RUHR
ECONOMIC PAPERS

Said Benjamin Bonakdar

Segregation, Housing and Neighborhood Dissimilarities: A Case Study for the City of Bochum

RUB

#914

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Manuel Frondel, Prof. Dr. Torsten Schmidt, Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #914

Responsible Editor: Thomas Bauer

ISSN 1864-4872 (online) – ISBN 978-3-96973-059-1

Ruhr Economic Papers #914

Said Benjamin Bonakdar

Segregation, Housing and Neighborhood Dissimilarities: A Case Study for the City of Bochum

RUB

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973059
ISSN 1864-4872 (online)
ISBN 978-3-96973-059-1

Said Benjamin Bonakdar[1]

Segregation, Housing and Neighborhood Dissimilarities: A Case Study for the City of Bochum

Abstract

The rising concentration of low-income households and ethnic minorities has become an important policy issue in Germany. The Ruhr Area is particularly interesting, because it is one of the largest conurbations in Europe and experienced radical structural changes in the past, which are connected to the boom and the deindustrialization of the coal mining and steel industry. Since there is no empirical evidence about the extent of residential segregation within the cities of the Ruhr Area, I use micro data on house-coordinates levels to investigate the urban structure of Bochum, which is located in the center of the conurbation. The results show that Bochum can be characterized by four different clusters under consideration of socioeconomic variables, dwelling rents and psychological indicators, e.g. captured by the so-called Sinus Milieus and Limbic Types. With a CHAID decision tree on dwelling rents I learned that especially dwelling size is a key separator between higher and lower dwelling prices in Bochum. The variables identified in the decision tree, like e.g. number of rooms or construction year, are used for hedonic price estimations within all clusters and show positive effects on dwelling rents. Finally, I found that, across all clusters, a rise in different satisfaction types also increase the odds of houses to be located in neighbourhoods with positive moving balance, which is a proxy for a high willingness-to-stay.

JEL-Code: R21, R23, R31

Keywords: Urban housing; housing prices; rent; neighbourhood characteristics; segregation

June 2021

1 Said Benjamin Bonakdar, RUB. – All correspondence to: Said Benjamin Bonakdar, Ruhr-University Bochum, Institute for Macroeconomics Universitätsstr. 150, 44801 Bochum, Germany, e-mail: benjamin.bonakdar@rub.de

1 Introduction

Social inequalities have increased significantly in many European countries since the mid-1970s (Nolan et al., 2014). It appears that exisiting institutions and policies have been rather unsuccessful in tackling poverty (Atkinson, 2016). This inequality partly manifests itself in urban structures, which are often characterized by spatially separated concentrations of different population groups with varying socioeconomic status (Malmberg et al., 2013). Musterd et.al (2017) find that especially capital cities in Europe show an increasing gap between poor and rich, which goes along with strong segregation patterns. The authors, furthermore, claim that socioeconomic segregation is caused by at least five key factors: social inequalities, changing economic structures and levels of global connectedness, welfare regimes, and housing systems (ibid.).

In Germany and in other parts of Western Europe, demographic change, international mobility and the rising concentration of low-income households and ethnic minorities in certain urban areas have become important policy issues (Smets and Salman, 2008; Musterd, 2005). Reuschke & Weck (2013, p.191) argue that this political debate on social mix and neighbourhood effects in segregated areas is particularly important for the Ruhr Area, because historically, there has been a lot of international mobility. Nowadays, the Ruhr Area has a population size of more than five million inhabitants and represents one of the largest conurbations in Europe.

From the early 19th century to the mid-20th century, the Ruhr Area boomed and flourished due to coal mining and steel industries. The growth of these branches required a large amount of workers, which resulted in an influx of many migrants, who were accomodated closely to the industrial locations (Shrestha et al., 2016, p.3). These movements led to an emergence of residential segregation. However, in the mid-20th century, deindustrialization began, for which reason the Ruhr Area suffered under radical structural changes with high amounts of job losses and several abandoned and derelict spaces (ibid.). Due to the history of the Ruhr Area, residential segregation and social inequality became more important over the years, since these processes have a strong influence on the individual life chances and integration within the society (Reuschke and Weck, 2013, p.19). Even though, the authors analyzed how the concentration of specifically Turkish individuals in certain neighbourhoods can be explained in the Ruhr Area, as a whole, there is no empirical evidence about the extent of residential segregation in the various cities in the Ruhr Area.

Therefore, this paper investigates empirically the urban structure of the city of Bochum, which is located in the center of the Ruhr Area and which is one of the largest cities in North Rhine Westphalia. The purpose of this investigation is to identify how strong potential segregation patterns are in Bochum, but also to learn about the socioeconomic structure of the population within potential clusters. Here, I also consider psychological indicators like the so-called Sinus-Milieus and Limbic Types in order to get a better understanding of the clustering process within Bochum. Due to the data availability, the analysis is limited only to the city of Bochum.

Furthermore, I want to get an understanding of the housing market in Bochum and how various factors may influence dwelling rents in potential clusters. Aside from socioeconomic and housing characteristics, I also include residential satisfaction to the analysis, because it may be an indicator for the moving decision of a household and thus, can provide information on the current composition of various neighbourhoods.

In order to tackle these issues, I ask the following research questions:

Q1: Does the city of Bochum show patterns of segregation and consequently, different average dwelling rents?

Q2: Which variables of interest show the greatest importance for dwelling rents in Bochum?

Q3: How does residential satisfaction in particular neighbourhoods affect the moving decision of households in various homogenous clusters (if any) as well as in the entire city of Bochum?

Q4: What are the effects of the variables of interest on dwelling rents in the homogenous clusters (if any) and in Bochum as a whole?

In Section 2, I provide some theoretical background and present the respective hypotheses, which need to be tested. In Section 3, I show the data description and the empirical strategies. The main results are presented in Section 4 and finally, I conclude in Section 5.

2 Theoretical Background

If households are willing to move, the relevant factors for the moving decision, and thus for urban mobility, are based on different factors of residential satisfaction. According to Speare (1974), residential satisfaction

is defined by the influence of individual and residence characteristics as well as by the satisfaction with ones neighbours and potential social bonds within the neighbourhood. The foundation of the evaluation of neighbourhood similarity goes back to Festinger's (1954) social comparison theory, which claims that individuals constantly compare themselves with their respective peer group. According to the theory, similar peers are preferred for comparison. Zanna et.al (1975) suggest that attribute similarity is crucial for social comparison. Consequently, individuals evaluate various economic, but also non-economic factors within their neighbourhood to determine their satisfaction (Bourdieu, 1984). Clark and Coulter (2015) show that individuals who feel similar to their neighbours have a lower desire to move to another neighbourhood. Schelling (1971; 1978), Clark (1992) and Luttmer (2005), among many others, claim that specifically the preference for living among similar neighbours affects the moving decision. The cluster formation and thus, the similarity to others depends on critieria like ethnicity or income, e.g. shown by the existence of various status symbols. Based on that, I propose the first hypothesis H1: Satisfactions across all clusters in Bochum show positive effects on the 'Moving Balances' (as proxy for the willingness-to-stay of households and attraction indicator of neighbourhoods) and thus, influence the moving decisions of households.

Aside from residential satisfaction, households also need to deal with the evaluation of other factors (Ioannides, 2011, p.1283). These include properties of the house itself, amenities and disamenities in the neighbourhood and affordability (Baker et al., 2015; Sirmans et al., 2006). Rosen (1974) claims that house prices capitalize the houses' attributes and the respective neighbourhood in a competitive market. Based on that, a lot of research has been done on hedonic pricing in the housing market by estimating the value of homes on these attributes (Sirmans et al., 2005). The authors have analyzed 125 different studies and found that specifically dwelling characteristics, like e.g. lot size, square feet or no. of rooms, were important for the analysis. Malpezzi (2002) notes that different individuals may value these various housing characteristics in a different manner. Based on that, I propose the second hypothesis H2: Dwelling characteristics have the largest effect on the house price determination in Bochum and are of different importance for households in different clusters under consideration of interpersonal similarity.

Malik et.al (2015) showed that education can also lead to various homogenous areas, which are specifically characterized by high or low creativity of the inhabitants. Narrowing it down, Guilford (1950, p.444) defines

creative behavior as manifestation of creative patterns. Individuals, who are characterized by these behavioral patterns typically follow activities like inventing, designing, contriving, composing and planning (ibid.). Individuals, who prefer other activities have other patterns of traits and thus, other psychological mindsets. According to the social comparison theory, these different psychological mindsets may have an influence on the moving decision of households in order to find their respective peer group and thus, may influence the clustering in urban areas similar to what Malik et.al.(2015) suggest. There are several approaches, which operationalize these different mindsets. In this study, I focus on the approaches of the so-called Sinus-Milieus and Limbic Types, which were originally meant as instruments for geomarketing and more personalized advertisement[1].

The categorization into Sinus-Milieus is one of the leading lifestyle approaches in Europe in the field of Marketing (Schwarz and Ernst, 2009, p.499). It was developed by the Sinus Institute for Market and Social Research (*Sinus: Markt-und Sozialforschung*), which is one of the major marketing companies in Europe. Under consideration of values and attitudes from 40 questions, the company divides the German population into ten Sinus-Milieus. A classification can be found in Figure (1).

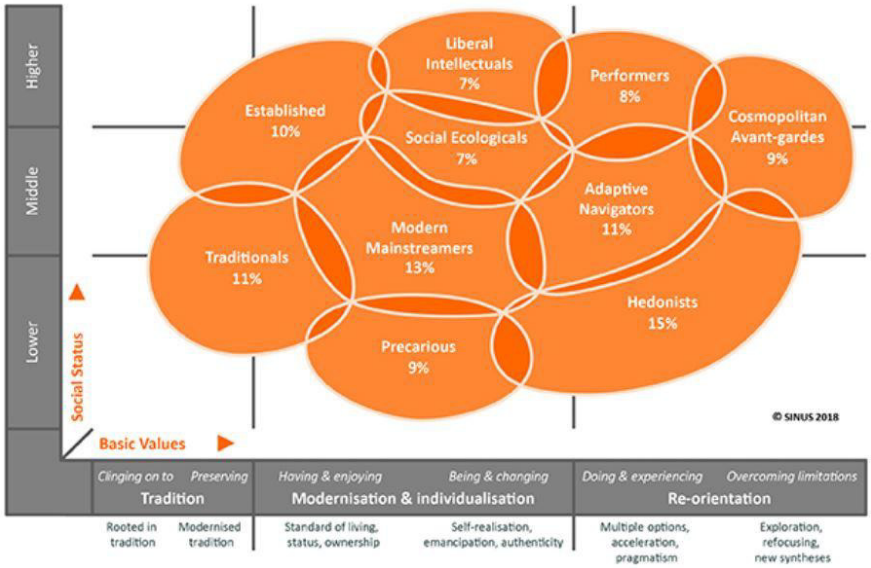

Figure 1: The Sinus-Milieus in Germany 2018 - Social Status and Basic Orientation

Source: Sinus: Markt- und Sozialforschung (2018)

[1]Even though, these categorizations are rather unique and are meant to be used for enterprises, they may be useful for various analyses, since they capture different values and attitudes of the heterogenous population within Germany. So far, they are barely used in Economic research.

According to the Sinus Institute, the vertical axis describes the social status (measured by income / education / profession) of an individual, whereas the horizontal axis refers to basic normative orientation such as tradition, modernization and re-orientation. Where "Traditionals" are characterized by lower to middle social status with a strong tendency to the value "Tradition", "Cosmopolitian Avantgardes" have a middle to higher social status and are typically exploring and refocusing. Other classifications are "Established", "Precarious", "Modern Mainstreamers", "Social Ecologicals", "Liberal Intellectuals", "Performers", "Adaptive Navigators" and "Hedonists".

The Limbic Types refer to studies in Neuropsychology, where every human can be categorized into one of seven different categories depending on their values and attitudes (Haeusel, 2011). The difference to the Sinus-Milieus is that the categorization criteria are based on more factors than status and values. Even though, there may be some overlaps, both approaches are useful for this study, because both categorizations cover different psychological aspects and thus, may be an indicator, where an individual wants to move within an urban area.

Haeusel (2011, p.58) defines seven different Limbic types: "Harmonizer", "Open-minded", "Hedonist", "Adventurer", "Performer", "Disciplinarian" and "Traditionalist". "Harmonizers" are characterized by a high level of social and family orientation. These individuals have low ascendancy as well as low status-orientation and show rather a desire for security. "Open-minded" people show a great openness to new things, have a sense of well-being and are also characterized as tolerant. "Hedonists" actively look for new things and have a high level of individualism. Also, a high level of spontaneity is typical. The fourth Limbic type is "Adventurer", where individuals show a large appetite for risk combined with low impulse control. "Performers" are defined as people, who have a high level of performance orientation, ambition and finally, a high level of status orientation. "Disciplinarians" are characterized by a high sense of duty and a low desire for shopping. Also, they pay high attention to detail. Finally, "Traditionalists" show a low level of future-orientation. However, they are characterized by a desire for order and security. Putting it together, I propose Hypothesis 3: The clusters of the city of Bochum can be characterized by socioeconomic variables, prices and the Sinus-Milieus and Limbic Types.

3 Data and Empirical Strategies

3.1 Data Description and Processing

This study makes use of three different data sources, which are cross-section data sets for the year 2018 [2]. At first, I need data on the socioeconomic characteristics of Bochum's population. This data was provided on the house coordinates-level by the geomarketing company *microm*[3] as categorical data as well as ordinal data in form of 8-point and 9-point Likert scales. It is important to mention that the data represents classifications within the houses, or to be more specific, they represent categories for the respective distribution within one house. E.g. there is no actual data on the age, but rather a classification according to the interval of the ages of households' heads across all households within the observed house. For simplicity reasons, the Likert scales get transformed into 5-point Likert scales in order to have a consistent evalution for the variables of interest (e.g. from 1: lowest to 5: highest). The transformation occurs by merging observations from different parameter characteristics, which imply a similar interpretation and which are not clearly distinguishable, e.g strongly negative and very strongly negative. Overall, the data set contains information on the debt default probability as proxy for income, status as well as the share of foreigners and shares of children in each house. Also, there is data on the average age of each household's head, probability of associated Sinus-Milieus, the so-called Primary Limbic type according to Haeusel (2011) and the moving balance, which is positive, if these houses and neighbourhoods attract many households to live and remain there. It becomes negative, if the area suffers from high fluctuation. The variable can be interpreted as proxy for more attractive or rather unpopular neighbourhoods, where individuals have a higher or lower willingness-to-stay.

Furthermore, housing characteristics as well as house price data are necessary in order to gain insights on the housing market situation in Bochum. The data was gathered by *F+B research and consultancy (original name: F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH)* from various real estate platforms on the object level and contains information on the absolute offer price in EUR, object size, construction year, number of rooms, various object endowments and full address of the object. Also, there

[2] The combination of those three data sets is unique and has not been used for any geospatial analyses to the best of my knowledge

[3] *microm* does not collect the data themselves, but rather obtains the data from various different partners, like e.g. Creditreform, CEG Consumer Reporting, German Postal Service, Federal and State Offices of Statistics, Federal Employment Agency and many more - a list of partners can be found on https://microm.de/partner/. The company then processes the data and creates categorizations on different geo-levels, e.g. house-level, street section-level and others.

is the differentiation between purchase and rent of houses and apartments. From the object endowments, I create an endowment index (E-Index) as proxy for dwelling quality based on a text mining approach by using the relative frequencies of keywords with a minimum value of 1% across all offers. Therefore, the index for each object is the sum of 20 keywords weighted by their relative frequency appearing in all offers[4]. Figure (2) shows the word cloud of the single characteristic keywords for Bochum in 2018, weighted by the relative frequency. It seems that "central heating", "cellar", "balcony/patio", "bright" and "garage-/parking lot" are most common for dwelling ads in Bochum.

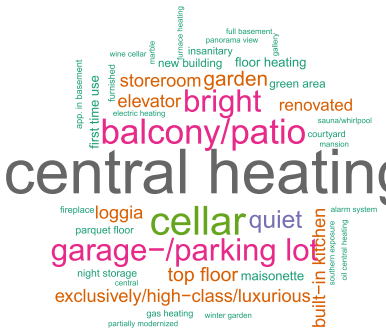

Figure 2: Word Cloud for Housing Characteristics in Bochum in 2018

Source: F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own illustration

Finally, the role of residential satisfaction within Bochum plays a role for the analysis. For this reason, the research institute *InWIS (original name: Institut für Wohnungswesen, Immobilienwirtschaft, Stadt- und Regionalentwicklung)* provided a sample on the neighbourhood level with various satisfaction indicators. The relevant data is represented as means over ordinal Likert-scale variables between 1 and 5. It contains information on satisfaction with e.g. the apartment itself, respective apartment characteristics, the neighbourhood and the total living situation. Finally, there is data on the total satisfaction level, which is provided in percentages.

In order to get a single data set containing the information, the data sets get merged on the neighbourhood level by the respective coordinates using the WGS84 system. For the data from *F+B research and consultancy* and from *InWIS* the respective georeferenced data gets transformed into the WGS84 coordinate system by

[4]Since the characteristical keywords were added individually, it can be possible that some dwellings have actually more characteristics than stated. However, the frequency of posting various words is an indicator of what dwelling owners and potential renters may prioritize. Furthermore, the index controls for doublings, e.g. any kind of heating system has the same influence on the index.

using *OpenStreetMap*.

3.2 Empirical Approaches

In order to get meaningful results for the city of Bochum, I conduct the following analyses with respect to the research questions and the hypotheses.

1. **Descriptive Statistics**. For the descriptive statistics, different kinds of visualizations are chosen. The three data sets are shown separately[5]. For the continuous variables like e.g. price and dwelling size the descriptive statistics are shown tabularly. For the Likert Scale data like e.g. Status, Debt Default Probability, Moving Balance and Satisfaction types as well as for the categories of the so-called Sinus-Milieus and Limbic Types a graphical representation is chosen, which specifically covers the relative frequencies of the respective characteristic. Note that the analysis is performed only with apartment rent data.

2. **Cluster Analysis**. The cluster analysis contains socioeconomic variables, the Dominant Sinus Milieus, the Primary Limbic Types and the price of the respective unit. For this purpose, the data sets from *microm* and *F+B research and consultancy* get merged on the neighbourhood level [6], resulting in a sample size of $N = 45,859$[7]. Here, I want to analyze if and how the clusters get generated and if the city of Bochum is segregated. Since the data set contains various types of data, like continuous data and ordinal data, the method of the Euclidean distance within the cluster analysis is not directly applicable. Also, the variables of interest as well as the so-called Dominant Sinus Milieus and Primary Limbic Types need to be considered as unordered factor variables, because the variables are not supposed to be scaled numerically from e.g. 1 to 10. Therefore, the appropriate approach is to use the Gower metric to determine a matrix $D = (d_{k,l})$, which contains the dissimilarities between entities, here factor variables. This dissimilarity matrix is square and symmetric $(N \times N)$ and contains the pairwise dissimilarities/distances between the observations (Gower, 1971). Equation (1) shows the mathematical

[5] For each analysis, another merged data set is required in order to obtain appropriate sample sizes
[6] This represents a precision of roughly 111m
[7] I assume that houses in the same neighbourhood have similar characteristics and therefore have a similar price - this assumption holds for any other merge within this analysis

formulation of Gower's metric.

$$S_{i,j} = \frac{\sum_{k=1}^{N} w_{i,j,k} S_{i,j,k}}{\sum_{k=1}^{N} w_{i,j,k}} \quad , \tag{1}$$

where $S_{i,j,k}$ defines the difference between the instances $X_{i,k}$ and $X_{j,k}$ and $w_{i,j,k}$ denotes the weight for variable k between the observations X_i and X_j. With matrix D, the agglomerative hierarchical cluster analysis with the WARD approach and Euclidean distance can be used. The WARD method is suitable, because it minimizes the total within-cluster variance. The analysis itself proceeds from an initial partition in N single-entity clusters and leads to successive mergings of clusters until all entities belong to its respective cluster(Hansen and Jaumard, 1997). The results are shown tabularly and cartographically by visualizing the different characteristics within each cluster in comparison to the entire city of Bochum. The data for the cartographic representation gets aggregated to the so-called *PLZ8*-level, which represent subcategories from the different zip-codes within Bochum.

3. **CHAID Decision Tree model**. Decision Tree models belong to the category of Machine-Learning algorithms, which are used to classify data (Quinlan, 1986) and to determine the predictor variables, which have the most influence on the target variable. In general, Decision Trees are hierarchical models, in which the data gets split at each node into distinct categories. The top of the Decision Tree represents the root node and is characterized as highest influence on the target. In this paper, the data set contains different data types, namely ordinal, categorical and continuous data. Therefore, the CHAID algorithm is the most suited approach, because it is able to capture all the various data types. For simplicity reasons, the continuous variables as well as the ordinal scaled variables in the data set get transformed into binary variables. The threshold is determined by the median of each variable so that any value which is larger or equal to the median is denoted as "high" and values below the median are denoted as "low" (see Equation (2)).

$$X_{i,k} = \begin{cases} \text{high} & \text{if } X_{i,k} \geq median(X_k) \\ \text{low} & \text{if } X_{i,k} < median(X_k) \end{cases} \tag{2}$$

In order to get robust results from the estimated performance of the CHAID model, the method of "repeated cross-validation" and three different significance levels $\alpha \in [0.001, 0.01, 0.05]$ are chosen in order to find the most important influence factors on dwelling rents/prices for the city of Bochum. Therefore, the data sets from *microm* and *F+B research and consultancy* get merged on the neighbourhood level. The resulting sample size is N = 45,859.

4. **Regression Approaches**. In order to determine the effects of satisfactions on the moving decision, a logistic regression approach is chosen, where the variable "Moving Balance" serves as dependent proxy dummy variable. The proxy variable is equal to 0, if a house has a negative moving balance and equal to 1, if a house shows a positive moving balance. Therefore, I make the following assumption: A negative moving balance implies high fluctuation of inhabitants, which furthermore, implies that many households want to move away and thus, make a moving decision. In the other case, households are likely to stay in their current dwelling, if the moving balance is positive. Here, there is no moving decision. For this analysis, dwelling satisfaction, area satisfaction, neighbourhood satisfaction and total satisfaction get considered. The data from *microm* and *F+B research and consultancy* get merged with the data from *InWIS* on the level of living quarters. The respective sample sizes after dropping the missings are: N_{C1} = 3,727; N_{C2} = 4,179; N_{C3} = 936; N_B = 1,992. The regression equation in terms of log odds is

$$ln\left(P(MovBal_i = 1)\right) = \beta_0 + \beta_1 DwelSatis_i + \beta_2 AreaSatis + \beta_3 NeighSatis_i + \beta_4 TotSatis_i \quad , \qquad (3)$$

where $ln(P(MovBal_i = 1))$ represents the log odds of attaining the probability of having a positive moving balance, $DwelSatis_i$ denotes dwelling satisfaction, $AreaSatis_i$ area satisfaction, $NeighSatis_i$ neighbourhood satisfaction and $TotSatis_i$ total satisfaction.

Finally, in order to measure the actual effects of the variables on dwelling prices, this part of the analysis shows a hedonic price estimation. Again, the data from *microm* and *F+B research and consultancy* get merged on the neighbourhood level, resuling in a sample size of N =45,859. Under consideration of the cluster analysis, the regression will be conducted for each cluster found in Analysis 2, as well as

for the entire city of Bochum by dropping all missings. This leads to the following subsamples: $N_{C1} =$ 12,779; $N_{C2} =$ 13,068; $N_{C3} =$ 3,216; $N_{C4} = 7,033$; $N_B =$ 36,096.

Due to the various different data types, the Likert scaled data as well as the cateogorical data get split into dummy variables, where each dummy represents a characteristic of the respective variable, e.g. "Debt Default Probability: very low" or "Sinus Milieu: Established". For the analysis, a Log-Level OLS approach is chosen, because it is rather implausible that the number of rooms or the construction year vary in percentage units. The regression equation looks as follows:

$$\begin{aligned} ln(P_{i,h}) =& \beta_0 + \beta_1 DS_h + \beta_2 NR_h + \beta_3 ConY_h + \beta_4 EI_h + \\ & \sum_{\pi}(\beta_\pi DebtDef_{\pi,i}) + \sum_{\delta}(\beta_\delta Stat_{\delta,i}) + \sum_{\eta}(\beta_\eta DomSinusM_{\eta,h}) + \epsilon_{i,h} \quad , \end{aligned} \tag{4}$$

where $ln(P_{i,h})$ denotes the natural logarithm of the dwelling prices, DS_h is the dwelling size, NR_h the number of rooms, $ConY_h$ the construction year, EI_h the endowment index, $DebtDef_{\pi,i}$ several dummies for the evaluation of debt default probability of household i, $Stat_{\delta,i}$ the respective status and $DomSinusM_{\eta,h}$ the different dominant Sinus Milieus. The dummies with the characteristic "very low" and the Sinus-Milieu "Adaptive Navigators" are used as references. The latter is chosen, because the group "Adaptive Navigators" has the smallest share in the city of Bochum. $\epsilon_{i,h}$ represents the error term.

4 Findings

In this section, the results of the various analyses are shown. Subsection 4.1 shows the descriptive statistics, 4.2 presents the results from the cluster analysis, 4.3 shows the outcome of the CHAID Decision Tree model and 4.4 discusses the output of the Hedonic Price Estimates. Finally, 4.5 closes with a discussion of the results.

4.1 Descriptive Statistics

Within the *microm* data set, the data does not capture the characteristics for each single party within a housing unit, but represents them as aggregates on the house level. For a basic understanding of Bochum's population characteristics, Figure (3) shows the relative frequencies for each of the five occurrences for the

variables "Debt Default Probability", "Status", "Share of Children" and "Share of Foreigners" in the city of Bochum.

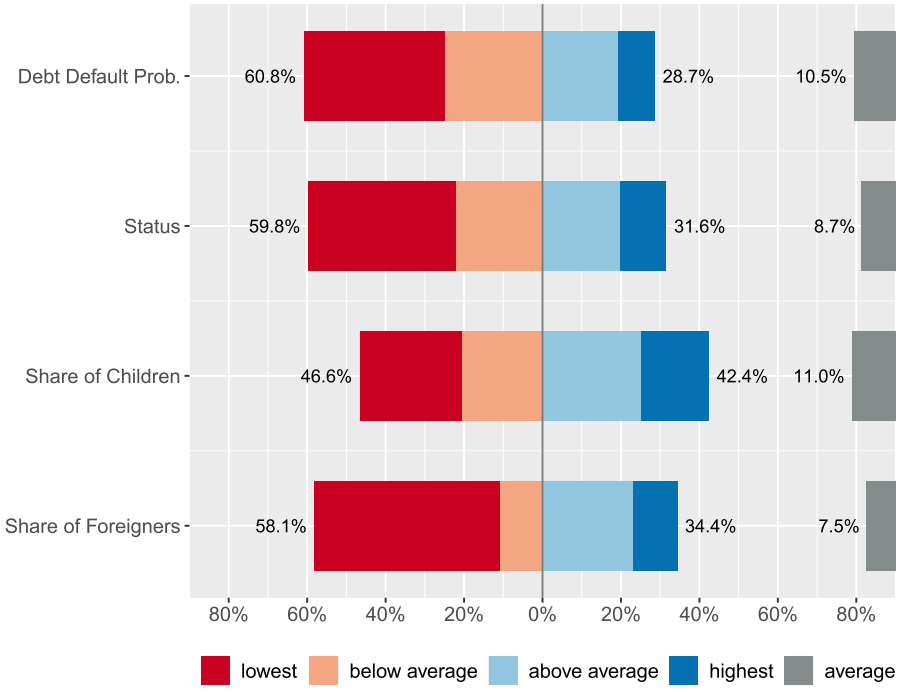

Figure 3: Relative Frequencies of each occurrence in different Likert scaled variables

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration

The graph shows that in 60.8% of all observed houses, Bochum's population has a Debt Default Probability below average and 28.7% above average. Furthermore, the plot shows that 59.8% of all houses have inhabitants with a Status below average and 31.6% of a Status above average. These variables suggest that roughly one third of Bochum's population may suffer under a higher probability of debt default even though almost two thirds are characterized as having a lower status. This could suggest implications for the labor market situation in Bochum, where one third of the "lower-Status" inhabitants may have a more permanent employment.

Figure (3) also shows that the Share of Children is almost equally distributed within Bochum. 46.6% of houses have a share of children below average whereas 42.4% of houses are characterized as having higher amounts of children compared to the average amount. Finally, 58.1% of housing units denote a Share of Foreigners below average and 34.4% a Share of Foreigners above average.

For a more detailed picture, Figure (4) and Figure (5) show the relative frequencies of the respective occurrences for the average age of household's heads and for the family structures in Bochum. Figure (4) shows that 32.19% of all houses show an average age of the household's head between 45 and 55 years. This is the largest category. There are also large shares of houses with an average age of the head between 55 and 65 (25.09%) and over 65 (24.67%). On the other hand, the share of the age of the household's head in the category below 35 years old is very small with 2.95%.

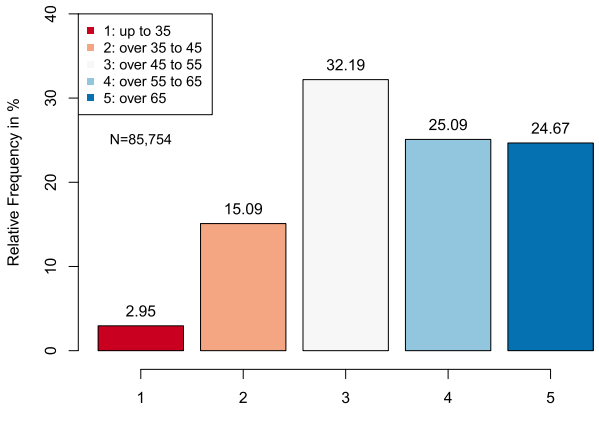

Figure 4: Relative Frequencies of the Average Age of Household's Head

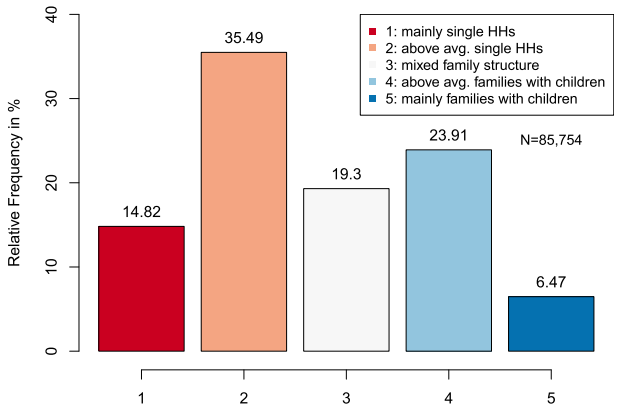

Figure 5: Relative Frequencies of Family structures

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration

Figure (5) shows that the share of single households outweigs the share of families with children. 50.31% of all houses show a share of single households above average or higher whereas only 30.38% have a share of families with children above average or higher [8]. Figure (4) and Figure (5) suggest that Bochum has a rather old population composition with mainly single households.

Beside the socioeconomic factors, the distribution of the Dominant Sinus Milieus and Primary Limbic Types within the city of Bochum are also important. Therefore, Figure (6) and Figure (8) show the relative frequencies of the various categorization types.

Figure (6) shows that the most Dominant Sinus Milieus in Bochum are Established (18.48%), Modern Mainstreamers (16.13%) and Traditionals (12.84%). According to Figure (1), they are all characterized by

[8] Couples without children are not included in the data

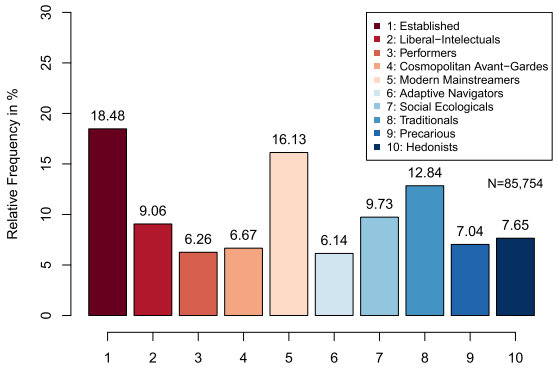

Figure 6: Dominant Sinus Milieus in Bochum

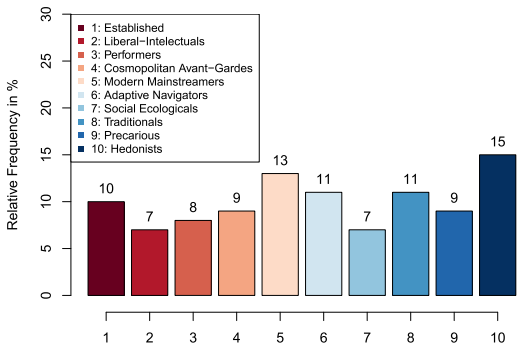

Figure 7: Dominant Sinus Milieus in Germany

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018); Sinus: Markt- und Sozialforschung (2018), own illustration

values of having and enjoying modernisation and individualisation by keeping a standard of living, status and ownership. Additionally, Established and Traditionals share the values of preserving tradition and only differ in social status. Established are characterized by middle to higher social status, whereas Traditionals have lower to middle social status. Modern Mainstreamers have a similar social status compared to Traditionals, but are not characterized by modernised tradition but rather by self-realisation, emancipation and authenticity. Other rather frequent Milieus are Social Ecologists (9.73%) and Liberal-Intelectuals (9.06%). Both share the same values as Modern Mainstreamers, but are characterized by higher Social Status. These five Milieus cover 66.24% of the entire city. The other five Milieus are also observable in Bochum, however their appearance accounts only for 33.76% and vary between 6.14% to 7.65%.

In comparison to Germany, the Dominant Sinus Milieus in Bochum show a different distribution across the population. Even though, some values are rather comparable between Bochum and Germany, like e.g. the share of Liberal-Intelectuals (9.06% compared to 7%), the share of Performers (6.26% compared to 8%) or the share of Precarious (7.04% compared to 9%), there are also some greater differences. Where Bochum is mainly inhabited by Established individiuals, this share in Germany is lower by 8.48%-points. On the other hand, the share of Hedonists is substantially lower than in entire Germany (7.65% compared to 15%). Also, the share of Adaptive Navigators in Bochum is almost equal to the half of the share in Germany (6.14% compared to 11%). These differences show that the city of Bochum is quite interesting for the analysis, since

the distribution of Dominant Sinus Milieus differs from the German numbers.

Figure (8) shows the relative frequencies of the Primary Limbic Types. Here, 42.69% of households are categorized as Harmonizers and 16.04% as Traditionalists. The former are characterized by a high level of social and family orientation as well as low status-orientation. The latter shows a low level of future-orientation. However, both share the desire for order and security. The third most frequency can be found in the Type of Performers (12.07%), who are characterized by a high level of performance and status orientation and ambition. These three types account for 70.8% of Bochum's population. The other frequencies vary between 2.41% (Adventurers) and 10.08% (Disciplinarians).

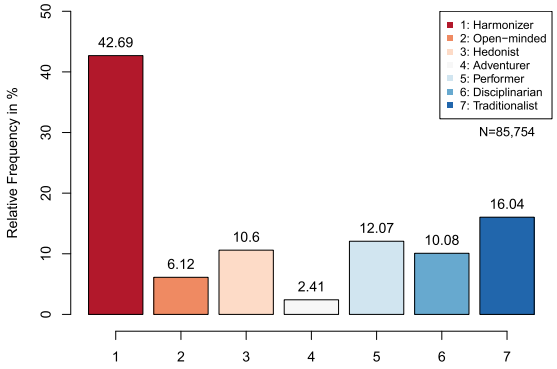

Figure 8: Primary Limbic Types

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration

Figure (9) shows the moving balance across all houses. The graph shows that 30.38% of all houses in Bochum show a negative moving balance and are thus, less attractive. On the other hand, there are 55.47% of all houses which show a positive balance and have less fluctuation. 14.15% of all houses are balanced.

In order to take a closer look at the housing characteristics, Table (1) shows the descriptive statistics of the data in Bochum from *F+B research and consultancy*[9]. The average monthly rent / price across all offers ($N = 8512$) is equal to 441.18 EUR. This is larger than the median of 400 EUR, which is an indicator for a right-skewed distribution, which is confirmed by the skewness value of 2.776. The prices vary between 50 EUR and 2,481 EUR. The mean of the dwelling size across all apartment offers in 2019 is equal to 65.01

[9]Table 1 only shows the characteristics for apartment rent. The other cases can be found in the appendix.

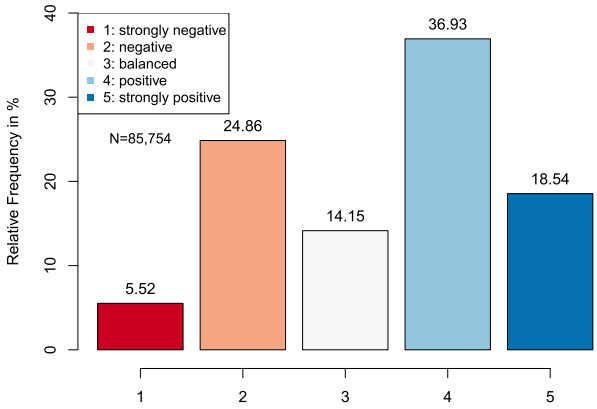

Figure 9: Relative Frequencies of the Moving Balance

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration

square metres. Similar to the prices, the distribution is right-skewed and varies between 12 and 491 square metres.

Furthermore, Table (1) shows that the average construction year of the observed houses was 1960. The construction years vary between 1753 and 2020, allowing the population of Bochum to inhabit rather old, but also rather new dwellings, both with a similar share with the median of 1961. The average number of rooms for all offers is equal 2.5, whereas the median lies at 2. This also implies a right-skewed distribution. The number of rooms vary between 1 and 12.

The self-constructed E-Index is shown in the table. Across all offers it has a mean of 31.437 and varies between 0 and 94.915 [10]. If an apartment offer contained all keywords from the text mining approach, the score would be 100. The single keywords are listed below the E-Index and represent the relative frequency of appearance. Central Heating (0.583) and Cellar (0.343) show the highest average relative frequencies across all offers.

Finally, Table (2) shows the descriptive statistics for different satisfactions. The first three variables vary between 1 and 5, where 1 is denoted as highly unsatisfied and 5 as highly satisfied. 'Total Satisfaction' varies between 0 (highly unsatisfied) and 100 (highly satisfied). In the mean, all agents are rather satisfied with their living conditions ($\bar{x} > 3$) and their total satisfaction ($\bar{x} = 69.288$). Especially, the satisfaction with the

[10]The E-Index controls for doublings, which means that e.g. all kinds of heating systems have the same influence on the index, the same holds for balcony/loggia. The frequencies were added up for the index, but are shown separately in Table 1.

Table 1: Descriptive Statistics for object types "apartment rent" from *F+B research and consultancy*

	n	mean	sd	median	min	max	skew	kurtosis	se
Price	8,512	441.188	185.667	400	50	2,481	2.776	13.274	2.012
Dwelling Size	8,512	65.014	22.084	62	12	491	2.625	24.966	0.239
Construction year	6,934	1,960.254	26.079	1,961	1,753	2,020	-0.467	1.563	0.313
No. of Rooms	8,466	2.501	0.820	2	1	12	0.785	5.134	0.009
E-Index	8,512	31.437	15.162	32.577	0	94.915	0.038	-0.038	0.164
Central Heating	8,512	0.583	0.493	1	0	1	-0.337	-1.887	0.005
Cellar	8,512	0.343	0.475	0	0	1	0.659	-1.565	0.005
Balcony/Patio	8,512	0.272	0.445	0	0	1	1.024	-0.951	0.005
garage-/parking lot	8,512	0.214	0.410	0	0	1	1.397	-0.047	0.004
bright	8,512	0.246	0.431	0	0	1	1.179	-0.609	0.005
quiet	8,512	0.182	0.386	0	0	1	1.644	0.703	0.004
garden	8,512	0.113	0.317	0	0	1	2.439	3.947	0.003
top floor	8,512	0.120	0.325	0	0	1	2.344	3.496	0.004
elevator	8,512	0.107	0.309	0	0	1	2.540	4.452	0.003
self-contained central heating	8,512	0.119	0.324	0	0	1	2.353	3.536	0.004
built-in kitchen	8,512	0.114	0.318	0	0	1	2.429	3.902	0.003
storeroom	8,512	0.117	0.321	0	0	1	2.384	3.686	0.003
exclusively/high-class/luxurious	8,512	0.093	0.291	0	0	1	2.799	5.835	0.003
loggia	8,512	0.103	0.304	0	0	1	2.609	4.808	0.003
renovated	8,512	0.098	0.297	0	0	1	2.702	5.301	0.003
floor heating	8,512	0.034	0.180	0	0	1	5.176	24.790	0.002
maisonette	8,512	0.054	0.227	0	0	1	3.929	13.438	0.002
first time use	8,512	0.043	0.204	0	0	1	4.477	18.046	0.002
new building	8,512	0.025	0.156	0	0	1	6.081	34.979	0.002
green area	8,512	0.038	0.191	0	0	1	4.836	21.387	0.002

Source: F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own predictions

Area is best on the mean with a value of 3.966. Here, the values across neighbourhoods vary between 3.184 and 4.669, where other kinds of satisfactions vary between 2.184 and 4.525 (Dwelling Satisfaction) and 2.526 and 4.669 (Neighbourhood Satisfaction).

Table 2: Descriptive Statistics for Satisfaction Types

	n	mean	sd	median	min	max	skew	kurtosis	se
Satisf. overall Apartment	50	3.784	0.377	3.836	2.184	4.525	-1.873	5.880	0.053
Satisf. Neighbourhood	50	3.743	0.348	3.778	2.526	4.669	-0.551	2.167	0.049
Satisf. Area	50	3.966	0.261	3.949	3.184	4.669	-0.022	1.062	0.037
Total Satisfaction	50	69.288	5.510	69.520	50.110	78.650	-0.818	1.303	0.779

Source: InWIS - Institut für Wohnungswesen, Immobilienwirtschaft, Stadt- und Regionalentwicklung (2018), own predictions

4.2 Cluster Analyses with socioeconomic variables and indicators for psychological mindsets

Figure (10) shows the Scree Plot of the analysis. The vertical red line represents the so-called elbow criterion and stands for the appropriate amount of clusters for the city of Bochum, namely 4.

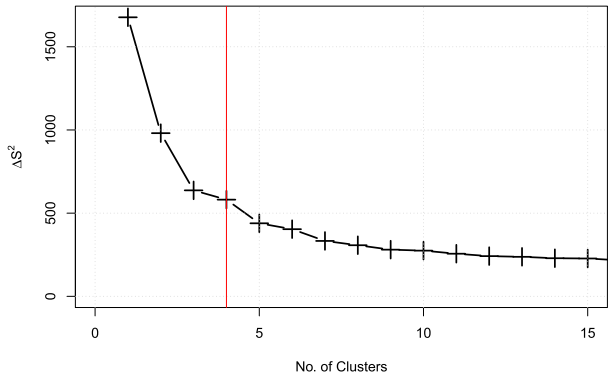

Figure 10: Scree plot of Cluster Analysis

Accordingly, Table (3) shows the average values of the agglomerative hierarchical cluster analysis of dissimilarity matrix D, which contains socioeconomic variables as well as indicators for psychological mindsets. The socioeconomic scores refer to the averages of the ordinal 5-scaled Likert data and rent scores are means of metric data. Cluster 1 can be characterized as follows: The cluster has a size of 9,468 housing units and shows the highest score of debt default probability (=4.86) compared to the entire city of Bochum (=3.40). The average value of the age of households' heads variable is equal to 2.83 and is the lowest value in comparison to the other clusters and the city (=3.32). It suggests that Cluster 1 is inhabited by a comparatively younger population. Furthermore, the share of foreigners (=3.71) is higher than the average share of foreigners in Bochum (=2.88) and in the other clusters. The share of children is right on average in Cluster 1 with a score of 2.53 (compared to 2.59 in Bochum). This suggests that on average, there are more single households than families with children in Cluster 1. Lastly, the average rent is lower than the average rent in Bochum (422.82 EUR compared to 452.09 EUR)

The barplots of the Sinus Milieus and Limbic Types show a different distribution as Figure (6) and Figure (8), which describe the Dominant Sinus Milieus and the Primary Limbic Types for the entire city of Bochum.

Table 3: Socioeconomic and value based characteristics for each cluster in Bochum

	Debt Default Prob.	Avg. Age of HH's Head	Share of Foreigners	Share of Children	Price	Cluster size
Cluster1	4.86	2.83	3.71	2.53	422.82	9,468
Cluster2	3.70	3.10	3.07	2.32	440.26	16,682
Cluster3	2.48	4.84	1.74	2.00	454.18	3,845
Cluster4	2.44	3.46	2.47	3.06	481.55	15,826
Bochum	3.40	3.32	2.88	2.59	452.09	45,821

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own predictions

Especially, the Hedonist group in the Sinus Milieus is strongly represented with 54.21%. The other frequent groups are Cosmopolitan Avant-Gardes (12.34%) and Precarious (12.32%). The other groups are either below 10% or close to 0% and thus, barely represented in Cluster 1. On the other hand, Figure (12) shows that Cluster 1 inhabits 31.71% of Open-minded, 17.78% of Adventurers, 17% of Disciplinarians and 16.43% of Hedonists. Thus, Cluster 1 is also characterized by a high share of lower-status individuals with barely varying intrinsic values.

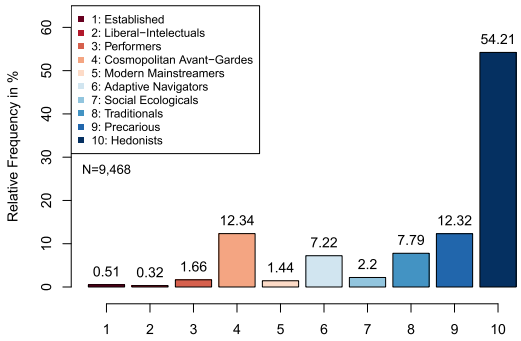

Figure 11: Dominant Sinus Milieus in Cluster 1

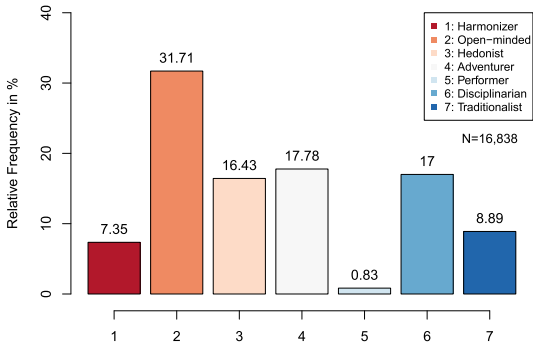

Figure 12: Primary Limbic Types in Cluster 1

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration/predictions

Cluster 2 can be characterized as follows: The cluster has a size of 16,682 housing units and shows a high score of debt default probability (=3.70) compared to the entire city of Bochum (=3.40) and Cluster 1 (=4.86). The average value of the age of households' heads variable is equal to 3.10 and still lower than in the entire city (=3.32). It suggests that Cluster 2 is also inhabited by a comparatively younger population, however with some more older people. Furthermore, the share of foreigners (=3.07) is high in comparison to the other clusters and the average score for Bochum (=2.88). The share of children is low in Cluster 2

with a score of 2.32 (compared to 2.59 in Bochum). This suggests that on average, there are almost as many single households as there are families with children. Lastly, the average rent is lower than the average rent in Bochum (440.26 EUR compared to 452.09 EUR).

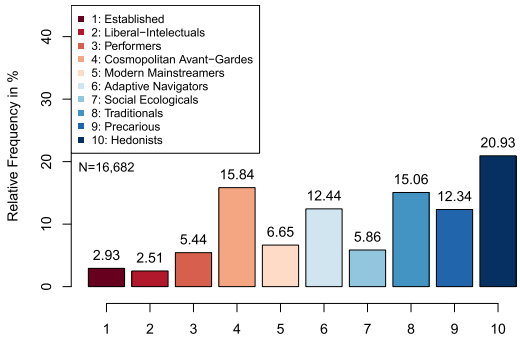

Figure 13: Dominant Sinus Milieus in Cluster 2

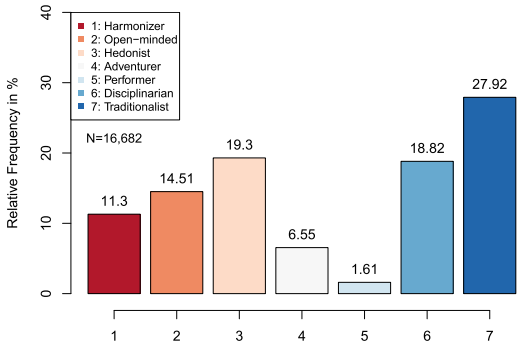

Figure 14: Primary Limbic Types in Cluster 2

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration/predictions

Figure (13) shows that Cluster 2 is mainly inhabited by Traditionals, Precarious and Hedonists ($\sum 48, 33\%$), where Hedonists have the highest share (20.93%). However, there are also many households, who refer to the Sinus-Milieu of Cosmopolitan Avant-Gardes (15.84%) and Adaptive Navigators (12.44%). Thus, the Dominant Sinus Milieus show a broader picture than Cluster 1, even though similar groups are rather dominant. According to Figure (1) these three Milieus are characterized by individuals with lower-status and varying intrinsic values with respect to tradition, modernisation and individualisation as well as re-orientation. Additionally, Figure (14) shows that Cluster 2 contains 27.92% of Traditionalists, who show a low level of future-orientation, but also 19.3% of Hedonists and 18.82% of Disciplinarians.

Cluster 3 can be characterized as follows: The cluster has the smallest size and contains 3,845 housing units. It shows the a low score of debt default probability (=2.48) compared to the entire city of Bochum (=3.40) and the other clusters. The average value of the age of households' heads variable is equal to 4.84 and is the highest value in comparison. This suggests that Cluster 3 contains a very high share of older individuals. Furthermore, the share of foreigners (=1.74) is lowest in comparison to the other clusters and the average score for Bochum (=2.88). The share of children is also lowest in Cluster 3 with a score of 2.00 (compared to 2.59 in Bochum). Lastly, the average rent is slightly higher in Cluster 3 compared to the

average rent in Bochum (454.18 EUR compared to 452.09 EUR).

Figure (15) and Figure (16) show the respective distributions within Cluster 3. Compared to the other clusters, the distribution of the Dominant Sinus Milieus shows that the highest share of inhabitants are characterized by Traditionals (50.95%) and 23.15% refer to Modern Mainstreamers. Figure (16) shows a similar picture by containing 50.85% of Traditionalists and 32.64% of Harmonizers.

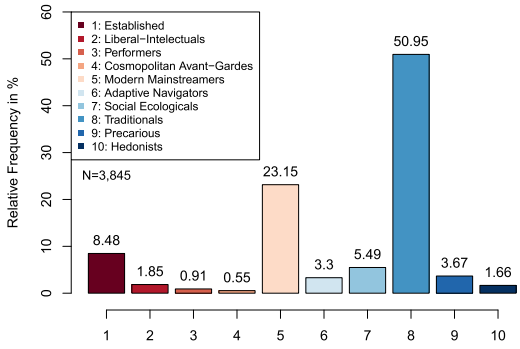

Figure 15: Dominant Sinus Milieus in Cluster 3

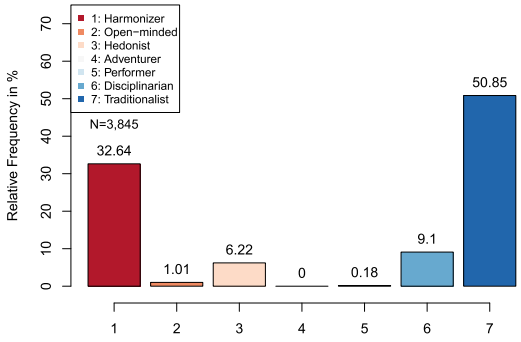

Figure 16: Primary Limbic Types in Cluster 3

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration/predictions

Finally, Cluster 4 can be characterized as follows: The cluster has a size of 15,826 housing units. It shows the the lowest score of debt default probability (=2.44) compared to the entire city of Bochum (=3.40) and the other clusters. The average value of the age of households' heads variable is equal to 3.46 and is higher than the average score in Bochum (=3.32). This suggests that Cluster 4 contains a higher share of older individuals. Furthermore, the share of foreigners (=2.47) is rather low in comparison to the other clusters and the average score for Bochum (=2.88). The share of children is highest in Cluster 4 with a score of 3.06 (compared to 2.59 in Bochum). Lastly, the average rent is highest in Cluster 3 compared to the average rent in Bochum (481.55 EUR compared to 452.09 EUR).

Figure (17) and Figure (18) show the respective distributions within Cluster 4. Compared to the other clusters, the distribution of the Dominant Sinus Milieus is similar to the average of the city of Bochum. It is dominated by Established (16.33%), Traditionals (14.13%) and Modern Mainstreamers (13.76%). However, there are also significant shares of the other groups. Figure (16) shows that even though, the Sinus Milieus show a rather diverse picture, Cluster 4 still inhabits 48.81% of Harmonizers, 15.23% of Performers and

14.24% of Traditionalists.

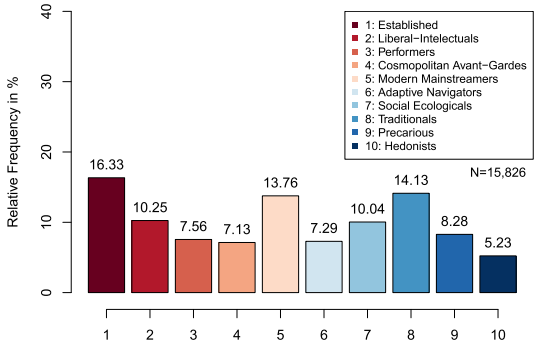

Figure 17: Dominant Sinus Milieus in Cluster 4

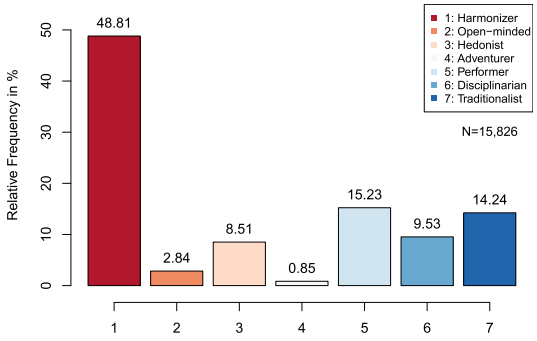

Figure 18: Primary Limbic Types in Cluster 4

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration/predictions

Figure (19) shows the aggregate cartographic visualization of the agglomerative hierarchical cluster analysis of dissimilarity matrix D, which contains socioeconomic variables and indicators for psychological mindsets. Cluster 1 (red) can mainly be found in the center of the city, Cluster 2 (orange) can be found in the East and West, Cluster 3 (lightblue) is more spread in the outer ring of the city and Cluster 4 can mainly be found in the South in the city with some areas East from the city center on a small strip towards the North.

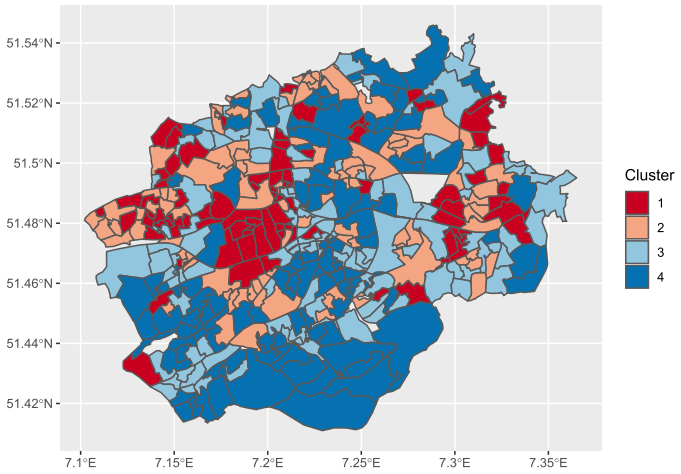

Figure 19: Cartographic visualization (aggregates) of the clusters of the city of Bochum with socioeconomic and psychological factors

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018), own illustration/predictions

As a final remark, the cluster analysis shows that not only socioeconomic variables, but also the respective Status and individual values and attitudes characterize the clusters in Bochum. In fact, they generate an image of an urban area, which can be split into four different homogenous areas, where e.g. high-status individuals live among other high-status individuals, but also conservative individuals seem to prefer living among other conservatives. The analysis shows that the city of Bochum is segregated in different dimensions. Therefore, Hypothesis 3 needs to be confirmed.

4.3 CHAID Decision Tree Approach to identify key drivers for housing rents

Figure (20) shows an examplary CHAID Decision Tree, where $N = 45,859$. The dependent variable is dwelling price/rent and the tree is cut after four levels (top-down). The graph shows that "Dwelling Size" is at the root node and has the highest impact on prices. This can be seen in the bar plots at the bottom of the decision tree, where there is a clear distinction between high shares of higher prices (high dwelling sizes at the left branches) and high shares of lower prices (low dwelling sizes at the right branches). For a more detailed price distinction, it is necessary to follow the different nodes. For example, to come to Terminal Node 13 (n=746), houses have the following characteristics (in a hierarchical order) and are grouped together: high dwelling size, the inhabiting households have low debt default probability, a Primary Limbic Type of 1 (=Harmonizer) or 3 (=Hedonist) and the house can be considered as old building. This group can be characterized by having a share of 75% of higher prices and a share of 25% of lower prices. Another example is Terminal Node 30 (n=640), where households inhabiting the observed dwellings have the following characteristics: low dwelling size, low number of rooms, new construction year and high Share of Foreigners. This group shows roughly a share of 96% of lower prices and only 4% of higher prices. The other groupings can be interpreted similarly.

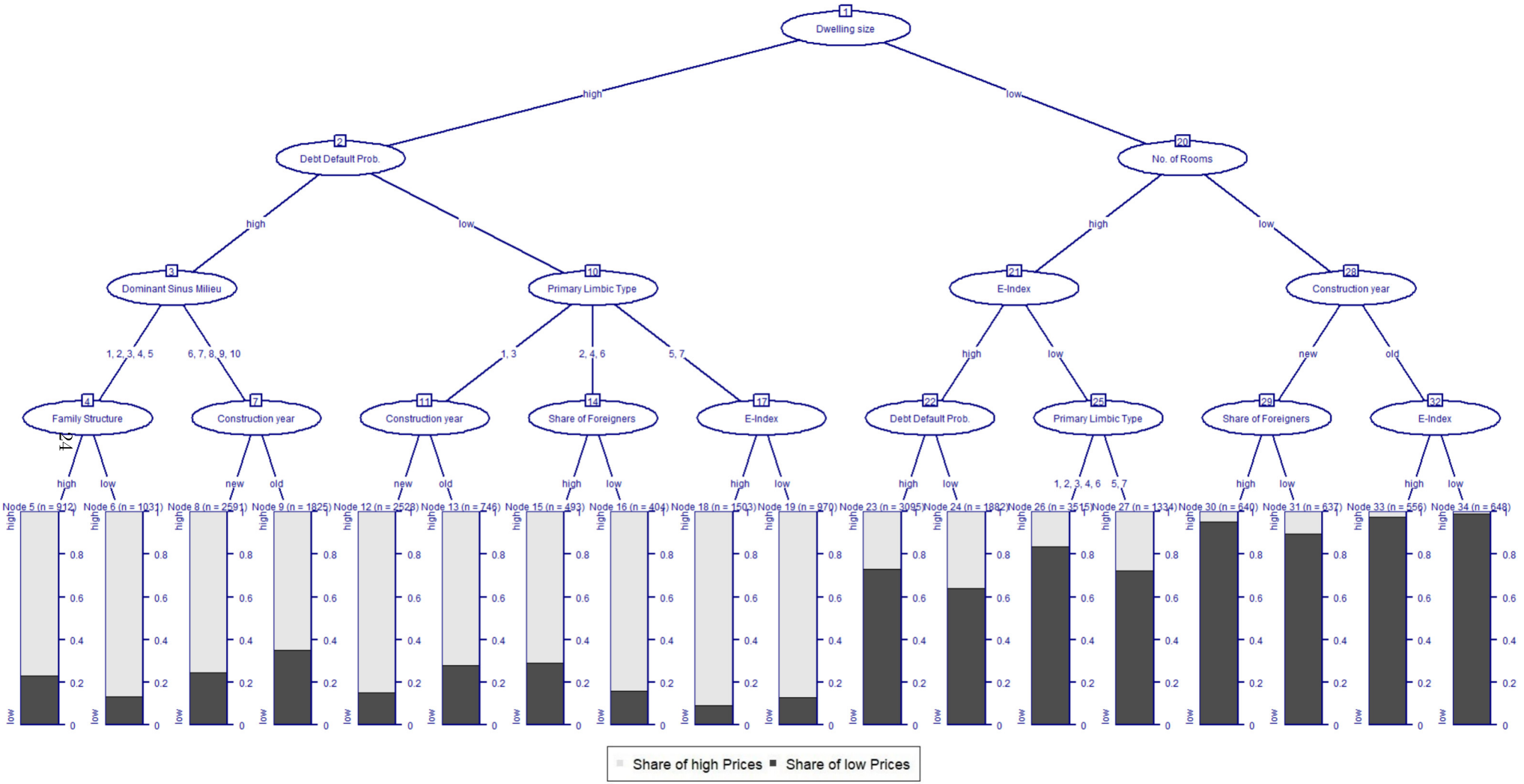

Figure 20: Examplary CHAID Decision Tree: Dwelling Prices/Rent as dependent variable

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018); F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own illustration/predictions

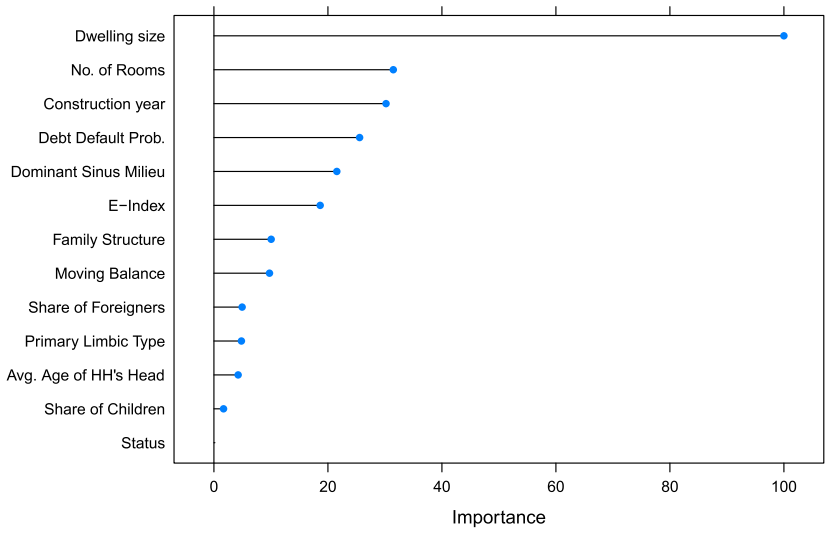

Figure 21: Importances on Bochum's Dwelling Rents

Source: (microm Micromarketing-Systeme und Consult GmbH, Neuss, 2018; F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH, 2018), own illustration/predictions

Concerning the importance of the variables for dwelling prices/rents, Figure (20) already provides some relevant information. However, the importance of the other variables becomes less distinctive as soon as I compare the variables with each other on the third and fourth level. Therefore, Figure(21) shows the importance scores generated from the model. Here, a value of 0 means that the variable has no effect on the distinction between higher and lower prices. If the value is equal to 100, the variable is a key identifier for the actual price split.

The graph shows that Dwelling Size (score = 100) and No. of Rooms (score = 31) have the highest importance for the dwelling rents/prices. However, Construction Year, Debt Default Probability and the Dominant Sinus Milieus also seem to influence rents/prices on a higher level. The lowest influence has the Share of Children and Status.

As a final step, the performance of the categorization model needs to be tested. In that context, a relative confusion matrix is shown in Table (4). The entries are relative shares with respect to high and low shares of prices. The table shows that the share of lower priced dwellings accounts predictively for 47.4% and actually for 49.9%. The share of higher priced dwellings is slightly higher than one third of all observations and

Table 4: Confusion Matrix for CHAID Decision Tree Model

		Prediction		
		high	**low**	**total**
Reference	**high**	TP 40.9	FN 9.2	50.1
	low	FP 11.7	TN 38.2	49.9
	total	52.6	47.4	100.0

Note: TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018); F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own illustration/predictions

accounts for 52.6% in the predictions and for 50.1% in the actual data. With the entries of Table (4), Table (5) shows the model performance. The indicator "Accuracy" shows that 79.1% of the classifier has been correct. Accordingly, the Misclassification is equal to 20.9% in this modelling framework. The True Positive Rate shows a value of 81.6%, which means that if the dwelling prices are high, the model predicts also high with a probability of more than 80%. Similar is the interpretation of the True Negative Rate, where actual low prices get predicted to be low by 76.6%.

Table 5: Performance Indicators for CHAID Decision Tree Model

Indicator	Determination Formula	Score
Accuracy	$\frac{TP+TN}{\sum}$	0.791
Misclassification	$\frac{FP+FN}{\sum}$	0.209
True Positive Rate	$\frac{TP}{\sum_{reference}^{high}}$	0.816
False Negative Rate	$\frac{FP}{\sum_{reference}^{low}}$	0.234
True Negative Rate	$\frac{TN}{\sum_{reference}^{low}}$	0.766

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018); F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own predictions

Finally, Figure (22) shows the different Accuracy scores by using the method of repeated cross-validation. I built a grid of the parameters and tested all the permutations considering the different significance levels for merging the various groups. The graph shows that the accuracy levels differ only on a minor basis. When $\alpha \in [0.001; 0.01; 0.05]$, the accuracy is almost identical at the beginning of the merging thresholds. By going one 0.01-unit forward, the accuracy drops from 0.7913 to 0.7908. Then $\alpha = 0.001$ remains constant, where the curves for $\alpha = 0.01$ and $\alpha = 0.05$ keep dropping slightly, however, the former remains on a higher level. Since the results are similar to each other, it implies that the model results are quite robust.

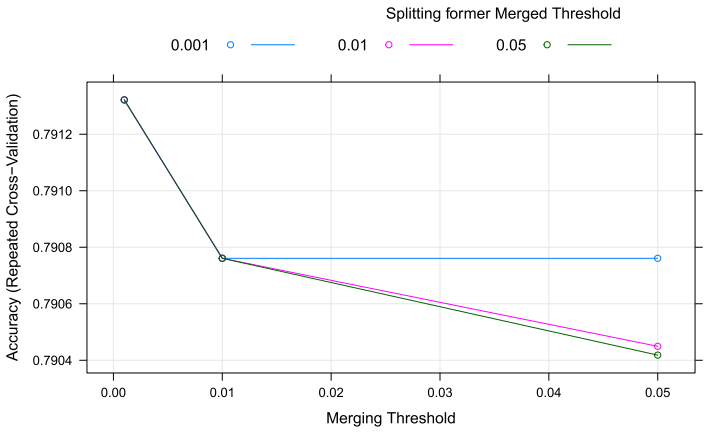

Figure 22: Accuracy Scores of CHAID Model using different significance levels

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018); F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own illustration/predictions

As a final remark, the CHAID Decision Tree Model shows that not only housing characteristics play a substantial role for the price determination, but also individual characteristics like the Dominant Sinus Milieu and the Debt Default Probability. The actual effect sizes are determined in the regression analysis in section 4.4.

4.4 Effects of Satisfaction on Moving Decisions and Hedonic Price Estimates

Table (6) shows the log odds coefficients of various satisfactions on the dummy proxy variable 'Moving Balance' for each cluster as well as for the entire city of Bochum. The coefficients are mainly significant on

the 1%-level, except for neighbourhood satisfaction. The output shows that especially the satisfaction with one's dwelling has the largest positive effect on the odds of living in a house with positive 'Moving Balance'. For every one unit change in 'Dwelling Satisfaction', the log odds of living in a house with positive Moving Balance increase between 2.432 (Cluster 2) and 4.134 (Cluster 4). The overall effect in Bochum is equal to 3.073.

The analysis shows that 'Area Satisfaction' has also positive effects on a positive 'Moving Balance'. For every one unit change in 'Area Satisfaction', the log odds of being in a house with positive Moving Balance increase between 0.954 (Cluster 3) and 2.252 (Cluster 4). The effect in Bochum is 1.045. 'Neighbourhood Satisfaction' shows different results. In Cluster 1, the effect of Neighbourhood Satisfaction is equal to -0.241 in log odds, however the coefficient is not significant. This means that increasing 'Neighbourhood Satisfaction' by one unit leads to a decrease in the log odds of being in a house with positive Moving Balance. A possible interpretation might be that households in Cluster 1 (with the cheapest dwellings on average) gain their satisfaction rather from their dwelling, area and total satisfaction and might be not interested in finding their peers. In the other clusters, the effect of neighbourhood satisfaction is positive with log odds coefficients of 0.018 (Cluster 2), 0.728 (Cluster 3), 0.987 (Cluster 4) and 0.122 (Bochum). However, the coefficients for Cluster 2 and Bochum are not significant and for Cluster 3 and 4 only on the 10% and 5%-level.

Finally, 'Total Satisfaction' has also positive effects on the living in a house with positive 'Moving Balance'. Here, the coefficients vary between 0.157 (Cluster 1) and 0.345 (Cluster 4). The log odds coefficient for Bochum is 0.188. Overall, the analysis shows that higher satisfaction levels lead to higher odds of being in houses with positive 'Moving Balances'. This implies that households in Bochum, who feel satisfied with their current living situation do not make a moving decision and rather stay in their current accomodation. Therefore, I can confirm Hypothesis 1 that satisfactions play an important role for the moving decision, since almost all types of satisfactions show higher odds of living in houses with positive Moving Balance, when increasing them by one unit.

Table 6: Logistic Regression Results of Satisfactions on Moving Balance as proxy for Moving Decisions

	Dependent variable:				
	‘Moving Balance‘				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Bochum
‘Dwelling Satisfaction‘	3.020***	2.432***	3.643***	4.134***	3.073***
	(0.255)	(0.281)	(0.532)	(0.868)	(0.164)
‘Area Satisfaction‘	1.294***	1.260***	0.954*	2.252***	1.045***
	(0.237)	(0.238)	(0.514)	(0.633)	(0.142)
‘Neighbourhood Satisfaction‘	−0.241	0.018	0.728*	0.987**	0.122
	(0.174)	(0.190)	(0.380)	(0.459)	(0.110)
‘Total Satisfaction‘	0.157***	0.190***	0.184***	0.345***	0.188***
	(0.016)	(0.017)	(0.034)	(0.054)	(0.010)
Constant	−19.695***	−22.024***	−23.547***	−42.421***	−22.833***
	(1.904)	(2.054)	(3.951)	(6.646)	(1.235)
Observations	3,727	4,179	936	1,992	10,834
Log Likelihood	−2,471.960	−2,415.640	−551.450	−526.986	−6,744.531
Akaike Inf. Crit.	4,953.920	4,841.280	1,112.900	1,063.971	13,499.060

Note: *p<0.1; **p<0.05; ***p<0.01

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018); InWIS - Institut für Wohnungswesen, Immobilienwirtschaft, Stadt- und Regionalentwicklung (2018), own predictions

Table (7) shows the coefficients of the Log-Level OLS regression for each cluster as well as for the entire city of Bochum. 'Dwelling Size', 'No. of Rooms', 'Construction year' and 'E-Index' are all significant on the 1%-level for each Cluster as well as for Bochum and have all positive effects on average on the dwelling prices. In Cluster 1, the effect of 'Dwelling Size' is largest: When 'Dwelling Size' increases by 1 unit, the dwelling rent/price increases on average by 1.18%. The lowest increase is shown in Cluster 3, where increasing 'Dwelling Size' by 1 unit leads to a price increase of 1.04%. The effect sizes of these four variables are largest for the 'No. of Rooms', meaning that a 1-unit increase in 'No. of Rooms' leads to price increases of 4.34% in

Cluster 1 and up to 9.50% in Cluster 3. For Bochum, this effect is equal to 5.00%.

The effects for 'Construction year' and 'E-Index' are smaller than for 'Dwelling Size' or 'No. of Rooms'. On average, dwellings show higher prices, if 'Construction year' increases by 1 unit. The highest effect can be seen in Cluster 3 with an increase of 0.18%. The effects of 'E-Index' are similar across the clusters and Bochum. On average, the price increases by roughly 0.24%, if 'E-Index' increases by 1 unit.

The dummies for 'Debt Default Probability' show that higher 'Debt Default Probability' leads to lower dwelling prices. Here, the reference category is set to the characteristic with the lowest value, in all cases. However, some coefficients are not significant in these estimations. Individuals with a very high 'Debt Default Probability' compared to a very low 'Debt Default Probability' (reference) show lower dwelling prices between 0.46% (in Cluster 4, but insignificant) and 5.51% (in Cluster 3), on average. For Bochum itself, the price decrease can be seen step-by-step by increasing the respective values for 'Debt Default Probability': low compared to very low shows a decrease of 0.76% and very high compared to very low shows a decrease of 2.23%.

The reference category for 'Status' is, in all cases, the lowest category 'Status very low'. Here, the results are all significant with some exceptions in Cluster 3. The coefficients show that increasing 'Status' leads to increasing dwelling prices. If households have a high 'Status', the prices increase between 2.20% in Cluster 3 and 7.25% in Cluster 2. If the Status is considered to be 'very high', the dwelling prices are even more expensive than in the reference category. Here, the coefficients vary on average between 4.90% in Cluster 4 and 16.70% in Cluster 2, compared to the reference category.

For the 'Dominant Sinus Milieus', the reference category is set to "Adaptive Navigators" across the clusters and for Bochum, because their frequence is lowest in Bochum. The analysis shows that in Cluster 1,2 and in Bochum, there are specifically two 'Dominant Sinus Milieus', which show a price increase compared to the reference category, namely Traditionals and Cosmopolitan Avant-Gardes. In Cluster 1, these two groups show higher prices than the reference category by 2.16% (Cosmopolitan Avant-Gardes) and 2.62% (Traditionals). The other groups show lower prices. Compared to 'Adaptive Navigators' the price decreases in Cluster 1 vary between -0.53% (Modern Mainstreamers) and -4.82% (Social Ecologicals). Cluster 2 shows a similar picture, however, the coefficients for 'Traditionals' and 'Cosmpolitan Avant-Gardes' are insignificant.

Table 7: OLS Log-Level Regression Results with Dominant Sinus Milieus

	Dependent variable:				
	Dwelling Rent/Price				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Bochum
‘Dwelling Size‘	0.0118^{***}	0.0114^{***}	0.0104^{***}	0.0108^{***}	0.0114^{***}
	(0.0001)	(0.0001)	(0.0003)	(0.0002)	(0.0001)
‘No. of Rooms‘	0.0434^{***}	0.0509^{***}	0.0950^{***}	0.0445^{***}	0.0500^{***}
	(0.0034)	(0.0032)	(0.0063)	(0.0044)	(0.0020)
‘Construction year‘	0.0017^{***}	0.0005^{***}	0.0018^{***}	0.0015^{***}	0.0012^{***}
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.00004)
‘E-Index‘	0.0023^{***}	0.0022^{***}	0.0027^{***}	0.0023^{***}	0.0024^{***}
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
‘Debt Default very low‘	Ref.	Ref.	Ref.	Ref.	Ref.
‘Debt Default low‘	-0.0441^{***}	-0.0043	-0.0424^{***}	0.0028	-0.0076^{**}
	(0.0129)	(0.0051)	(0.0159)	(0.0139)	(0.0030)
‘Debt Default average‘	-0.0405^{***}	0.0107	-0.0473^{***}	0.0013	-0.0057
	(0.0129)	(0.0078)	(0.0157)	(0.0146)	(0.0039)
‘Debt Default high‘	-0.0393^{***}	0.0218^{***}	-0.0034	0.0147	0.0059
	(0.0127)	(0.0071)	(0.0152)	(0.0182)	(0.0037)
‘Debt Default very high‘	-0.0539^{***}	-0.0250^{***}	-0.0551^{***}	0.0046	-0.0223^{***}
	(0.0131)	(0.0088)	(0.0169)	(0.0272)	(0.0045)
‘Status very low‘	Ref.	Ref.	Ref.	Ref.	Ref.
‘Status low‘	0.0159^{***}	0.0283^{***}	-0.0124	0.0580^{***}	0.0170^{***}
	(0.0052)	(0.0054)	(0.0101)	(0.0122)	(0.0033)
‘Status average‘	0.0208^{**}	0.0211^{**}	0.0003	0.0779^{***}	0.0201^{***}
	(0.0082)	(0.0091)	(0.0162)	(0.0187)	(0.0054)
‘Status high‘	0.0689^{***}	0.0725^{***}	0.0220	0.0638^{***}	0.0655^{***}
	(0.0064)	(0.0084)	(0.0138)	(0.0202)	(0.0045)
‘Status very high‘	0.1651^{***}	0.1670^{***}	0.0813^{***}	0.0490	0.1612^{***}
	(0.0078)	(0.0172)	(0.0223)	(0.0359)	(0.0062)
Established	-0.0277^{***}	-0.0393^{***}	-0.0184	0.0592^{*}	-0.0300^{***}
	(0.0083)	(0.0109)	(0.0228)	(0.0312)	(0.0053)
‘Liberal Intelectuals‘	-0.0431^{***}	-0.0706^{***}	0.0346	-0.1357^{***}	-0.0453^{***}
	(0.0091)	(0.0117)	(0.0319)	(0.0437)	(0.0061)
Performers	-0.0415^{***}	-0.0615^{***}	-0.0418	0.0188	-0.0411^{***}
	(0.0099)	(0.0089)	(0.0439)	(0.0194)	(0.0060)
‘Cosmopolitan Avant-Gardes‘	0.0216^{**}	0.0067	0.0219	0.0736^{***}	0.0247^{***}
	(0.0099)	(0.0063)	(0.0532)	(0.0105)	(0.0047)
‘Modern Mainstreamers‘	-0.0053	-0.0136^{*}	-0.0138	0.0653^{***}	-0.0089^{*}
	(0.0084)	(0.0077)	(0.0209)	(0.0204)	(0.0049)
‘Adaptive Navigators‘	Ref.	Ref.	Ref.	Ref.	Ref.
‘Social Ecologicals‘	-0.0482^{***}	-0.0535^{***}	-0.0335	0.0235	-0.0437^{***}
	(0.0090)	(0.0085)	(0.0243)	(0.0178)	(0.0054)
Traditionals	0.0262^{***}	0.0025	-0.0125	0.0276^{**}	0.0087^{**}
	(0.0085)	(0.0063)	(0.0202)	(0.0115)	(0.0043)
Precarious	-0.0097	-0.0439^{***}	-0.0250	0.0164	-0.0242^{***}
	(0.0095)	(0.0067)	(0.0270)	(0.0108)	(0.0048)
Hedonists	-0.0225^{**}	-0.0354^{***}	0.0222	0.0184^{**}	-0.0212^{***}
	(0.0106)	(0.0060)	(0.0326)	(0.0090)	(0.0043)
Constant	1.8223^{***}	4.1314^{***}	1.4421^{***}	2.2564^{***}	2.7031^{***}
	(0.1364)	(0.1293)	(0.3047)	(0.1662)	(0.0790)
Observations	12,779	13,068	3,216	7,033	36,096
R^2	0.7160	0.6742	0.6917	0.6782	0.6952
Adjusted R^2	0.7156	0.6737	0.6897	0.6773	0.6950
Residual Std. Error	0.2057 (df = 12757)	0.1882 (df = 13046)	0.1914 (df = 3194)	0.1864 (df = 7011)	0.1957 (df = 36074)
F Statistic	$1{,}531.7380^{***}$ (df = 21; 12757)	$1{,}285.7370^{***}$ (df = 21; 13046)	341.3095^{***} (df = 21; 3194)	703.7099^{***} (df = 21; 7011)	$3{,}917.3740^{***}$ (df = 21; 36074)

Note: *p<0.1; **p<0.05; ***p<0.01

Source: microm Micromarketing-Systeme und Consult GmbH, Neuss (2018); F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own predictions

The price decreases for the other groups compared to the reference group vary between -1.36% (Modern Mainstreamers) and -7.06% (Liberal Intelectuals).

In Cluster 4, there is a different picture. Except for the group of 'Liberal Intelectuals', all groups within the 'Dominant Sinus Milieus' show price increases compared to the reference group of 'Adaptive Navigators'. However, not all coefficients are significant. The values vary between 1.64% (Precarious and insignificant) and 7.36% (Cosmopolitan Avant-Gardes). The only group, who shows lower prices than the reference category is the group of 'Liberal Intelectuals', who show a price decrease in comparison of -13.57%. In Cluster 3, all coefficients are insignificant, which might be due to the smallest amount of observations.

As a final remark, the constant values in the regressions refer to different 'base' prices. Furthermore, the hedonic price estimations show different effect sizes across the clusters and the entire city of Bochum, where especially Status shows great impact. This is an indicator that households of different population groups assign different values to various aspects. Accordingly, the price determination differs, as well. With respect to these estimations as well as the CHAID Decision Tree approach, I find that especially the housing characteristics are responsible for the clear distinction between higher priced and lower priced dwellings. However, both analyses showed there is also a correlation between the renters' Debt Default Probability and Status[11], for which reason I can partly confirm hypothesis H2. It can be confirmed in a way that in different clusters, and thus with interpersonal similarity among households, there are different valuations of the explanatory variables. However, it can be rejected in a way that mainly housing characteristics drive the price determination. Here, there is the necessity for some further analyses.

5 Conclusion

This paper investigates empirically the population structure and the effects on the housing market in the city of Bochum in Germany. The purpose of the analyses is to gain insights on segregation patterns in Bochum, but also on how socioeconomic factors, psychological indicators/values like the so-called Sinus Milieus and Limbic Types and residential satisfactions affect urban mobility and market prices within different clusters and within Bochum as a whole. Accordingly, several different empirical approaches like cluster analyses, a

[11]There is the possibility that not only debt default probability and status have an effect on rents, but also that rents influence status and debt default probability. The potential endogeneity problem here provides, at least, information on the interconnection of these variables, but not on any causal effect.

CHAID decision tree and regression models have been chosen.

The analysis shows that Bochum can be defined by four clusters, as the outcome of the Scree Plot suggests. The clusters are mainly located in the Center (Cluster 1), in the West and in the East (Cluster 2), in the outer ring of the city center (Cluster 3) and in the South (Cluster 4). Cluster 1 is mainly inhabited by individuals with a very high debt default probability, which is a rather young and contains the highest amount of foreigners, an average share of children and the lowest rent. Furthermore, Cluster 1 inhabits mainly individuals, who are assigned to be Hedonists *(Sinus Milieus)* and Open-minded *(Limbic Types)*. Cluster 2 is similar to Cluster 1, however, the debt default probability is second-highest, the population is older and contains less shares of foreigners as well as less shares of children. The overall rent is slightly lower than the average in Bochum. Furthermore, the inhabiting households are mainly characterized by Traditionalists, Precarious and Hedonists *(Sinus Milieus)*. Cluster 3 shows a low debt default probability with the oldest population and the lowest share of foreigners and children. The overall rent is comparable to the price average of the city. Here, mainly Traditionals and Modern Mainstreamers *(Sinus Milieus)* as well as Traditionalists and Harmonizers *(Limbic Types)* can be found. Finally, Cluster 4 is characterized by the lowest debt default probability and an average age of households' heads. The share of foreigners is lower than the average in Bochum and the share of children is higher. In this cluster, the rent is most expensive. Within the Sinus Milieus, the cluster is quite diverse. Concerning the *Limbic Types*, the cluster contains mainly Harmonizers. These results imply that the city of Bochum shows patterns of segregation, where the different clusters can be characterized by different socioeconomic factors and prices as well as psychological indicators like the so-called Sinus-Milieus and Limbic Types. Policy makers can use more selective measures for the urban development based on information about individual values and needs in accordance with socioeconomic endowments. Hypothesis 3 can be confirmed.

The second research question refers to the importance of various variables on dwelling prices within the city of Bochum. The main separator for dwelling prices in Bochum is dwelling size. According to the CHAID decision tree, there is a clear distinction between lower and higher prices depending on lower or higher dwelling sizes. According to the importance scores for prices, the analysis shows that not only dwelling size and the number of rooms have a large importance score, but also debt default probability and the so-called Sinus

Milieus. An interpretation of this result would be that a household's income as well as values and status play a substantial role, where individuals prefer to live, when taking dwelling prices as measurement. This also may be an indicator on how satisfied individuals are within their current neighbourhood and if they prefer to stay or move elsewhere. The categorization according to the CHAID decision tree shows an accuracy of 79.1%.

In accordance with this machine learning approach, I conduct a hedonic price estimation as log-level OLS regression for all clusters and for the city of Bochum, for the fourth research question. Here, the samples are chosen from the cluster analysis and the variables of interest are based on the CHAID Decision Tree. The regression results show that all continuous variables concerning housing characteristics, which were identified by the CHAID algorithm, are significant on the 1%-level and influence rents positively. Here, the highest impact on prices can be found in the variable 'No. of Rooms', where households with one more room show higher prices of up to 9.50% on average, c.p. The Likert scaled data show also interesting effects. With increasing Debt Default Probability, the prices become lower. Furthermore, Status shows the largest price spread, where high status individuals have higher dwelling rents of up to 16.70% compared to low status. Overall, the effect sizes of all variables of interest differ across clusters leading to the conclusion that in different clusters, there are also different weights for the price determination. This might be an indicator that households in the various clusters attach importance to different attributes. The CHAID algorithm as well as the hedonic price estimation showed that dwelling characteristics do have a positive effect on dwelling rents, but debt default probability and status show also high correlations with the price. The actual effect is not measurable due to potential endogeneity, for which reason I can only confirm Hypothesis 2, partly.

In order to analyse the third research question, I conducted a logistic regression on a dummy variable for moving balance of each house. If the variable for moving balance is equal to 0, the balance is negative and implies more fluctuation in that particular house. Contrarily, if the variable is equal to 1, the balance is positive and implies that many households stay in the current dwelling. The analysis shows that in all clusters, found in the cluster analysis, as well as in the entire city of Bochum, almost all coefficients of the satisfaction types show higher odds of observing a house with positive moving balance. This means that individuals prefer to live in the current dwelling, if they have a high degree of residential satisfaction. The regression results

show that especially dwelling satisfaction has the largest impact on staying in the neighbourhood. Area satisfaction shows the second highest effect. Accordingly, policy makers should ensure that their tenants are primary satisfied with their living unit and secondary, with their surroundings. Hypothesis 1 can be confirmed.

The findings in this paper suggest that not only the 'hard facts' like socioeconomic and housing characteristics have an influence on the clusters or on the dwelling rents in an urban area, but also psychological indicators/values, e.g. captured by the so-called Sinus Milieus and Limbic Types. Furthermore, different kinds of satisfactions may have an effect on the affiliation to a social group, on hedonic price estimations and on urban mobility.

References

Atkinson, R. (2016). Limited exposure: Social concealment, mobility and engagement with public space by the super-rich in london. *Environment and Planning A: Economy and Space*, 48(7):1302–1317.

Baker, E., Mason, K., and Bentley, R. (2015). Measuring housing affordability: A longitudinal approach. *Urban policy and research*, 33(3):275–290.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Harvard university press.

Clark, W. A. (1992). Residential preferences and residential choices in a multiethnic context. *Demography*, 29(3):451–466.

Clark, W. A. and Coulter, R. (2015). Who wants to move? the role of neighbourhood change. *Environment and Planning A*, 47(12):2683–2709.

F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018). Data on housing characteristics and prices for bochum.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2):117–140.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4):857–871.

Guilford, J. P. (1950). Creativity. *American psychologist*, 5(9):444.

Haeusel, H.-G. (2011). *The scientific Foundation of the Limbic Approach*. Gruppe Nymphenburg Consult AG.

Hansen, P. and Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical programming*, 79(1):191–215.

InWIS - Institut für Wohnungswesen, Immobilienwirtschaft, Stadt- und Regionalentwicklung (2018). Data on residential satisfaction in bochum.

Ioannides, Y. M. (2011). Neighborhood effects and housing. In *Handbook of Social Economics*, volume 1, pages 1281–1340. Elsevier.

Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly journal of economics*, 120(3):963–1002.

Malik, A., Crooks, A., Root, H., and Swartz, M. (2015). Exploring creativity and urban development with agent-based modeling. *Journal of Artificial Societies and Social Simulation*, 18(2):1–12.

Malmberg, B., Andersson, E., and Östh, J. (2013). Segregation and urban unrest in sweden. *Urban geography*, 34(7):1031–1046.

Malpezzi, S. (2002). Hedonic pricing models: a selective and applied review. *Housing economics and public policy*, pages 67–89.

microm Micromarketing-Systeme und Consult GmbH, Neuss (2018). Data on socioeconomic characteristics of bochum's population.

Musterd, S. (2005). Social and ethnic segregation in europe: Levels, causes, and effects. *Journal of urban affairs*, 27(3):331–348.

Musterd, S., Marcińczak, S., Van Ham, M., and Tammaru, T. (2017). Socioeconomic segregation in european capital cities. increasing separation between poor and rich. *Urban Geography*, 38(7):1062–1083.

Nolan, B., Salverda, W., Checchi, D., Marx, I., McKnight, A., Tóth, I. G., and van de Werfhorst, H. G. (2014). *Changing inequalities and societal impacts in rich countries: thirty countries' experiences*. OUP Oxford.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.

Reuschke, D. and Weck, S. (2013). Residential segregation of turkish migrants in the ruhr area - reasons, patterns and policies. In *The Economies of Urban Diversity*, pages 191–215. Springer.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York, Norton.

Schwarz, N. and Ernst, A. (2009). Agent-based modeling of the diffusion of environmental innovationsan empirical approach. *Technological forecasting and social change*, 76(4):497–511.

Shrestha, R., Flacke, J., Martinez, J., and Van Maarseveen, M. (2016). Environmental health related socio-spatial inequalities: identifying hotspots of environmental burdens and social vulnerability. *International journal of environmental research and public health*, 13(7):691.

Sinus: Markt- und Sozialforschung (2018). `https://www.sinus-institut.de/en/sinus-solutions/sinus-milieus/`. Accessed: 2020-12-31.

Sirmans, G. S., MacDonald, L., Macpherson, D. A., and Zietz, E. N. (2006). The value of housing characteristics: a meta analysis. *The Journal of Real Estate Finance and Economics*, 33(3):215–240.

Sirmans, S., Macpherson, D., and Zietz, E. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, 13(1):1–44.

Smets, P. and Salman, T. (2008). Countering urban segregation: Theoretical and policy innovations from around the globe. *Urban Studies*, 45(7):1307–1332.

Speare, A. (1974). Residential satisfaction as an intervening variable in residential mobility. *Demography*, 11(2):173–188.

Zanna, M. P., Goethals, G. R., and Hill, J. F. (1975). Evaluating a sex-related ability: Social comparison with similar others and standard setters. *Journal of Experimental Social Psychology*, 11(1):86–93.

Appendix

Table 8: Descriptive Statistics for the data type "House Purchase"

	n	mean	sd	median	min	max	skew	kurtosis	se
Price	174	377,922.500	136,790.300	356,674.500	109,900	1,150,000	1.480	4.949	10,370.050
Dwelling Size	174	156.052	75.960	138.500	75	640	3.745	17.694	5.759
Construction year	154	1,985.058	36.206	1,993.500	1,895	2,020	-0.777	-0.533	2.918
No. of Rooms	171	5.608	2.438	5	3	20	2.497	8.478	0.186
E-Index	174	27.653	12.616	27.352	0	57.949	-0.034	-0.581	0.956
Central Heating	174	0.351	0.479	0	0	1	0.621	-1.624	0.036
Cellar	174	0.356	0.480	0	0	1	0.595	-1.656	0.036
Balcony/Patio	174	0.466	0.500	0	0	1	0.137	-1.993	0.038
garage-/parking lot	174	0.713	0.454	1	0	1	-0.932	-1.138	0.034
bright	174	0.178	0.384	0	0	1	1.668	0.786	0.029
quiet	174	0.178	0.384	0	0	1	1.668	0.786	0.029
garden	174	0.557	0.498	1	0	1	-0.229	-1.959	0.038
top floor	174	0.437	0.497	0	0	1	0.253	-1.947	0.038
elevator	174	0.017	0.131	0	0	1	7.354	52.376	0.010
self-contained central heating	174	0.023	0.150	0	0	1	6.311	38.048	0.011
built-in kitchen	174	0.138	0.346	0	0	1	2.082	2.348	0.026
storeroom	174	0.029	0.168	0	0	1	5.593	29.453	0.013
exclusively/high-class/luxurious	174	0.178	0.384	0	0	1	1.668	0.786	0.029
loggia	174	0.046	0.210	0	0	1	4.298	16.571	0.016
renovated	174	0.034	0.183	0	0	1	5.059	23.726	0.014
floor heating	174	0.339	0.475	0	0	1	0.674	-1.555	0.036
maisonette	174	0.006	0.076	0	0	1	12.964	167.034	0.006
first time use	174	0.247	0.433	0	0	1	1.162	-0.652	0.033
new building	174	0.236	0.426	0	0	1	1.235	-0.477	0.032
green area	174	0.080	0.273	0	0	1	3.058	7.396	0.021

Source: F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own predictions

Table 9: Descriptive Statistics for the data type "House Rent"

	n	mean	sd	median	min	max	skew	kurtosis	se
Price	156	773.538	427.188	550.500	262	2,251	0.874	-0.229	34.202
Dwelling Size	156	92	36.950	79	51	237	1.224	1.354	2.958
Construction year	142	1,974.979	35.359	1,972	1,822	2,020	-1.578	4.831	2.967
No. of Rooms	152	3.441	1.306	3	2	9	1.405	2.799	0.106
E-Index	156	26.914	13.192	26.987	0	63.598	0.129	-0.350	1.056
Central Heating	156	0.532	0.501	1	0	1	-0.127	-1.996	0.040
Cellar	156	0.250	0.434	0	0	1	1.144	-0.696	0.035
Balcony/Patio	156	0.513	0.501	1	0	1	-0.051	-2.010	0.040
garage-/parking lot	156	0.417	0.495	0	0	1	0.335	-1.900	0.040
bright	156	0.231	0.423	0	0	1	1.266	-0.400	0.034
quiet	156	0.103	0.304	0	0	1	2.595	4.764	0.024
garden	156	0.346	0.477	0	0	1	0.641	-1.600	0.038
top floor	156	0.135	0.342	0	0	1	2.121	2.513	0.027
elevator	156	0.160	0.368	0	0	1	1.834	1.374	0.029
self-contained central heating	156	0.051	0.221	0	0	1	4.030	14.330	0.018
built-in kitchen	156	0.167	0.374	0	0	1	1.772	1.146	0.030
storeroom	156	0.103	0.304	0	0	1	2.595	4.764	0.024
exclusively/high-class/luxurious	156	0.295	0.457	0	0	1	0.891	-1.214	0.037
loggia	156	0.122	0.328	0	0	1	2.291	3.268	0.026
renovated	156	0.077	0.267	0	0	1	3.145	7.942	0.021
floor heating	156	0.122	0.328	0	0	1	2.291	3.268	0.026
maisonette	156	0.013	0.113	0	0	1	8.578	72.042	0.009
first time use	156	0.109	0.313	0	0	1	2.486	4.205	0.025
new building	156	0.141	0.349	0	0	1	2.043	2.188	0.028
green area	156	0.051	0.221	0	0	1	4.030	14.330	0.018

Source: F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own predictions

Table 10: Descriptive Statistics for the data type "Apartment Purchase"

	n	mean	sd	median	min	max	skew	kurtosis	se
Price	700	168,825.000	144,836.700	119,950	12,000	1,800,000	3.482	25.838	5,474.314
Dwelling Size	700	88.964	88.052	75	16	1,650	10.683	158.391	3.328
Construction year	655	1,974.344	35.241	1,972	1,575	2,020	-2.398	23.894	1.377
No. of Rooms	695	3.252	3.667	3	1	71	12.317	194.341	0.139
E-Index	700	30.210	13.471	31.310	0	66.651	0.059	-0.020	0.509
Central Heating	700	0.556	0.497	1	0	1	-0.224	-1.953	0.019
Cellar	700	0.437	0.496	0	0	1	0.253	-1.939	0.019
Balcony/Patio	700	0.361	0.481	0	0	1	0.576	-1.671	0.018
garage-/parking lot	700	0.537	0.499	1	0	1	-0.149	-1.981	0.019
bright	700	0.309	0.462	0	0	1	0.827	-1.318	0.017
quiet	700	0.176	0.381	0	0	1	1.701	0.893	0.014
garden	700	0.269	0.444	0	0	1	1.042	-0.915	0.017
top floor	700	0.171	0.377	0	0	1	1.740	1.029	0.014
elevator	700	0.237	0.426	0	0	1	1.233	-0.479	0.016
self-contained central heating	700	0.110	0.313	0	0	1	2.488	4.194	0.012
built-in kitchen	700	0.117	0.322	0	0	1	2.376	3.650	0.012
storeroom	700	0.097	0.296	0	0	1	2.715	5.378	0.011
exclusively/high-class/luxurious	700	0.239	0.427	0	0	1	1.224	-0.502	0.016
loggia	700	0.110	0.313	0	0	1	2.488	4.194	0.012
renovated	700	0.106	0.308	0	0	1	2.559	4.556	0.012
floor heating	700	0.251	0.434	0	0	1	1.143	-0.693	0.016
maisonette	700	0.100	0.300	0	0	1	2.661	5.088	0.011
first time use	700	0.136	0.343	0	0	1	2.123	2.510	0.013
new building	700	0.159	0.366	0	0	1	1.865	1.482	0.014
green area	700	0.043	0.203	0	0	1	4.505	18.317	0.008

Source: F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH (2018), own predictions