58. DOI: 10.5038/2375-0901.7.2.3.

Liu, Zhili; Jia, Xudong; Cheng, Wen (2012): Solving the Last Mile Problem: Ensure the Success of Public Bicycle System in Beijing. In *Procedia - Social and Behavioral Sciences* 43, pp. 73–78. DOI: 10.1016/j.sbspro.2012.04.079.

Marquardt, Nadine; Fueller, Henning; Glasze, Georg; Puetz, Robert (2013): Shaping the Urban Renaissance: New-Build Luxury Developments in Berlin. In *Urban Studies* (50 (8)), pp. 1540–1556.

Martin, Elliot W.; Shaheen, Susan A. (2014): Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. In *Journal of Transport Geography* 41, pp. 315–324. DOI: 10.1016/j.jtrangeo.2014.06.026.

Mathä, Thomas Y.; Porpiglia, Alessandro; Ziegelmeyer, Michael (2017): Household wealth in the euro area. The importance of intergenerational transfers, homeownership and house price dynamics. In *Journal of Housing Economics* 35, pp. 1–12. DOI: 10.1016/j.jhe.2016.12.001.

Médard de Chardon, Cyrille; Caruso, Geoffrey; Thomas, Isabelle (2017): Bicycle sharing system 'success' determinants. In *Transportation Research Part A: Policy and Practice* 100, pp. 202–214. DOI: 10.1016/j.tra.2017.04.020.

Melly, Blaise (2005): Public-private sector wage differentials in Germany. Evidence from quantile regression. In *Empir Econ* 30 (2), pp. 505–520. DOI: 10.1007/s00181-005-0251-y.

Miedema, H. M.; Vos, H. (1998): Exposure-response relationships for transportation noise. In *The Journal of the Acoustical Society of America* 104 (6), pp. 3432–3445.

Mohammad, Sara I.; Graham, Daniel J.; Melo, Patricia C.; Anderson, Richard J. (2013): A meta-analysis of the impact of rail projects on land and property values. In *Transportation Research Part A: Policy and Practice* 50, pp. 158–170. DOI: 10.1016/j.tra.2013.01.013.

Mok, Henry M. K.; Chan, Patrick P. K.; Cho, Yiu Sun (1995): A Hedonic Price Model for Private Properties in Hong Kong. In *Journal of Real Estate Finance & Economics* 10 (1), pp. 37–48.

Molnar, Vigar (2010): The Cultural Production of Locality: Reclaiming the 'European City' in Post-Wall Berlin. In *International Journal of Urban and Regional Research* 34 (2), pp. 281–309. DOI: 10.1111/j.1468-2427.2010.00894.x.

Moral-Arce, Ignacio; Sperlich, Stefan; Fernández-Sainz, Ana I.; Roca, Maria J. (2012): Trends in the Gender Pay Gap in Spain. A Semiparametric Analysis. In *J Labor Res* 33 (2), pp. 173–195. DOI: 10.1007/s12122-011-9124-7.

Mummolo, Jonathan; Peterson, Erik (2018): Improving the Interpretation of Fixed Effects Regression Results. In *PSRM* 6 (4), pp. 829–835. DOI: 10.1017/psrm.2017.44.

Munn, Ian A.; Hussain, Anwar (2010): Factors determining differences in local hunting lease rates. Insights from blinder-oaxaca decomposition. In *Land Economics* 86 (1), pp. 66–78.

Nankervis, Max (1999): The effect of weather and climate on bicycle commuting. In *Transportation Research Part A: Policy and Practice* 33 (6), pp. 417–431. DOI: 10.1016/S0965-8564(98)00022-6.

Navrud, Ståle (2002): The State-of-the-art on economic valuation of Noise. In *European Commission*.

Nelson, Jon P. (2008): Hedonic Property Value Studies of Transportation Noise: Aircraft and Road Traffic. In Andrea Baranzini, José Ramirez, Caroline Schaerer, Philippe Thalmann (Eds.): Hedonic Methods in Housing Markets. New York, NY: Springer New York, pp. 57–82.

Neumark, David (1988): Employers' discriminatory behavior and the estimation of wage discrimination. In *Journal of human resources : JHR*.

Neumark, David; Simpson, Helen (2015): Place-Based Policies. In, vol. 5: Elsevier (Handbook of Regional and Urban Economics), pp. 1197–1287.

Newman, Sandra; Holupka, Scott; Ross, Stephen L. (2018): There's no place like home: Racial disparities in household formation in the 2000s. In *Journal of Housing Economics* 40, pp. 142–156. DOI: 10.1016/j.jhe.2018.04.002.

Newton, Isaac (1687): Philosophiæ Naturalis Principia Mathematica. Londini: Jussu Societatis Regiæ ac typis Josephi Streatii.

Nickell, Stephen (1981): Biases in Dynamic Models with Fixed Effects. In *Econometrica* 49 (6), p. 1417. DOI: 10.2307/1911408.

Oaxaca, Ronald (1973): Male-Female Wage Differentials in Urban Labor Markets. In *International Economic Review* 14 (3), p. 693. DOI: 10.2307/2525981.

Oliveira, Gustavo de L.T.; McKay, Ben; Plank, Christina (2017): How biofuel policies backfire: Misguided goals, inefficient mechanisms, and political-ecological blind spots. In *Energy Policy* 108, pp. 765–775. DOI: 10.1016/j.enpol.2017.03.036.

Olsen, Edgar O. (1972): An Econometric Analysis of Rent Control. In *Journal of Political Economy* 80 (6), pp. 1081–1100. DOI: 10.1086/259959.

Ospino, Carlos G.; Roldan Vasquez, Paola; Barraza Narvaez, Nacira (2010): Oaxaca-Blinder Wage Decomposition. Methods, Critiques and Applications: A Literature Review. In *Revista de Economia del Caribe*.

Otero, I.; Nieuwenhuijsen, M. J.; Rojas-Rueda, D. (2018): Health impacts of bike sharing systems in Europe. In *Environment international* 115, pp. 387–394. DOI: 10.1016/j.envint.2018.04.014.

Parkes, Stephen D.; Marsden, Greg; Shaheen, Susan A.; Cohen, Adam P. (2013): Understanding the diffusion of public bikesharing systems. Evidence from Europe and North America. In *Journal of Transport Geography* 31, pp. 94–103. DOI: 10.1016/j.jtrangeo.2013.06.003.

Partridge, Mark D.; Boarnet, Marlon; Brakman, Steven; Ottaviano, Gianmarco (2012): INTRODUCTION: WHITHER SPATIAL ECONOMETRICS? In *Journal of Regional Science* 52 (2), pp. 167–171. DOI: 10.1111/j.1467-9787.2012.00767.x.

Passchier-Vermeer, Willy; Passchier, Wim F. (2000): Noise Exposure and Public Health. Environmental Health Perspectives 108, pp. 123–131. DOI: 10.2307/3454637.

Pesaresi Martino, Melchiorri Michele, Siragusa Alice, Kemper Thomas (2016): Atlas of the Human Planet 2016. Mapping Human Presence on Earth with the Global Human Settlement Layer. Edited by Pesaresi Martino, Melchiorri Michele, Siragusa Alice, Kemper Thomas. Joint Research Centre. European Commission.

Phillips, Peter C. B. (1996): Econometric Model Determination. In *Econometrica* 64 (4), p. 763. DOI: 10.2307/2171845.

Phillips, Peter C.B.; Sul, Donggyu (2007): Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. In *Journal of Econometrics* 137 (1), pp. 162–188. DOI: 10.1016/j.jeconom.2006.03.009.

Pineda-Ríos, Wilmer; Giraldo, Ramón; Porcu, Emilio (2019): Functional SAR models: With application to spatial econometrics. In *Spatial Statistics* 29, pp. 145–159. DOI: 10.1016/j.spasta.2018.12.002.

Pinkse, Joris; Slade, Margaret E. (2010): THE FUTURE OF SPATIAL ECONOMETRICS. In *Journal of Regional Science* 50 (1), pp. 103–117. DOI: 10.1111/j.1467-9787.2009.00645.x.

Plümper, Thomas; Neumayer, Eric (2010): Model specification in the analysis of spatial dependence. In *European Journal of Political Research* 49 (3), pp. 418–442. DOI: 10.1111/j.1475-6765.2009.01900.x.

Proost, Stef; Thisse, Jacques-François (2019): What Can Be Learned from Spatial Economics? In *Journal of Economic Literature* 57 (3), pp. 575–643. DOI: 10.1257/jel.20181414.

Quality of transport. Report (2014). Luxembourg: Publications Office (Special Eurobarometer, 422a).

Radzimski, Adam; Dzięcielski, Michał (2021): Exploring the relationship between bike-sharing and public transport in Poznań, Poland. In *Transportation Research Part A: Policy and Practice* 145, pp. 189–202. DOI: 10.1016/j.tra.2021.01.003.

Rasmussen, David W.; Zuehlke, Thomas W. (1990): On the choice of functional form for hedonic price functions. In *Applied Economics* 22 (4), pp. 431–438. DOI: 10.1080/00036849000000002.

Redding, Stephen J.; Rossi-Hansberg, Esteban (2017): Quantitative Spatial Economics. In *Annu. Rev. Econ.* 9 (1), pp. 21–58. DOI: 10.1146/annurev-economics-063016-103713.

Ricci, Miriam (2015): Bike sharing. A review of evidence on impacts and processes of implementation and operation. In *Research in Transportation Business & Management* 15, pp. 28–38. DOI: 10.1016/j.rtbm.2015.03.003.

Rudloff, Christian; Lackner, Bettina (2014): Modeling Demand for Bikesharing Systems. In *Transportation Research Record* 2430 (1), pp. 1–11. DOI: 10.3141/2430-01.

Sah, Vivek; Conroy, Stephen J.; Narwold, Andrew (2016): Estimating School Proximity Effects on Housing Prices: the Importance of Robust Spatial Controls in Hedonic Estimations. In *J Real Estate Finan Econ* 53 (1), pp. 50–76. DOI: 10.1007/s11146-015-9520-5.

Santos, Georgina; Maoh, Hanna; Potoglou, Dimitris; Brunn, Thomas von (2013): Factors influencing modal split of commuting journeys in medium-size European cities. In *Journal of Transport Geography* 30, pp. 127–137. DOI: 10.1016/j.jtrangeo.2013.04.005.

Sarafidis, Vasilis; Yamagata, Takashi; Robertson, Donald (2009): A test of cross section dependence for a linear dynamic panel model with regressors. In *Journal of Econometrics* 148 (2), pp. 149–161. DOI: 10.1016/j.jeconom.2008.10.006.

Schafer, Philipp; Hirsch, Jens (2017): Do Urban Tourism Hotspots Affect Berlin Housing Rents? In *International Journal of Housing Markets and Analysis* (10 (2)), pp. 231–255.

Seidman, Michael D.; Standring, Robert T. (2010): Noise and quality of life. In *International journal of environmental research and public health* 7 (10), pp. 3730–3738. DOI: 10.3390/ijerph7103730.

Shaheen, Susan; Chan, Nelson (2016): Mobility and the Sharing Economy: Potential to Facilitate the First- and Last-Mile Public Transit Connections. In *built environ* 42 (4), pp. 573–588. DOI: 10.2148/benv.42.4.573.

Sirmans, G. Stacy; Macpherson, David A.; Zietz, Emily N. (2005): The composition of hedonic pricing models. In *Journal of Real Estate Literature* 13 (1), pp. 3–46.

Small, K. (2013): Urban Transportation Economics. Hoboken: Taylor and Francis. Available online at https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=1273224.

Smith, Keith (1993): The influence of weather and climate on recreation and tourism. In *Weather* 48 (12), pp. 398–404. DOI: 10.1002/j.1477-8696.1993.tb05828.x.

Standing, Craig; Standing, Susan; Biermann, Sharon (2019): The implications of the sharing economy for transport. In *Transport Reviews* 39 (2), pp. 226–242. DOI: 10.1080/01441647.2018.1450307.

Swoboda, Aaron; Nega, Tsegaye; Timm, Maxwell (2015): HEDONIC ANALYSIS OVER TIME AND SPACE: THE CASE OF HOUSE PRICES AND TRAFFIC NOISE. In *Journal of Regional Science* 55 (4), pp. 644–670. DOI: 10.1111/jors.12187.

Tang, Kathy X.; Waters, Nigel M. (2005): The internet, GIS and public participation in transportation planning. In *Progress in Planning* 64 (1), pp. 7–62. DOI: 10.1016/j.progress.2005.03.004.

Thanos, Sotirios; Dubé, Jean; Legros, Diègo (2016): Putting time into space. The temporal coherence of spatial applications in the housing market. In *Regional Science and Urban Economics* 58, pp. 78–88. DOI: 10.1016/j.regsciurbeco.2016.03.001.

Theebe, MarcelA. J. (2004): Planes, Trains, and Automobiles: The Impact of Traffic Noise on House Prices. The Journal of Real Estate Finance and Economics. In *The Journal of Real Estate Finance and Economics* 28 (2-3), pp. 209–234. DOI: 10.1023/B:REAL.0000011154.92682.4b.

Thisse, Jacques-François (2010): How Transport Costs Shape the Spatial Pattern of Economic Activity. In : The Future for Interurban Passenger Transport: OECD, pp. 25–51.

Thomschke, Lorenz (2015): Changes in the distribution of rental prices in Berlin. In *Regional science & urban economics* 51 (2015), pp. 88–100.

Tin Tin, Sandar; Woodward, Alistair; Robinson, Elizabeth; Ameratunga, Shanthi (2012): Temporal, seasonal and weather effects on cycle volume: an ecological study. In

Environmental health : a global access science source 11, p. 12. DOI: 10.1186/1476-069X-11-12.

Tobler, W. R. (1970): A Computer Movie Simulating Urban Growth in the Detroit Region. In *Economic Geography* 46, p. 234. DOI: 10.2307/143141.

Tran, Tien Dung; Ovtracht, Nicolas; d'Arcier, Bruno Faivre (2015): Modeling Bike Sharing System using Built Environment Factors. In *Procedia CIRP* 30, pp. 293–298. DOI: 10.1016/j.procir.2015.02.156.

Ulrich Leth; Tadej Brezina; Bertram Ludwig; Christina Birett (2017): Is Bike Sharing Competitor, Relief or Supplement to Public Transport?

United Nations (2019): World urbanization prospects. The 2018 revision. New York: United Nations.

V E, Sathishkumar; Cho, Yongyun (2020): A rule-based model for Seoul Bike sharing demand prediction using weather data. In *European Journal of Remote Sensing* 53 (sup1), pp. 166–183. DOI: 10.1080/22797254.2020.1725789.

Wang, Kailai; Akar, Gulsah; Chen, Yu-Jen (2018): Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City's bike share. In *Transportation Research Part A: Policy and Practice* 116, pp. 1–14. DOI: 10.1016/j.tra.2018.06.001.

Watkins, Craig (1999): Property valuation and the structure of urban housing markets. In *J of Property Inv & Finance* 17 (2), pp. 157–175. DOI: 10.1108/14635789910258543.

Wei, Wenhui (2005): China and India. Any difference in their FDI performances? In *Journal of Asian Economics* 16 (4), pp. 719–736. DOI: 10.1016/j.asieco.2005.06.004.

Wells, Samantha; Graham, Kathryn; Purcell, John (2009): Policy implications of the widespread practice of 'pre-drinking' or 'pre-gaming' before going to public drinking establishments: are current prevention strategies backfiring? In *Addiction (Abingdon, England)* 104 (1), pp. 4–9. DOI: 10.1111/j.1360-0443.2008.02393.x.

Wikipedia (2020): Fahrradverleihsystem. Deutschland. Edited by Wikipedia. Available online at https://de.wikipedia.org/wiki/Fahrradverleihsystem#Deutschland, checked on 5/7/2020.

Wilhelmsson, Mats (2002): Spatial Models in Real Estate Economics. In *Housing, Theory & Society* 19 (2), pp. 92–101. DOI: 10.1080/140360902760385646.

Ye, Xinyue; Dang, Anrong (2013): Spatial Analysis and Modeling of Urban and Regional Development. In *Annals of GIS* 19 (3), p. 129. DOI: 10.1080/19475683.2013.827448.

Zhang, Jian; Pan, Yaozong (2018): Planning Station Capacity and Bike Rebalance Based on Visual Analytics of Taxi and Bike-Sharing Data. In : 2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). 2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). Zhengzhou, China, 2018: IEEE, pp. 305–3054.

Zhang, Lei; Yi, Yimin (2018): What contributes to the rising house prices in Beijing? A decomposition approach. In *Journal of Housing Economics* 41, pp. 72–84. DOI: 10.1016/j.jhe.2018.04.003.

Zorlu, Aslan; Mulder, Clara H.; van Gaalen, Ruben (2014): Ethnic disparities in the transition to home ownership. In *Journal of Housing Economics* 26, pp. 151–163. DOI: 10.1016/j.jhe.2014.01.004.

6. General Appendix

6.1 Summary

Spatial and urban economics are mainly concerned with explaining the spatial distributions of economic amenities and activities as well as the interactions between these amenities and activities. This dissertation contains three essays focused on the effects of amenities, disamenities and the socioeconomic structure on demand in urban areas.

In Chapter 2, we use real estate data from Berlin to simultaneously analyze the effects of alternative traffic noise sources. We compare the most frequent traffic noise sources i.e., flight noise, train noise and road noise. We do this to isolate the relative harms of each source. This research adds to the literature, which has only analyzed one noise source or has aggregated the noise levels of different sources. We find that all noise sources have a strong negative effect on real estate prices. Flight noise had the most negative effect on housing prices, and road and train noises had similar but more minor effects. Our general results are in line with the literature regarding single noise sources.

Chapter 3 provides the first decomposition of the price gap between prices for single family houses and apartments into their microeconomic determinants. Using individual data from the collection of real estate transactions data for Berlin, Germany, we find a positive price gap for houses in Berlin in 1990 that decreased up to 1998. When controlling for structural characteristics, no positive price gap can be identified between 1996 and 2006. The application of Oaxaca decomposition reveals a growing preference for size (favoring houses), which is – since 2006 - offset by a growing preference for a more central location. In Berlin, this central location is primarily reserved for apartments. In summary, a change in preferences can be observed for Berlin over the 25 years.

Chapter 4 analyses the effects between the social structure of the general population surrounding a bike-sharing station and the usage of that station. Compared to other European countries, Germany has a rather big bike-sharing market with systems in more than 40 cities. The most extensive German bicycle sharing system (StadtRAD) is placed in the city of Hamburg and used for this case study. The trip demand is estimated using daily trip starts for every station. Our empirical approach consists of different configurations of semi-log models. The models control for standard explanatory variables shown in the literature to affect bike-sharing, i.e., weather, season, lagged demand, travel distance, and infrastructure. The model extends the literature by including real time bicycle availability and socioeconomic structure variables like age and gender. Our results reinforce findings in the literature for baseline variables. In addition, most notably, a higher share of young adults and females positively impacts demand. Bicycle availability also has a positive impact.

6.2 Zusammenfassung (German Summary)

Sowohl die Raum- als auch Stadtökonomie befassen sich hauptsächlich mit der Erklärung der räumlichen Verteilung wirtschaftlicher Annehmlichkeiten/Leistungen und Aktivitäten. Ebenso wird die Wechselwirkungen zwischen diesen Annehmlichkeiten/Leistungen und Aktivitäten betrachtet. Diese Dissertation enthält drei Aufsätze, die sich mit den Auswirkungen von Annehmlichkeiten/Leistungen, Unannehmlichkeiten und der sozioökonomischen Struktur auf die Nachfrage in städtischen Gebieten befassen.

In Kapitel 2 verwenden wir Immobiliendaten aus Berlin, um gleichzeitig die Auswirkungen alternativer Verkehrslärmquellen zu analysieren. Wir vergleichen die häufigsten Verkehrslärmquellen, d.h. Fluglärm, Zuglärm und Straßenlärm. Wir tun dies, um die relativen Auswirkungen der einzelnen Quellen zu isolieren. Diese Untersuchung ergänzt die Literatur, in der nur eine Lärmquelle analysiert oder die Lärmpegel verschiedener Quellen zusammengefasst wurden. Wir stellen fest, dass alle Lärmquellen einen starken negativen Einfluss auf die Immobilienpreise haben. Fluglärm hat den stärksten negativen Effektt auf die Immobilienpreise, während Straßen- und Bahnlärm vegleichbare, aber verhältnismäßig geringere Effekte haben. Unsere allgemeinen Ergebnisse stehen im Einklang mit der Literatur zu einzelnen Lärmquellen.

In Kapitel 3 wird der Preisunterschied zwischen den Preisen für Einfamilienhäuser und Wohnungen erstmals in seine mikroökonomischen Determinanten zerlegt. Unter Verwendung von Einzeltransaktionsdaten aus der Sammlung von Immobilientransaktionsdaten für Berlin, Deutschland, finden wir einen positiven Preisaufschlag für Häuser im Vergleich zu Wohnungen in Berlin im Jahr 1990. Dieser Aufschlag nimmt bis 1998 kontinuierlich ab. Sofern für strukturelle Merkmale kontrolliert wird kann zwischen 1996 und 2006 kein Preisaufschlag für Häuser oder Wohnungen im Vergleich zum anderen festgestellt werden. Die Anwendung der Oaxaca-Dekomposition zeigt eine wachsende Präferenz für Größe (zugunsten von Häusern), die seit 2006 durch eine wachsende Präferenz für eine zentralere Lage ausgeglichen wird. Diese zentrale Lage ist in Berlin vor allem den Wohnungen vorbehalten. Zusammenfassend lässt sich für Berlin über den Zeitraum von 25 Jahren ein Wandel in den Präferenzen beobachten.

In Kapitel 4 wird der Zusammenhang zwischen der Sozialstruktur der Bevölkerung im Umfeld einer Fahrradverleihstation und der Nutzung dieser Station untersucht. Im Vergleich zu anderen europäischen Ländern hat Deutschland einen relativ großen Fahrradverleih Markt mit Systemen in mehr als 40 Städten. Das größte deutsche Fahrradverleihsystem (StadtRAD) befindet sich in der Stadt Hamburg und wird für diese Fallstudie verwendet. Die Fahrtennachfrage wird anhand der täglichen Fahrten für jede Station geschätzt. Unser empirischer Ansatz besteht aus verschiedenen Konfigurationen von Semi-Log-Modellen. Die Modelle kontrollieren die Standardvariablen, die in der Literatur gezeigt haben, dass sie einen Einfluss auf die Nutzung von Fahrradverleih haben, d.h. Wetter, Jahreszeit, frühere Nachfrage, Fahrdauer und Infrastruktur. Das Modell erweitert die Literatur, indem es die Fahrradverfügbarkeit in Echtzeit und sozioökonomische Strukturvariablen wie Alter und Geschlecht einbezieht. Unsere Ergebnisse untermauern die Erkenntnisse der Literatur insbesondere für Basisvariablen. Darüber hinaus hat vor allem ein höherer Anteil junger Erwachsener und Frauen einen positiven Einfluss auf die Nachfrage. Auch die Verfügbarkeit von Fahrrädern am dem Verleihstationen hat einen positiven Einfluss.

6.3 List of Publications

Beimer, Waldemar; Maennig, Wolfgang (2017):
Noise effects and real estate prices.
A simultaneous analysis of different noise sources.
In *Transportation Research Part D: Transport and Environment* 54, pp. 282–286.
https://doi.org/10.1016/j.trd.2017.05.010

Beimer, Waldemar; Maennig, Wolfgang (2020):
On the Price Gap between Single Family Houses and Apartments.
In *Journal of Housing Economics* 49, p. 101686.
https://doi.org/10.1016/j.jhe.2020.101686

6.4 Erklärung

Hiermit erkläre ich, Waldemar Beimer, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Ort/Datum Unterschrift Doktorand

6.5 **Eidesstattliche Versicherung:**

Ich, Waldemar Beimer, versichere an Eides statt, dass ich die Dissertation mit dem Titel: „EXTERNAL EFFECTS ON DEMAND IN EMPIRICAL URBAN ECONOMICS“ selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern gemäß den beigefügten Darlegungen nach § 6 Abs. 3 der Promotionsordnung der Fakultät für Wirtschafts- und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt.

Ort/Datum Unterschrift Doktorand