Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

Betina Stefansen & Tuva Wille Kvarme

An Examination of Hyper-Locality Variables in Norwegian Real Estate Transactions – a Structural Equation Modeling Approach

Master's thesis in Economics and Business Administration
Supervisor: Are Oust & Randi Hammervold
May 2022

NTNU
Norwegian University of
Science and Technology

Betina Stefansen & Tuva Wille Kvarme

An Examination of Hyper-Locality Variables in Norwegian Real Estate Transactions – a Structural Equation Modeling Approach

Master's thesis in Economics and Business Administration
Supervisor: Are Oust & Randi Hammervold
May 2022

Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of
Science and Technology

Preface

This thesis was written as a part of our Master of Science in Economics and Business Administration at Norwegian University of Science and Technology (NTNU), during the spring semester of 2022. The thesis is written by Betina Stefansen and Tuva Wille Kvarme within the major Business Analytics, and has a scope of 30 ECTS. All work is independent and original.

We would like to express our gratitude to our supervisors at NTNU, Associate Professor Are Oust and Associate Professor Randi Hammervold for their valuable guidance and feedback during the semester. We thank Are for his helpful input of expert knowledge in the area of real estate economics and property valuation. The methodological part related to the statistical software LISREL was particularly challenging, and we would like to thank Randi for her expertise within statistics and excellent guidance through the methodological process. Furthermore, we would like to extend our thanks to the team at VIRDI AS for an extensive dataset and for answering questions regarding the dataset. Lastly, we want to thank friends and family for their help and support along the way.

Norwegian University of Science and Technology
Trondheim, May 2022

Betina Stefansen — Tuva Wille Kvarme

Abstract

The purpose of this thesis is to examine the impact of multiple aspects of location on apartment prices. This is elucidated by highlighting specific location related attributes, so-called hyper-locality features, valued by apartment buyers in Oslo. In addition, the socio-economic aspect of location is studied. The real estate market of Oslo is characterized by strong price growth. However, the growth differs between the city districts and the prices within the districts vary, indicating that location plays a substantial role on housing prices.

The rise of Business Intelligence systems and access to big data enables a more thorough investigation of location related to hyper-locality and socio-economic status. The dataset consists of 62,134 apartment transactions in the Norwegian capital Oslo, in the time period 2017 to 2021. In this thesis, hedonic methodology and Structural Equation Modeling (SEM) is combined to achieve a deeper understanding of the underlying effects of different location variables, by utilizing detailed locational unit feature data. A comprehensive model is composed of components regarding hyper-locality, travel costs, and the socio-economic status of the districts, in order to examine the importance of location on apartment prices.

The thesis finds that proximity to hyper-locality features explain differences in apartment prices within the Oslo districts. The SEM analysis disclosed three location factors, and concluded that distance to leisure activities is the most important factor. In addition, the thesis concludes that the socio-economic status of the district plays a significant role for people's willingness to pay when purchasing an apartment in Oslo. The findings are of great importance with regards to urban issues, planning, and site selection, especially from a real estate developer's perspective.

Sammendrag

Hensikten med denne oppgaven er å undersøke effekten av flere aspekter ved beliggenhet på leilighetspriser. Dette belyses ved å fremheve spesifikke stedsrelaterte attributter, såkalte hyperlokalitetsvariabler, som verdsettes av leilighetskjøpere i Oslo. I tillegg studeres det sosioøkonomiske aspektet ved beliggenhet. Eiendomsmarkedet i Oslo er preget av en sterk prisvekst, men prisutviklingen er imidlertid forskjellig mellom bydelene, samtidig som at prisene innenfor bydelene varierer. Dette indikerer at beliggenhet spiller en vesentlig rolle for boligprisene.

Fremveksten av Business Intelligence-systemer og tilgang til big data muliggjør en grundigere undersøkelse av beliggenhet knyttet til hyperlokalitet og sosioøkonomisk status. Datasettet består av 62 134 leilighetstransaksjoner i den norske hovedstaden Oslo, for tidsperioden 2017 til 2021. I denne oppgaven er hedonisk metode og strukturmodellering (SEM) kombinert for å få en dypere forståelse av de underliggende effektene av ulike beliggenhetsvariabler, gjennom bruk av detaljert beliggenhetsdata på enhetsnivå. En omfattende modell er satt sammen av komponenter knyttet til hyperlokalitet, reisekostnader og bydelenes sosioøkonomiske status for å forstå betydningen av beliggenhet på leilighetspriser.

Oppgaven finner at nærhet til hyperlokalitetsattributter er med på å forklare forskjeller i leilighetspriser innad i Oslos bydeler. SEM-analysen avdekket tre beliggenhetsfaktorer, og konkluderte med at avstand til fritidsaktiviteter er den viktigste faktoren. I tillegg konkluderer oppgaven med at bydelens sosioøkonomiske status spiller en signifikant rolle for kjøperes betalingsvillighet ved anskaffelse av leilighet i Oslo. Funnene er av stor betydning for spørsmål knyttet til byplanlegging og områdevalg, spesielt fra en eiendomsutviklers perspektiv.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Location, location, location" is a recognized and often emphasised slogan among real estate practitioners. In the standard practice of property valuation, the location of the property is considered the most important factor when determining the value of the dwelling. In real estate prospecting, walking distance to various amenities and services are often highlighted, but they tend not to be included as predictor variables in house pricing models (Heyman and Sommervoll, 2019). These can also be referred to as relative location or hyper-locality variables, where the latter formulation will be further used. Some examples of hyper-locality features are public transportation systems, schools, museums, libraries, and retail. Postcode and district dummies are typically used to deal with price differences across locations, and can be defined as absolute descriptions of residential locations (National Geographic, 2012). Models that focus solely on absolute location capture only average effects for a given postcode or district, and thus provide limited information (Heyman and Sommervoll, 2019). Therefore, considering the impact of hyper-locality variables on housing prices can expand current knowledge, and perhaps give more accurate property value estimations.

Present day's changing and complex environment characterized by globalization, rapid technological development and well-developed Business Intelligence systems, has lowered extraction costs and ensured greater data access. As a result of this, it has become easier to implement more detailed location data in property value estimation models. Nevertheless, it is observed that few researchers have studied the effect of distance-measured hyper-locality features on the property value. In addition, it does not appear that previous research within property value estimation largely has focused on the socio-economic aspect. A literature review revealed that only Heyman and Sommervoll (2019) have conducted a detailed study of hyper-locality variables, where in addition a socio-economic aspect was included. An inclusion of a socio-economic aspect can be beneficial for several reasons. For instance, districts can be a measure for both centrality and status, where the latter can be approxied by socio-economic variables. Moreover, significant relationships between factors are not necessarily unambiguous, but can be complex and need further explanation. On that basis, the aim of this thesis is twofold. Firstly, a wider set of hyper-locality variables is included to outperform earlier real estate price estimation models. In order to illuminate this, the real estate market in Oslo is analyzed. Secondly, the study seeks to explore the underlying effects of physical and socio-economic location factors on the sales price of apartments in Oslo.

Since 2005, fewer homes have been built in the Norwegian capital, Oslo, relative to the increase in the number of households (Mæhlum, Pettersen and Xu, 2018, as cited in Lindquist et al., 2021). This, in addition to the low interest rate level, has contributed to increased house price growth. Several empirical studies have found that interest rate changes have a stronger impact on the residential market in Oslo than on a national basis (Hov, 2021; Midtgaard, 2019, as cited in Lindquist et al., 2021). Despite the fact that the total price level in Oslo has increased, the relative price growth between the districts in Oslo varies, as central districts have had an especially strong price growth. On the other hand, the outer southern and eastern districts of the city have had a clearly weaker price development than the outer western districts (Öye, 2019). This indicates that location, related to both hyper-locality and socio-economic factors, potentially plays a substantial role when people purchase a home in Oslo. This makes the Oslo real estate market an interesting case for the thesis.

The dataset used in the thesis consists of 62,134 apartment transactions in Oslo for the five-year period 2017 to 2021. To answer the problem statement, three different hedonic models are estimated: a baseline model, a hyper-locality model, and a socio-economic model. In the analyses, the hedonic price method is supplemented by Structural Equation Modeling (SEM) in order to examine more complex relationships between hyper-locality variables. The study attempts to combine hedonic and SEM methodology to achieve a deeper understanding of the underlying effects of different location variables. According to what is known, this study is the first to conduct a SEM model built on detailed, high-quality hyper-location data. Generally, it appears that earlier studies tend to include fewer hyper-locality variables when examining dwelling prices compared to this study. An inclusion of previously omitted hyper-locality variables, and a self-compiled socio-economic variable, can potentially improve the model estimations and interpretability of various location factors. As a result of this, the thesis provides a useful and unique contribution within the field of property value estimation.

The thesis finds that there are differences within the Oslo districts in apartment prices, depending on the proximity to hyper-locality features. The SEM analysis disclosed three location factors, and concluded that distance to leisure activities is the most important location factor. In other words, the hyper-locality features distance to library, museums, theaters, and gyms, collectively affect the apartment sales price per square meter the most. The results from the study will be of importance for several decision makers, such as politicians, urban planners, and private individuals, in regards to urban issues, planning and housing. The findings can highlight specific location related attributes valued by apartment buyers in Oslo, and not just the district location in itself.

The structure of the master thesis is as follows: Chapter 2 describes the background and presents the relevant literature for the study. Thereafter, the real estate market on a national basis and in Oslo is accounted for, with an emphasis on critical concepts. In Chapter 3, the dataset and data cleaning process is described, in addition to descriptive statistics of relevant variables. Chapter 4 introduces the hedonic price models and the theoretical SEM research model. Next, in Chapter 5, the estimated models and results from the analyses are presented, in addition to evaluation of model performances. In Chapter 6, findings are compared with previous literature and implications are discussed. Finally, Chapter 7 presents the conclusion and suggestions for further research within this field.

Chapter 2

Literature Review and Background

Geospatial data available to researchers has increased tremendously over the last decades, opening up opportunities to define residential location in multiple ways (Schirmer et al., 2014). Several methods are utilized in previous literature when investigating the effect of different location factors on dwelling prices. As mentioned in Chapter 1, relatively few earlier studies have performed a detailed evaluation of hyper-locality features, and those who have tend to include few of them in the analyses. In the following chapter, a literature review of relevant research is presented, followed by background information about the real estate market in Norway and Oslo.

2.1 Location and Hyper-locality

The use of hyper-locality data for estimating property value is not a new phenomenon. McLeod (1984) investigated the demand for proximity to four different local amenities in Perth, Australia, by the application of hedonic price theory to a sample of house transactions. He examined the importance of local amenities on housing prices, and found that proximity to the local park and the river had a positive effect on the sales price of the property. In contrast, he found that proximity to the highway reduced the sales price of the dwelling. The effect of the distance to the central business district on the sales price was also included, but the estimated coefficient was not statistically significant.

Dziauddin and Idris (2017) studied the effect of five hyper-location attributes on residential property values of Kuala Lumpur, Malaysia. The included hyper-locality features were distance to shopping malls, forests, rail stations, and primary and secondary schools. Through the use of a Geographically Weighted Regression (GWR) specification of a hedonic model, they found that proximity to shopping malls, forests and rail stations were likely to increase the property sales price. On the other hand, their study showed that for every meter the dwelling was located away from a primary and secondary school, the price increased.

Heyman and Sommervoll (2019) examined hyper-locality in order to understand house prices in the metropolitan area of Oslo, through a hedonic model for apartments transactions. Their analyses included walking distances to key locations, proximity measures to commercial services and cultural amenities. They found that proximity to features such as metro stations, highways, large parks, and forests, was expected to reduce the sales price of a dwelling, while proximity to the Oslo fjord, smaller parks, and supermarkets, were expected to result in a higher sales price. The study found that hyper-locality features could reveal more to the value of a location than the postcode itself, as they shed light on proximity valuations of consumers. Furthermore, a socio-economic variable based on employment level, income level, and education level, was included to capture status effect of living in an area generally considered attractive. However, a decrease in the model's explanatory power was seen after the inclusion of this variable.

Analyses regarding hyper-locality features is also prevalent in real estate development literature, in order to help real estate developers better meet the purchase preferences of buyers. Cheng (2020) examined what factors real estate developers should keep in mind with regard to area selection. The article proposed a novel site selection method of real estate developers. The study was based on a survey performed to measure latent factors such as cost-effectiveness, reputation, cultural, and environmental factors. A SEM model was then constructed with the factors influencing consumers' purchase choice to study how home buyers valuate a property. Liu and Wu (2009) built a SEM model in order to set up a more comprehensive indicator system to study the impact of various characteristics on the property value. Based on traditional hedonic price theory and qualitative analysis they disclosed four latent variables: location, district, house structure, and neighbourhood environment. The results showed that location ranked as the most important influencing factor.

2.2 The Real Estate Market in Norway and Oslo

The models in the analyses are estimated based on transaction data from Oslo, Norway. Hence, it is relevant to present some contextual information on the Norwegian real estate market. Norway is characterized by a high degree of home ownership, where as many as 81.8% of the population own their own home (SSB, 2022). The Norwegian home ownership rate is relatively high compared to the European Union average (69.7%), and is significantly higher than its neighbouring countries Sweden and Denmark, where the rates are 64.5% and 59.3%, respectively (Eurostat, 2021).

Usually, the seller will employ a real estate agent to help with the transaction. The real estate agent uses her expertise to assess the value of the dwelling and is supposed to set a reasonable asking price. The asking price should express the value of the dwelling, market conditions, and the price the seller is willing to sell for. Setting a strategic misleading price is not allowed, and will be a violation of several laws regulating this aspect (Forbrukertilsynet, 2022). Most dwelling announcements are published on Finn.no, which is the dominating platform for publication of housing ads in Norway (Huseierne, 2022).

Oslo, the capital of Norway, is the most populous city in Norway with approximately 700,000 inhabitants (Oslo Kommune, 2022). A relatively large proportion, around 34%, of Oslo's inhabitants are immigrants or Norwegian-born with immigrant parents (Oslo Kommune, 2021). The real estate market in Oslo is characterized by high square meter prices and a stronger price growth than the rest of the country (Krogsveen, 2022). The population density is highest in Grünerløkka and Sagene, followed by St. Hanshaugen, Frogner, and Gamle Oslo (Oust, 2012), which together make up the inner city. The outer districts consist of Ullern, Vestre Aker, Nordre Aker, Bjerke, Grorud, Stovner, Alna, Østensjø, Nordstrand, and Søndre Nordstrand. In its entirety, Oslo consists of 15 administrative city districts as seen in Figure 2.1. Marka and Sentrum are two additional districts, but they do not have their own administrative committee.

Roughly, the real estate market of Oslo can be divided in an east-west division. The districts in the east of Oslo, from now on referred to as Oslo East, are Gamle Oslo, Grünerløkka, Sagene, Bjerke, Grorud, Stovner, Alna, Østensjø, Nordstrand, and Søndre Nordstrand. The remaining five districts, St. Hanshaugen, Frogner, Ullern Vestre Aker, and Nordre Aker, are classified as Oslo West. Historically, square meter prices for apartments on average have been higher in Oslo West than in Oslo East (Bydelsfakta, 2022a). The east-west distinction is due to historical factors. Already in the late 19th century, the division between the rich western part of the city and the poorer eastern part took place, as factories were built alongside Akerselva on the east side of the river (Oust, 2012). Thus, the industrial workers settled in the eastern part of Oslo. The historic east-west distinction has persisted over time, and is still present through differences regarding socio-economic factors. Today, Oslo West is characterized by a higher income level, higher level of education and lower unemployment rate. Additionally, Oslo East has a significantly larger immigrant population compared to Oslo West (Bydelsfakta, 2022d). A more detailed description of various socio-economic rates for each district can be found in Appendix A.1.

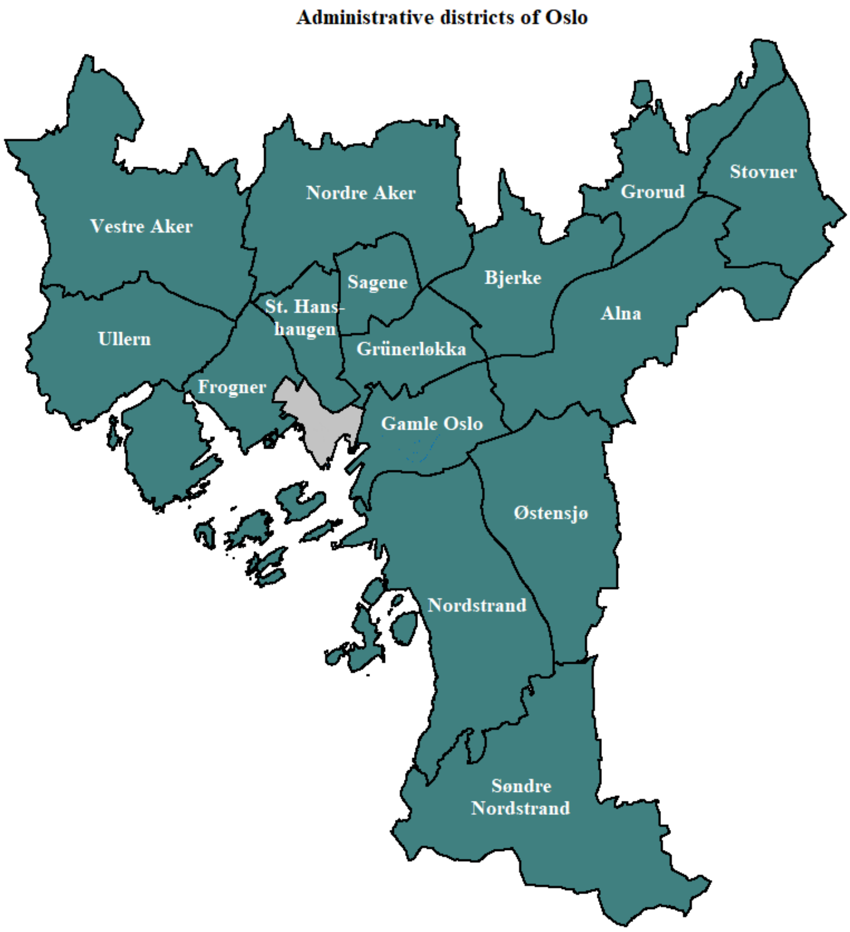


Figure 2.1: The 15 administrative districts of Oslo. District Sentrum is denoted by grey color. District Marka, that surrounds the northern and eastern districts of Oslo, is not represented in our dataset, and is thus not shown on the map.

Chapter 3

Data Analysis

The real estate transaction data used in the study was provided by VIRDI AS, a Norwegian proptech company. VIRDI has access to data collected mainly from Ambita, which registers all property and land transactions in Norway and sits on an unique and complex data base. The dataset provided contained VIRDI's reworked cadastre, all transactions from the official Norwegian transaction registry since 1991, and features related to the properties. Next, the dataset from VIRDI is supplemented by socio-economic statistics from Bydelsfakta. Bydelsfakta is a platform published by the City Council's department for finance in Oslo municipality, which is developed by Oslo Origo (Bydelsfakta, 2022b). On the platform, central statistics on population and living conditions in Oslo are published, both on an aggregated level and on a district level. Prior to utilizing the dataset for usage and interpretation purposes, modifications were needed. This chapter will present the process of data merging, data cleaning, and further exploration of the data through a descriptive examination of the variables included in the final dataset.

3.1 Data Pre-Processing

3.1.1 Data Selection

The analysis in the study is limited to transactions in Oslo in the time period 2017 to 2021. The original dataset consisted of 90,655 transactions. It is characterized by a cross-sectional structure, which implies that observations are mutually independent. In the dataset, the data from recent years is observed to be most complete considering feature values, and is in addition more comparable to the macroeconomic conditions today. Moreover, the study also utilities the public data source Bydelsfakta to construct a socio-economic status index for the districts, see Subsection 3.2.7. On that basis, it is also convenient to ensure that the chosen time period corresponds with the years available from the public data.

Furthermore, the study focuses exclusively on apartments in Oslo. For the chosen time period, the process of organizing data based on housing types showed that 89.0% of all the transactions in Oslo were apartments, see Table 3.1. The majority of transactions being apartments is as expected, as apartments account for about 76% of all housing types in Oslo (Bydelsfakta, 2022c). Apartments in general have a higher price per square meter compared to other types of housing, although the total price often is higher for detached houses (Oslo Kommune, 2018). Comparing prices per square meter in different areas of Oslo can therefore be deceptive if dwelling type is not taken into account. Different areas of the city have a different proportion of the several housing types. Thus, looking at one specific housing type will be an advantage as it provides more appropriate terms for valuation modeling and comparison (Oslo Kommune, 2018). In the case of Oslo, this housing type is preferably apartments, as it is the most common housing type in the capital. The remaining housing types from the dataset are not used for further analysis.

Table 3.1: Housing type distribution in the dataset for Oslo from 2017 to 2021.

Housing type	**Frequency**	**Percent**
Apartment	80,675	89.0%
Serial house	3,749	4.1%
House	3,456	3.8%
Semi-detatched house	2,567	2.8%
Other	205	0.2%
Holiday home	3	0.0%
Total	**90,655**	**100.0%**

3.1.2 Data Cleaning

The dataset contained missing values, typing errors, and unrealistic observations. Therefore, the process of data cleaning was focused around handling these, which made it necessary to filter out a number of observations. Removal of missing values was done carefully as this could have severe implications for further analysis. It turned out that if a transaction had one missing on a hyper-locality variable, it also had a missing on the remaining hyper-locality variables. Because the hyper-locality variables are the main focus of this master thesis, the transactions in question were deleted.

In order to exclude unrealistic observations, the cut-offs of Pollestad and Helgaker (2021) were utilized. Firstly, observations with sales price more than twice the list price and less than half of the list price were removed. Secondly, observations with apartment size equal or less than nine square meters and equal or more than 300 square meters were also removed from the dataset. Furthermore, only apartments built after year 1600 were kept. Next, transactions with more than 10 bedrooms or located on a floor under one or higher than 20 were excluded from the dataset. Pollestad and Helgaker (2021) argue that these cut-offs are appropriate because observations outside this range are generally not seen as realistic based on common sense. Finally, descriptive statistics were calculated for key variables from a dataset including missing observations and one without, to examine the pattern of the remaining missing values. When comparing the two datasets, systematic missing data was not observed. Thus, it could be argued that entries contained missing values mainly because of failure in the recording of transactions. Consequently, it was safe to remove the transactions containing missing values. The data cleaning process resulted in 62,134 observations, and an overview of the process can be seen in Table 3.2. Even after a thorough data cleaning process, the final dataset is still deemed sufficiently large for estimating models.

Table 3.2: A detailed description of the data cleaning process of the original dataset. The process resulted in a dataset consisting of 62,134 apartment transactions.

Data cleaning process	**Observations left**
Original data (Oslo, 2017 to 2021)	90,655
Apartments only	80,675
Remove missing from hyper-locality variables	76,682
Keep apartments with sales price higher than 0.5*list price and lower than 2*list price	73,423
Keep apartments with size larger than 9 square meters and less than 300 square meters	73,414
Keep apartments with build year newer than 1600	73,406
Keep apartments with less than 10 bedrooms	71,226
Keep apartments with floor larger than 0 and less than 21	70,639
Listwise missing deleting procedure	62,134

3.2 Descriptive Statistics

3.2.1 Sales Price Variable

One approach in earlier research is to estimate the value of a given dwelling by the sum of its sales price and the sum of its common debt, divided by the house area (Oust et al., 2020). By dividing the total sales price with a size variable, the dependent variable is adjusted for variations in apartment sizes, and gives a more comparable measure for studies of other Norwegian cities. In statistics related to real estate prices, the living area is commonly used as a size variable, and is thus the preferred size variable here. Living area makes up the total interior livable area excluded storage space, and usually consists of living room, kitchen, bedrooms, bathrooms, hallway, and stairs (Iversen, 2020). Consequently, the dependent variable in this study is the natural logarithm of the sales price per square meter represented by the total transaction price, including sales price and common debt, divided by the size based on living area, as presented in Equation 3.1. The logarithmic functional form of the response variable offers a simple and appealing interpretation, as the coefficient can be interpreted as the percentage change in the value given a unit change in the independent variable. The logarithmic transformation also mitigates the effect of any remaining outliers in the dataset.

$$ln(Price\,per\,square\,meter_i) = \frac{Sales\,price_i + Common\,debt_i}{Living\,area_i} \tag{3.1}$$

Descriptive statistics for the variable in Table 3.3 underline the differences in apartment square meter prices in the dataset. Minimum apartment sales price per square meter is NOK 19,938, while the most expensive apartment had a cost of NOK 237,000 per square meter. Included apartments give a median value of NOK 74,393. The histogram in Figure 3.1 shows the distribution of sales price per square meter for apartments in the data material. The graphical visualization indicates that there is a large spread in sales prices per square meter in Oslo.

Table 3.3: Descriptive statistics for sales price per square meter in NOK, presented with mean, standard deviation, median, 1st and 3rd quantile, and minimum and maximum values. N represents the number of observations.

Mean	**Standard deviation**	**Median**	**1st quantile**	**3rd quantile**	**Min**	**Max**	**N**
75,102	20,234	74,393	61,167	87,644	19,938	237,000	62,134

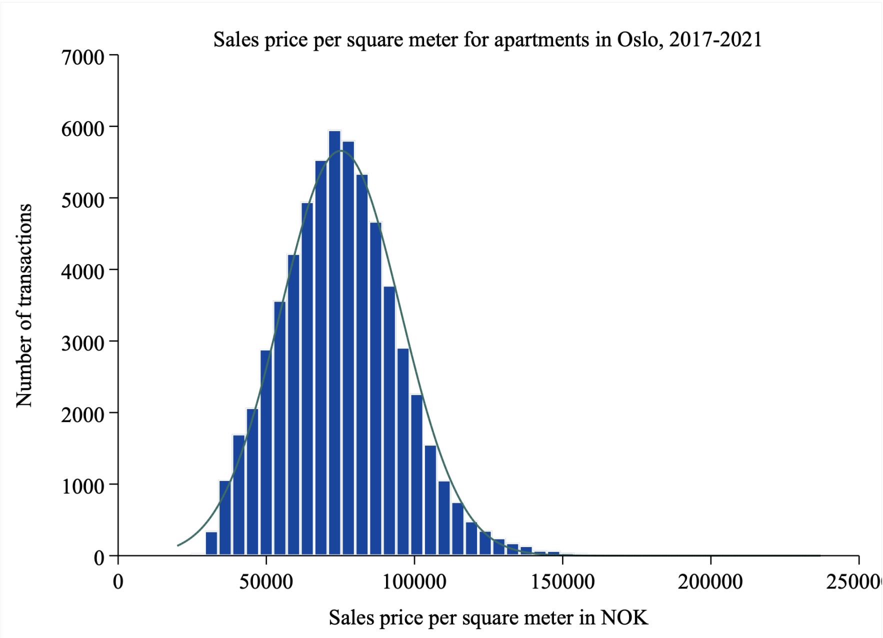


Figure 3.1: Sales price per square meter for apartments in Oslo in NOK for the time period 2017 to 2021. The histogram shows a right skewed distribution of sales price per square meter for an apartment in Oslo, with an average sales price of NOK 75,102 per square meter.

Table 3.4: Skewness and kurtosis tests for normality in the dependent variable, sales price per square meter. The left column shows a test for skewness, the middle column a test for kurtosis, and the column to the right show an overall test for skewness and kurtosis.

Skewness	**Kurtosis**	**Joint test**
Pr(skewness) = 0.000	Pr(kurtosis) = 0.000	chi2(2) = 2,962.53 Prob>chi2 = 0.000

Tests for skewness and kurtosis in Table 3.4 show significant skewness and significant kurtosis, with an associated p-value of 0.000 for both tests, in addition to a p-value of 0.000 for the joint test. This confirms significant non-normality for the sales price per square meter variable.

3.2.2 Size and Build Year Variables

Further, the independent variables relevant for the apartment price estimation are presented. The first included regressors address the size and build year of the apartments. The original dataset consisted of three different size variables. In addition to living area, there were variables related to gross floor area and usable square meters. The former was chosen as the basis for the size variable, due to the argumentation in Subsection 3.2.1, and that the living area is deemed more precise than the other size variables.

In the modeling, the size variable is expressed as the natural logarithm. The use of a continuous logarithmic size variable is reasonable due to the likelihood of diminishing returns as, *ceteris paribus*, the value of the variable increases (Bourassa et al., 2003). It is an informative and interpretable approach, which also mitigates the effect of any remaining outliers in the dataset, as mentioned in Subsection 3.2.1. In addition, LISREL models are better at handling continuous variables. An alternative, typically used approach is to use the size variable categorically in the hedonic analyses, which is typically used in previous research (Nesset and Oust, 2020; Oust et al., 2020; Pollestad and Helgaker, 2021). However, the size variable is here used logarithmically in both hedonic and SEM modeling, in order to be consistent. The build year variable will be used as a dummy for apartments built before and after year 2000. The division is included to distinguish between old and new apartments, as it can be argued that apartments built after year 2000 are relatively new.

Table 3.5: Descriptive statistics for the size and build year variables, presented with mean, standard deviation, median, 1st and 3rd quantile, and minimum and maximum values. N represents the number of observations.

	Mean	**Standard deviation**	**Median**	**1st quantile**	**3rd quantile**	**Min**	**Max**	**N**
Size	65.73	24.14	64	50	77	11	299	62,134
Build year	1959	38	1960	1936	1988	1697	2021	62,134

Table 3.5 shows the descriptive statistics for the size and build year variables. The median apartment in the dataset is 64 square meters and built in 1960. The size mean is 65.73, which is larger than the median value, and indicates that the size variable distribution is right skewed. Size varies from 11 to 299 square meters for the apartments in the data material. An overview of how apartment sizes vary in each district can be found in Appendix **??**. Further, the oldest apartment was built in 1697, and the newest apartments were built in 2021.

3.2.3 Facility Variables

The next category of variables concerns binary variables related to facilities of the dwelling. The dataset initially contained 18 facility variables. When browsing the housing platform at Finn.no, there are 10 facilities that are highlighted and the buyer can check off to refine the search. These are balcony, garage, elevator, no opposite neighbours, electric car charger, fireplace, beach line, hiking nearby, view, and janitor (Finn.no, 2022). To limit the number of facility variables, the aforementioned features were used as a starting point. However, after evaluating the importance and quality of each variable in the context of explaining sales prices, no opposite neighbours, electric car charger, and janitor, were excluded from the dataset. The data material did not contain information about whether the dwelling had a garage or not, and this facility was therefore unavailable. Furthermore, the variable quiet was included on the basis of its relevance in connection to location. The five facility variables are presented in Table 3.6.

Table 3.6: The percentage distribution of binary facility variables for apartments in Oslo in the dataset.

	Quiet	**View**	**Balcony**	**Fireplace**	**Elevator**
Yes	80.3%	45.4%	87.4%	26.5%	36.7%
No	19.7%	54.7%	12.6%	73.5%	63.3%

3.2.4 Hyper-Locality Variables

The dataset consists of 13 hyper-locality variables. The thesis proceeds with nine of the variables for further analysis, chosen on the basis of the variables' quality. All of the variables are measured in air distance to specific services in kilometers, and are presented in Table 3.7. In forthcoming analyses in Chapter 5, the distance variables are logarithmically transformed, due to the expectation that people appreciate proximity up to a certain distance and then the utility of proximity decreases. This is consistent with the theory of diminishing marginal returns, as discussed in Subsection 3.2.2. Additionally, this transformation is often used in econometrics, because the regression coefficients are given a useful interpretation as elasticities (Studenmund, 2017).

Table 3.7: Descriptive statistics for hyper-locality variables, presented with mean, standard deviation, median, 1st and 3rd quantile, and minimum and maximum air distance from an apartment to the service. N represents the number of observations. All distances are measured in kilometers.

Distance	**Mean**	**Standard deviation**	**Median**	**1st Qqantile**	**3rd quantile**	**Min**	**Max**	**N**
Primary school	0.708	0.400	0.630	0.414	0.921	0.021	2.522	62,134
Secondary school	1.032	0.718	0.841	0.536	1.318	0.033	4.813	62,134
Library	5.565	4.316	4.161	1.980	8.313	0.063	17.094	62,134
Museum	2.698	2.528	1.587	0.770	4.183	0.011	10.611	62,134
Theater	2.266	2.025	1.518	0.651	3.344	0.022	8.936	62,134
Gym	4.080	3.671	2.282	1.409	6.561	0.023	14.778	62,134
Retail	0.441	0.270	0.392	0.239	0.592	0.005	2.330	62,134
Subway and train	0.499	0.415	0.363	0.215	0.654	0.007	3.394	62,134
Industry	0.397	0.301	0.314	0.177	0.542	0.009	1.847	62,134

As seen from the overview in Table 3.7, none of the apartments in the final dataset are located more than approximately 1.8 kilometers away in air distance from the closest industrial building. In contrast to this, the apartment that is located the furthest away from a library, is located about 17.1 kilometers away from the closest one. Further, the mean air distance from the apartments in our dataset to the closest library is around 5.6 kilometers. This is the longest average distance to a service in our dataset, followed by average distance to the closest gym, which is nearly 4.1 kilometers. The median apartment in the dataset has the shortest distance to three following services: retail store, industrial building, and subway and train.

3.2.5 District Variable

Housing prices tend to vary geographically (De Bruyne and Van Hove, 2013), and thus the geographical location of the apartments in the dataset is a central factor in house price determination. Districts therefore ought to be included as explanatory variables in the models. The original dataset contained information on each apartments' full address with postal code and geographical coordinates, in addition to district information. Hence, each observation's geographical location could have been included in the dataset in several ways. Whether geographical divisions should be based on postal codes or districts, comes down to a trade-off between low and high spatial aggregation (Sommervoll and Sommervoll, 2018, as cited in Pollestad and Helgaker, 2021). However, this study examines price differences that stem from geographical location through looking at district divisions. This division is considered an appropriate spatial grouping able to capture spatial effects, because districts were found to have similar intra-regional location premiums (Pollestad and Helgaker, 2021). Furthermore, choosing districts spatial groupings is convenient, because the socio-economic data, see Subsection 3.2.7, were available at district level.

The dataset in the study originally consisted of 16 districts that differ in size, inhabitants, and average sales price per square meter. Observations belonging to the city center district, Sentrum, were reassigned with St. Hanshaugen, due to its limited number of transactions and the fact that Sentrum does not have its own administration, as mentioned in Section 2.2. This is a measure taken to reduce the possible impact of potential outliers. As the municipality of Oslo consists of 15 administrative districts (Oslo Byleksikon, 2022), the rest of the study consequently proceeds with these 15 districts. Descriptive statistics of the sales price per square meter for each district are shown in Table 3.8.

Table 3.8: Descriptive statistics of the sales price per square meter in NOK for each district, presented with mean, standard deviation, median, 1st and 3rd quantile, and minimum and maximum values. N is the number of observations within each of the 15 administrative districts in the dataset.

District	**Mean**	**Standard deviation**	**Median**	**1st quantile**	**3rd quantile**	**Minimum**	**Maximim**	**N**
Gamle Oslo	79,685	16,350	76,816	68,421	87,823	37,215	208,738	7,785
Grünerløkka	81,570	13,874	79,717	71,875	89,450	20,926	166,378	9,197
Sagene	85,707	16,250	83,281	74,083	94,271	33,326	172,750	7,546
St. Hanshaugen	87,793	16,082	85,770	76,392	97,255	42,089	175,818	4,391
Frogner	93,301	17,978	90,997	81,049	103,218	45,981	237,000	5,956
Ullern	81,972	16,719	80,412	69,754	92,430	37,942	189,815	2,241
Vestre Aker	71,630	14,258	69,404	61,547	79,197	30,918	148,402	2,111
Nordre Aker	81,540	15,739	80,245	70,000	90,196	45,692	174,786	2,727
Bjerke	64,016	14,255	62,676	53,438	73,152	34,712	154,444	3,026
Grorud	51,072	13,042	48,937	41,491	56,891	22,894	104,931	2,319
Stovner	45,771	11,918	42,969	37,927	50,639	22,105	149,383	1,449
Alna	53,623	12,257	52,046	44,641	60,632	22,048	119,271	4,602
Østensjø	60,037	10,409	58,493	53,000	65,382	22,904	115,767	4,702
Nordstrand	65,121	12,920	62,438	56,593	70,413	31,603	159,864	2,957
Søndre Nordstrand	44,935	9,785	43,541	38,063	50,091	19,938	90,334	1,125
Total	**75,102**	**20,234**	**74,393**	**61,167**	**87,644**	**19,938**	**237,000**	**62,134**

There are noticeable differences in average sales price per square meter, number of observations, and the minimum and maximum price per square meter, across districts in Oslo. This emphasizes the importance of examining the relationship between the square meter price and district. As seen in Table 3.8, the average sales price per square meter in Frogner was more than twice as high as in Søndre Nordstrand, respectively NOK 93,301 and NOK 44,935. The districts with mean sales price per square meter above the average in Oslo are Frogner, St. Hanshaugen, Sagene, Ullern, Grünerløkka, Nordre Aker, and Gamle Oslo. At the opposite end, Søndre Nordstrand, Stovner, Grorud, Alna, Østensjø, Bjerke, Nordstrand, and Vestre Aker, are found with average square meter prices below the average in Oslo. This is illustrated in Figure 3.2.

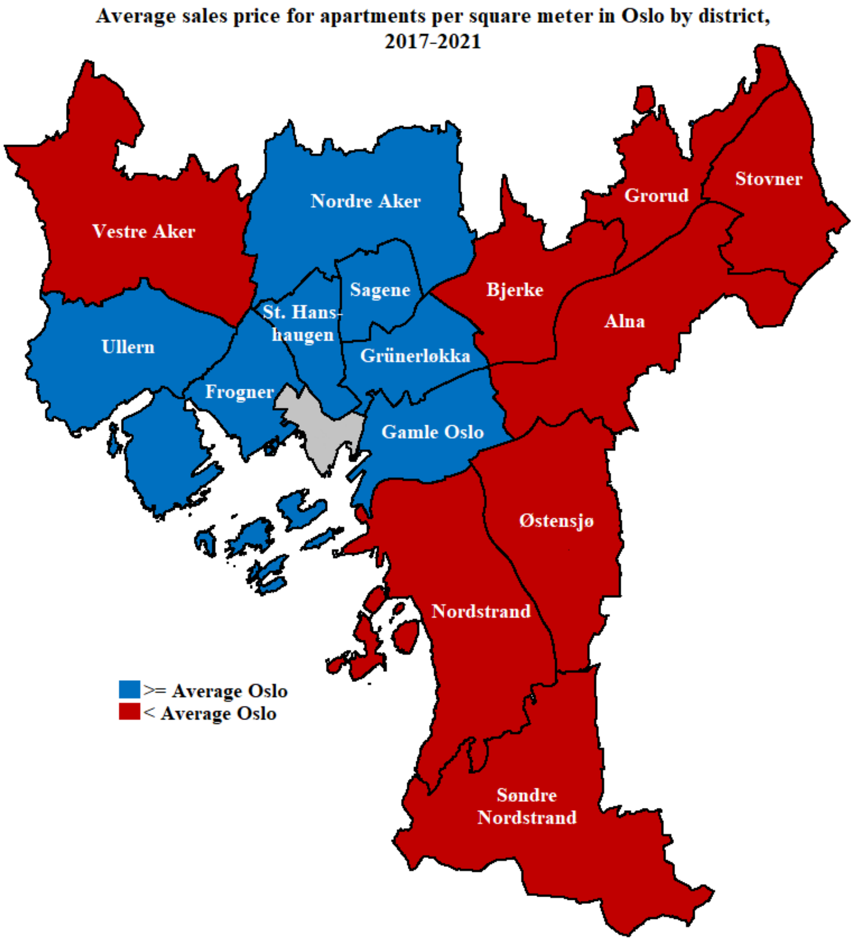


Figure 3.2: Average sales price for apartments per square meter in Oslo by district, for the time period 2017 to 2021. Districts with blue color have an average sales price per square meter above or equal to the average of NOK 75,102 in Oslo, while districts with red color have an average sales price per square meter below the average. District Sentrum is denoted by grey colour. District Marka is not represented in our dataset and is thus not shown on the map.

3.2.6 City Area Variable

An alternative approach to take into account the geographical location of a dwelling, is by grouping the districts into inner and outer city area, based on location. Although this division represents a lower spatial aggregation, it can provide useful information that is not captured by the district and hyper-locality variables. In this way, travel costs can be more thoroughly examined. Travel costs are associated with travel expenses, travel time, and other costs related to travelling.

The inner city consists of the following five districts surrounding the city centre: Gamle Oslo, Grünerløkka, Sagene, St. Hanshaugen, and Frogner. The remaining districts, which are Ullern, Vestre Aker, Nordre Aker, Bjerke, Grorud, Stovner, Alna, Østensjø, Nordstrand, and Søndre Nordstrand, make up the outer city of Oslo. Inner city accounts for 56.1% of the apartments in the dataset, while the remaining 43.9% of the apartments are located in the outer city. A general map of the city districts was presented in Figure 2.1. The city area variable is used as a dummy for inner and outer city in the model estimations.

3.2.7 Socio-Economic Variable

Variations in housing prices, beyond location and real estate characteristics, can be attributed to differences in socio-economic variables (De Bruyne and Van Hove, 2013). In order to study this in detail, a socio-economic index is compiled based on components that are indicators of living conditions and demographics. The statistics were available at district level, and the index therefore serves as a proxy for measuring the socio-economic status of the districts of Oslo. Four components are retrieved from Bydelsfakta and included in the index: unemployment level (Em), level of low education (Ed), level of low-income households (I), and the share of immigrants with long residence (Im). The three former components were chosen based on the analysis of Heyman and Sommervoll (2019), while the immigration component was added because this demographic level varies largely across the districts of Oslo. By utilizing data from the five most recent available years, an average value was calculated. Then the components were standardized into a continuous socio-economic index ranging from 0 to 1, where each district has an individual socio-economic score S, as seen in Equation 3.2. See Appendix A.1 for a more detailed display of the four standardized socio-economic components.

$$S_i = 1 - \left(\frac{\left(\frac{Em_i - min(Em)}{max(Em) - min(Em)} + \frac{Ed_i - min(Ed)}{max(Ed) - min(Ed)} + \frac{I_i - min(I)}{max(I) - min(I)} + \frac{Im_i - min(Im)}{max(Im) - min(Im)}\right)}{4}\right) \quad (3.2)$$

Low level of education is defined as the share of the citizens between the age 30 to 59 with low education, meaning people with only primary school, no, or unspecified education. Low income level is based on statistics on households with children under the age of 18 with low incomes, according to the EU scale. The unemployment level includes residents in Oslo aged 30 to 59, who are not registered as employed per the fourth quarter in the current year. To be employed is defined either as an employee or self-employed, performing income-generating work of at least one hour's duration during the reference week. The share of immigrants with long residence defines an immigrant as a person born abroad by two foreign-born parents, and four foreign-born grandparents. Residence time is calculated according the year of arrival in Norway. The aforementioned definitions are provided by Bydelsfakta (2022b), and more information about the components can be found on their online platform.

An overview of the districts and their respective socio-economic score is found in Appendix A.1. Figure 3.3 provides an illustrative ranking of the districts based on their socio-economic score. Stovner, Søndre Nordstrand, and Alna, are ranked as the districts with lowest socio-economic score among the Oslo districts, while Nordre Aker, Vestre Aker, and Ullern, have the highest scores.

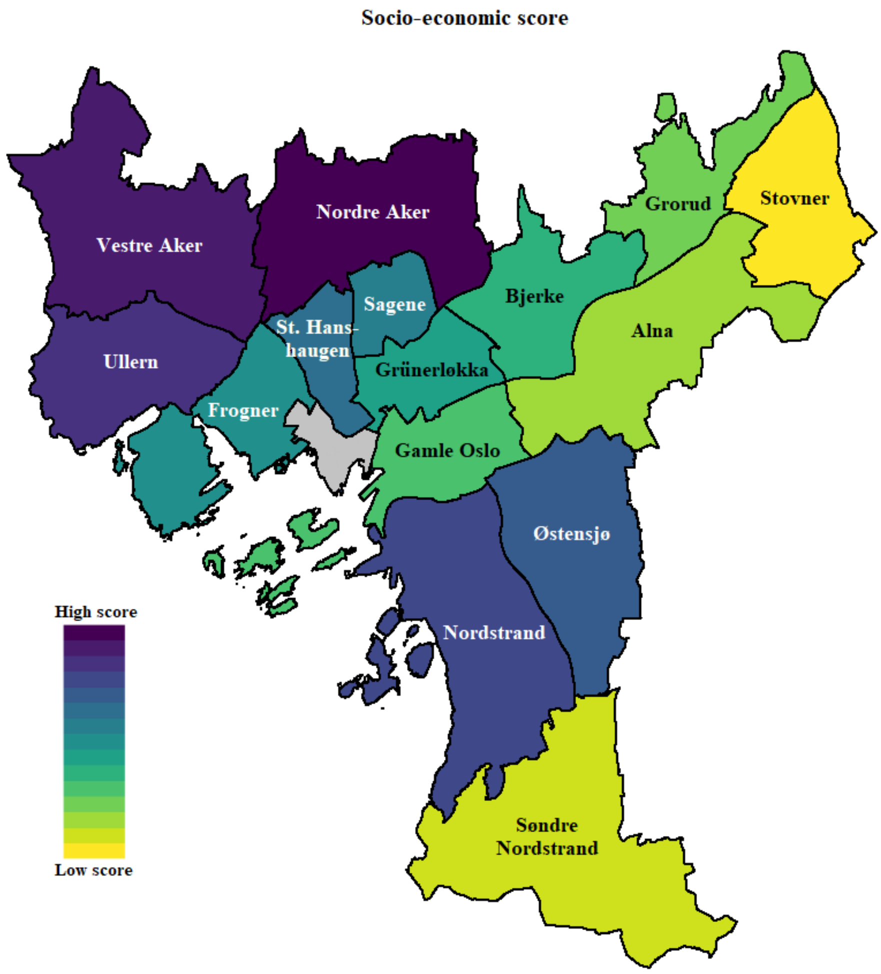


Figure 3.3: Districts ranked by socio-economic score, given by Equation 3.2. District Sentrum is denoted by grey color. District Marka is not represented in our dataset and is thus not shown on the map.

Chapter 4

Methodology

The literature review indicated that hedonic pricing models are the most commonly used within property value estimation that focuses on location. Machine learning methods are also becoming increasingly popular for this purpose, as they mitigate the strict assumptions of feature independence and linearity in the hedonic model (Venerandi et al., 2019). Structural equation modeling is another realistic, yet less widespread, analytical approach for estimating property prices based on locality variables. This study attempts to combine hedonic and SEM methodology to achieve a deeper understanding of the underlying effects of different location variables, such as various hyper locality variables. In the following chapter, the methodology of the two approaches will be presented. In addition, the research models to be estimated in Chapter 5 are introduced, as well as measures for evaluating model fit. Finally, relevant hypotheses associated with the SEM model are accounted for, and reliability and validity of the SEM model is presented.

4.1 Hedonic Regression

4.1.1 Hedonic Method

Hedonic regression models were first introduced by Rosen (1974) and are used to assess people's willingness to pay for various characteristics of complex, heterogeneous goods. Today, hedonic models are widely used in valuation of real estate to explain price differences. Dwellings are considered complex goods because they have distinctive features that make them unique. The hedonic approach takes this into account by determining property value through the sum of the market value of its various attributes. Hedonic models are considered intuitive and simple to use, and real estate is just one of several examples of heterogeneous goods that this method can be used to valuate.

The dataset in this thesis consists of a variety of hedonic attributes. Different attributes can be valued based on the benefit they provide (Rosen, 1974). Individuals can value dwelling attributes differently, and a variation in their willingness to pay for a property can therefore be observed.

In order to find an objective price for each property, regression can be used (Rosen, 1974). The approach is based on the attributes of the dwelling, which can affect the sales price and provide a result that is independent from the asking price set by the real estate agent. In this way errors related to mispricing are handled by hedonic regressions.

Several hedonic regression formulations can be employed when estimating the value of a dwelling. Palmquist (1984) emphasizes linear, semi-logarithmic, log-linear, and inverse semi-logarithmic as the most frequently used functional forms in valuation models. Therefore, this thesis specifies hedonic models with a log-linear functional form, based on Ordinary Least Squares (OLS) method, which is consistent with the argumentation of Malpezzi et al. (2003). As discussed in Subsection 3.2.1, the log-linear model offers a simple and appealing interpretation. Additionally, it mitigates heteroscedasticity, which is a common statistical problem. Thus, Equation 4.1 is applied:

$$ln(P_{it}) = \gamma_0 + \delta_t + \sum_k \alpha_k c_{kit} + e_{it} \tag{4.1}$$

where $ln(P_{it})$ is the natural logarithm of the sales price per square meter for dwelling i in time period t $(t = 1,...,T)$. γ_0 is the base year intercept of the model. Further, δ_t is the time dummy coefficient for period t, in relation to the base year, which here represents year 2017. c_{kit} is a set of explanatory variables for the presence of different attributes k, apartment i, in time period t, respectively. α represents the hedonic regression coefficients for each explanatory variable, while e_{it} is the error term of the model.

4.1.2 Hedonic Models

The further analysis will be based on three different hedonic models. These are the baseline model, the hyper-locality model, and the socio-economic model, from now on referred to as Model 1, Model 2, and Model 3, respectively. In Model 1, size, sales year, and build year of apartment i, are included as independent variables. Next, the five binary dwelling facility variables are added as predictors, in addition to the 15 administrative city districts of Oslo to take into account the geographical location of the transactions in the dataset. Furthermore, Model 2 is constructed, which is an extension of Model 1. Here, the nine distance measured hyper-locality features are included in order to more thoroughly examine the effect of locality on the property value. To analyze the underlying effects of locality variables on the sales price of apartments in Oslo, Model 3 is assembled. Model 3 corresponds to Model 2, but here districts are substituted with the socio-economic index and city area variable. This is done in order to capture the socio-economic differences, with respect to the east-west division, as discussed in Section 2.2, and the effect of travel costs. A detailed description of the variables included in the three models was presented in Chapter 3. A summary of the included variables in each of the three hedonic models is found in Table 4.1.

Table 4.1: A summary of the three hedonic models with associated independent variables included in each model. The * mark indicates that the variable is included in the respective model.

	Model 1	**Model 2**	**Model 3**
	Baseline model	Hyper-locality model	Socio-economic model
Districts	*	*	
Hyper-locality variables		*	*
Socio-economic index			*
City area			*
Size	*	*	*
Sales year	*	*	*
Build year	*	*	*
Dwelling facilities	*	*	*

4.1.3 Model Fit

Two available measures for evaluating the model fit of the three hedonic models, are adjusted explanatory power (adjusted R^2) and Root-Mean-Square Error *(RMSE)*. Adjusted R^2 is the amount of variation in the apartments' sales price per square meter explained by the included independent variables. RMSE, on the other hand, represents the standard deviation of the residuals, and can be used to compare alternative models on the same dataset.

4.1.4 Residual Analysis

Once the hedonic models are estimated with OLS, a residual analysis needs to be performed in order to check if the assumptions for the OLS estimation method are fulfilled and estimation results reliable. Here, the residual analysis is composed of assessments of multicollinearity, heteroscedasticity, and standardized residuals.

Multicollinearity is a problem due to relatively strongly correlated explanatory variables, which is a violation of the OLS assumptions. In the case of multicollinearity, standard errors of the regression coefficients may be overestimated. This gives smaller t-values, which may further lead to non-significant explanatory variables. To examine whether there is multicollinearity among the explanatory variables in the models, Variance Inflaction Factors (VIF) can be calculated. In the case of no multicollinearity, the VIF scores have lowest value equal to one, while VIF scores above five can indicate problems with multicollinearity (Hammervold, 2020). However, VIF scores do not give a definite answer. An assessment of how severe the case of multicollinearity is needs to be done. In general, smaller samples are more sensitive for multicollinearity. The large sample size in this thesis makes it reasonable to expect the implications of possible multicollinearity to not be serious.

In the case of heteroscedasticity in the estimated models, the residual variance is not constant across observations, which is a breach of the assumptions of OLS. Heteroscedasticity implications are misestimated, often underestimated, standard errors of the regression coefficients.

Consequently, t-tests and explanatory power become unreliable. White's test tests for possible heteroscedasticity. If the p-value associated with White's test is equal to 0.000, this indicates the presence of heteroscedasticiy in the models. A measure to address the problem of heteroscedasticity can be the use of robust standard errors, which are heteroscedasticity corrected.

Another way to detect problems with regression, is by studying a residual plot with standardized residuals. The residual plot shows standardized residuals plotted against predicted values. Only random variation, white noise, is desired in the residual plot. White noise in the plot indicates that all patterns in the data are captured by the model, and thus the model is correctly specified. In the case of heteroscedasticity, increasing variance for increasing predicted values will be seen.

Further examination of the standardized residuals can identify useful characteristics of their distribution. Standardized residuals are desired normally distributed. This is not required for OLS, but for the testing of hypotheses in the regression model (Studenmund, 2017). Descriptive statistics of the standardized residuals should therefore show values within +/-3 (Hammervold, 2020). In addition, large absolute values of standardized residuals may indicate outliers. Next, skewness and kurtosis tests for standardized residuals should give an associated p-value above 0.05, in order to show normally distributed standardized residuals. The distribution of standardized residuals can be illustrated with a histogram including a normal distribution curve. Lastly, a standardized normal probability plot (P-P plot) compares standardized residuals against the normal distribution. If the model is correct and data normally distributed, the estimated curve will be along the 45-degree line (Hammervold, 2020). The standardized residuals will then have a have variance equal to that of the normal distribution.

4.2 Structual Equation Modeling

4.2.1 Structual Equation Modeling Method

Sometimes variables are latent variables, and traditional statistical methods do not suffice to solve the investigated problem. As a result, SEM is a widely used method in the researching field of economics, psychology, and sociology today (Jöreskog et al., 2016). SEM is an established statistical method using multivariate data to investigate complex relationships among latent and observed variables (Javid et al., 2019). A SEM model is a powerful generalization of the traditional linear regression model adding theoretical constructs as well (Bollen, 1989, as cited in Fallan et al., 1995). The model allows for simultaneous estimation of measurements and structural parameters, accounts for measurement errors, and produces diagnostic statistics for the model as a whole (Fallan et al., 1995).

The SEM model estimated in this thesis is an application of the LISREL system developed by Karl G. Jöreskog and Dag Sörbom. In its most general form, the model can be thought of as a factor-analytic model consisting of two parts that are estimated simultaneously: a measurement model and a structural model (Titman and Wessels, 1988). The measurement model describes how latent variables are indicated by the observed variables, represented by the factor loadings. The factor loadings are desired to be as high as possible, because when the indicators show high factor loadings and are significant, they are considered good indicators of the associated latent factor. Ringdal (2001) suggests a minimum requirement for the factor loadings of 0.4. A high factor loading is equivalent to a strong correlation between the observed indicators X and associated latent factor. Further, the structural model specifies the causal relationships among the latent variables (Jöreskog et al., 2016), represented by structural parameters which also are desired high. A SEM model is said to fit the observed data to the degree that the model-implied covariance matrix is equivalent to the empirical covariance matrix (Schermelleh-Engel et al., 2003).

Maximum Likelihood (ML) is the most extensively used fitting function for SEM models, and is the default estimator in LISREL. ML provides two chi-squares which require multivariate normally distributed data, C1 and C2NT. However, Robust Maximum Likelihood (RML) is often used if data deviates from normality. RML provides ML parameter estimates, in addition to robust standard errors, chi-squares, and goodness-of-fit indices. With RML estimation, the following four additional non-normality corrected chi-squares are calculated: C2NNT, C3, C4, and C5. C3 provides the Satorra-Bentler scaled chi-square, which was proposed by Satorra and Bentler (1988), and is commonly used to correct for non-normality. C2NNT and C3 are the preferred chi-squares in the case of a large sample size, and a large to moderate sample size, respectively. Due to the extensive dataset and the assumption of non-normality, RML is likely the preferred estimation technique, and C2NNT and C3 the preferred chi-squares in this thesis. In Chapter 5, tests for non-normality will be performed to examine these assumptions.

Within SEM, factor analysis methods are used to analyze the relationships among measured variables to determine whether the observed variables can be clustered into a smaller set of underlying factors (Thompson, 2004; Worthington & Whittaker, 2006, as cited in Bowen and Guo, 2012). Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are two approaches for determining the factor structure of a set of observed variables. The aim of both EFA and CFA is to obtain data reduction. The purpose of EFA is to uncover few factors that are the cause of the correlation between a large number of variables (Hammervold, 2020). By extracting a few factors the problem of multicollinearity can be reduced. Selecting the number of factors to include in the SEM model can be done by utilizing the eigenvalue criterion. The criterion says that the number of factors with eigenvalue greater than or equal to one is retained. An eigenvalue of one states that the factor explains as much variance as a single variable does (Hammervold, 2020). To test if the correlation matrix is suitable for factor analysis, the Likelihood Ratio test (LR test) is used. The LR test investigates if the included variables are uncorrelated. Moreover, EFA gives explanatory power for each of the variables, and is desired over 0.5 (Hammervold, 2020).

CFA, on the other hand, involves testing a theoretical measurement model where both the number of factors, and the relationship between the factors and the observed variables are pre-determined (Hammervold, 2020). CFA should be based on theory and/or the results of EFAs, and other relevant tests (Bowen and Guo, 2012). As accounted for in Chapter 1, there is limited literature using SEM within property value estimation. For that reason, EFA of the hyper-locality variables is used in advance of CFA when building the SEM models in this thesis. Defining correct latent variables that fit the model is a challenging process, and can be a potential limitation of using SEM.

4.2.2 Goodness-of-Fit

To examine if the SEM model fits data well, several goodness-of-fit (GoF) indices can be interpreted. Chi-square (χ^2) is the most common fit statistic used to evaluate the appropriateness of SEM models (Bowen and Guo, 2012). If the p-value associated with the χ^2 value is larger than 0.05, the model fits the data and indicates a good fit. The χ^2 test is a strict test that often disregards the model, because it tests if the error of approximation is equal to zero. The test also has a strict assumption of multivariate normality, and in addition the χ^2 statistic is sensitive to sample size. As result of this, alternative GoF measures have been developed (Schermelleh-Engel et al., 2003).

Schermelleh-Engel et al. (2003) presents several available GoF indices to evaluate model fit. Although there are no well-established guidelines for what minimal conditions constitute an adequate fit, Schermelleh-Engel et al. (2003) provides some rules of thumb for cut-values. Root Mean Square Error of Approximation (RMSEA), p-value for Test of Close Fit (Close-Fit test), and Standardized Root Mean Square Residual (SRMR), are three frequently highlighted measures. RMSEA values less than or equal to 0.05 indicate good fit, and values between 0.05 and 0.08 indicate acceptable fit. The Close-Fit test is a less stringent test than the χ^2 test, because it considers an error of approximation less than or equal to 0.05 as close fit. SRMR should be less than 0.05 for a good fit, whereas values smaller than 0.10 can be interpreted as acceptable.

Three other indices commonly used for evaluating are Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), and Comparative Fit Index (CFI). These are incremental indices, and compare the fit of the model with a baseline model, usually the independence model, which is a model in which all observed variables are uncorrelated (Jöreskog et al., 2016). A NFI between 0.95 and 1.00 indicates a good model fit, while values between 0.90 and 0.95 are interpreted as acceptable. A rule of thumb for NNFI is that values between 0.97 and 1.00 are indicative of good fit, whereas values between 0.95 and 0.97 may be interpreted as an acceptable fit. Further, a CFI value between 0.95 and 0.97 indicates acceptable fit, and above 0.97 indicates good fit. Other indices available to explain model fit are Critical N, Goodness-of-Fit Index (GFI), and Adjusted Goodness-of-Fit Index (AGFI). Critical N should be higher than 200, GFI between 0.95 and 1.00, and AGFI between 0.90 and 1.00, for a good model fit. However, a GFI and AGFI between 0.90 and 0.95, and 0.85 and 0.90, respectively, are acceptable. Moreover, a large difference between GFI and NFI indicates presence of considerable noise in the model. Finally, it is important to have in mind that these rule of thumbs are guidelines for good practice and not absolute rules.

4.2.3 Hypotheses

As mentioned in the introduction, "Location, location, location" is a widely used slogan within real estate for a reason. Location is in general considered the most decisive factor when determining house prices. On that basis, it is reasonable to believe that proximity to attractive location amenities will increase apartment prices. An EFA of the hyper-locality variables will be executed in Chapter 5. Based on theory and common sense, the nine variables are expected to cluster into three factors: distance to school, distance to leisure, and distance to commerce. Families with school-age children most likely appreciate short distance to relevant schools and are willing to pay more, while others may be indifferent. Next, proximity to leisure related location amenities are expected to be valued by people in general, and thus increase the apartment price. In contrast to the school and leisure factor, the influence of distance to commerce on apartment prices is in lesser extent given. Some people love a bustling city life among stores, traffic, and industry, while others prefer to live in a more tranquil environment. Kauko (2003) finds that several studies confirm that proximity to industry affect property values negatively. However, the positive effects of proximity to commerce are expected to outweigh the negative effects for most people. This leads to the following hypotheses:

H_1: Increased distance to school has a negative effect on sales price per square meter.

H_2: Increased distance to leisure has a negative effect on sales price per square meter.

H_3: Increased distance to commerce has a negative effect on sales price per square meter.

The socio-economic variable presented in Subsection 3.2.7 is a scaled variable that serves as a proxy for socio-economic status. Segal (1979) found that the socio-economic characteristics of a neighbourhood can affect the attractiveness of a residential location, and thus the sales price. This is prevalent in neighbourhood formation because people believe they will find congenial friends if they live near people with similar social characteristics as themselves (Blair, 1995, as cited in Jordaan et al., 2004). Further, Balchin et al. (1995) presume that households with similar incomes are willing to incur the same price for traveling and housing. Thus, social desires can explain neighbourhood formation and housing prices, and the variable captures not only positive peer effects, but also a brand effect of living in an area considered attractive by the general public (Heyman and Sommervoll, 2019). A positive relationship between socio-economic status and apartments' sales price is therefore expected. This gives the following hypothesis:

H_4: Increased socio-economic status has a positive effect on sales price per square meter.

Some people will tend to move to costlier homes if traveling costs decrease. On the other hand, when traveling costs increase, certain groups will migrate to areas where the cost of housing is relatively cheaper (Jordaan et al., 2004). However, this potential inverse trade-off between travel costs and house prices are not given for all households, according to Jordaan et al. (2004). Anyhow, Hui et al. (2007) point out that earlier literature suggests that housing prices tend to be higher the closer the apartment is located to the central business district. In Oslo, the central business district is located in Sentrum, see Figure 2.1, and is surrounded by the inner city districts, defined in Subsection 3.2.6. In their own study, Hui et al. (2007) confirmed found theory, and concluded that traveling time from an apartment to the central business district was negatively correlated with the price of the dwelling. On this basis, it is hypothesized that:

H_5: Outer city has a negative effect on sales price per square meter.

As accounted for in Subsection 3.2.2, apartment size is expressed as the natural logarithm of living area due to the likelihood of diminishing returns as the apartment gets larger in square meters. Here, the dependent variable is sales price *per square meter* of the apartment. When the number of square meters increases, it makes sense to imagine that the sales price per square meter decreases. For this reason, the hypothesis is:

H_6: Increased apartment size has a negative effect on sales price per square meter.

As mentioned in Chapter 3, the analyses in this study is limited to transactions in the time period 2017 to 2021. Housing prices are characterized by fluctuations, but a relatively steady price growth has been witnessed over the time period in question. With 2017 as the reference year, the sales price development is expected to be positively correlated with increasing sales year. The house price growth is also seen in Statistics Norway's official price index for existing dwellings, where an increase in sales prices is seen each year in the relevant time period, see Appendix B.1. Following hypothesis is therefore formed:

H_7: Each sales year, from 2018 to 2021, has a positive effect on sales price per square meter, relative to reference year 2017. The positive effect are expected to be increasing relative to the year before.

It is easy to imagine that a newer apartment has a higher sales price per square meter than an old one. Nevertheless, it is important to keep in mind that older apartments may have been built at the most attractive locations. Despite of this, Rehák and Síbert (2017) found that newly built apartment buildings are more expensive. Newer apartments have better quality and are more energy-efficient, due to new standards and rules of construction. This, combined with higher material and construction costs the recent decades, gives the following hypothesis:

H_8: Apartments built after year 2000 have a higher sales price per square meter, compared to apartments built before year 2000.

Dwelling facilities are important elements when purchasing a residential property, as they can have a significant impact on sales price. The five dwelling facilities in question were presented in Table 3.6. Firstly, it is expected that apartments in quiet areas are preferred. However, Hui et al. (2007) found the opposite effect, and argue that some are willing to sacrifice serenity for centrality. Consequently, the total effect of a quiet location is not necessarily given, but is here expected to be positively correlated to property price. Secondly, Hui et al. (2007) observed that households are willing to pay more for apartments with a nice view. Further, it is expected that property buyers are willing to pay a sales price premium for units with balconies. Mesthrige and Ka (2017) found that balconies exert a strong and positive effect on residential property price. Next, the presence of a fireplace consistently has a significant positive effect on selling price (Sirmans et al., 2005). Lastly, apartments in a building equipped with an elevator are expected to be more expensive (Rehák and Síbert, 2017). Based on the expected effects of the five dwelling facilities, the following hypothesis is formed:

H_9: The presence of dwelling facilities has a positive effect on sales price per square meter.

Figure 4.1 shows the theoretical SEM research model summarizing the nine stated hypotheses. The SEM research model coincides with hedonic Model 3. This specific model is selected for further SEM analysis because it is more generalizable to other cities, as it does not directly contain the administrative districts of Oslo, but the continuous variable S and city area. On top of that, hedonic Model 3 is chosen on the basis of that the contribution of the thesis is to look closer at the socio-economic aspect of location. All features in the SEM model correspond to the variables in the hedonic model.

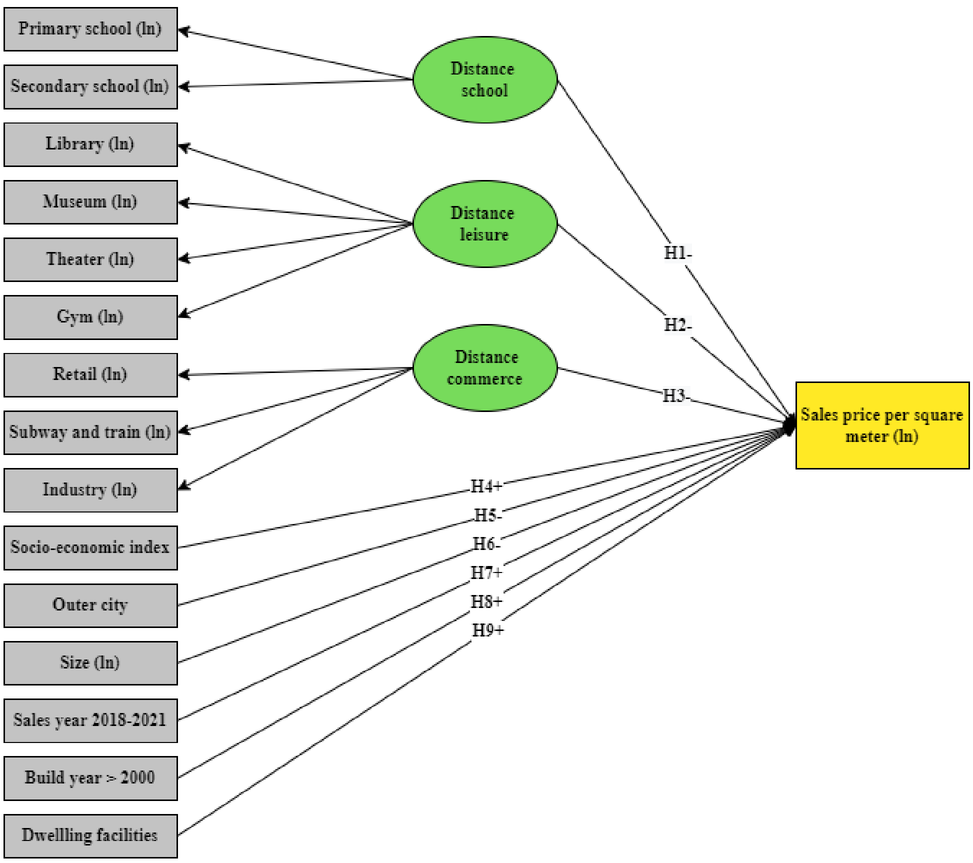


Figure 4.1: Theoretical SEM research model. Grey boxes denote the empirical indicators, green ovals denote hypothetical constructs, and the yellow box denotes the dependent variable in the model, sales price per square meter (ln). Measurement relations are represented with arrows from ovals to boxes. The arrows to the dependent variable express the hypotheses with associated expected signs.

4.2.4 Correlation Matrix for Latent and Observed Variables

When estimating a SEM model, LISREL also provides a covariance matrix for the variables included in the model. Additionally, standardized covariance, correlations, can be calculated. The calculated correlation matrix provides a better understanding of hyper-locality variables compared to the hedonic models. This is because it enables a bivariate examination of the correlation between the latent distance factors and the predicted sales price per square meter, where sales price per square meter is a function of all the independent variables. The correlations in the matrix thus express the correlations between the latent factors and all the variables combined.

4.2.5 Reliability and Validity in SEM

Before concluding with significant relationships in the structural model, the reliability and validity of the measurement model needs to be assessed in order to ensure high research quality. Reliability refers to the consistency of a measure, and several measures of reliability can be computed (Twycross and Shields, 2004). Individual item reliability is defined as the true score variance divided by the total variance. In LISREL, individual item reliability is computed directly and presented as squared multiple correlations for the X and Y variables. The R^2s are scaled from 0 to 1, and are desired as high as possible.

To further evaluate the SEM model's reliability, measures of the reliability of the latent factors can be examined. In this thesis, Composite Reliability (CR), Average Variance Extracted (AVE), and Cronbach's Alpha, are calculated by using unstandardized estimates for each of the three latent variables presented in Figure 4.1. Formulas for each measure are presented in Equation 4.2 to 4.4. Although individual item reliability captures reliability of a single indicator, it can be useful to expand the measure to include several indicators of a construct in order to more fully examine the shared variance in the measurement model. Therefore CR is calculated, which captures the conceptual reliability of the latent variable. Further, neither individual item reliability, nor CR, manages to measure the amount of variance that is captured by the construct in relation to the amount of variance due to measurement error (Fornell and Larcker, 1981). However, AVE provides this information and is an alternative measure calculated from a set of indicators of a latent variable. Lastly, Cronbach's Alpha is a useful coefficient for assessing internal consistency, meaning how closely related a set of indicators are as a group (UCLA Statistical Consulting Group, 2021).

$$CR = \frac{(\sum_i^r \lambda_i)^2}{(\sum_i^r \lambda_i)^2 + \sum_i^r var(\delta i)} \geq 0.6 \tag{4.2}$$

where λ_i is the unstandardized factor loading for item i and $var(\delta_i)$ is the variance of the error term for the ith indicator. CR examines the reliability for the composite of measures of a latent variable, and values above 0.6 are considered satisfactory (Fornell and Larcker, 1981).

$$AVE = \frac{\sum_i^r \lambda_i^2}{\sum_i^r \lambda_i^2 + \sum_i^r var(\delta i)} \geq 0.5 \tag{4.3}$$

where λ_i is the unstandardized factor loading for item i and $var(\delta_i)$ is the variance of the error term for the ith indicator. Values greater than 0.5 are desirable (Bagozzi and Yi, 1988), because this means that the variance due to measurement error is larger than the variance captured by the latent variable. This makes the validity of the individual indicators, as well as the latent variable, questionable (Fornell and Larcker, 1981).

$$Cronbach's\ Alpha = \frac{k * \overline{r}}{1 + (k-1) * \overline{r})} \geq 0.7 \tag{4.4}$$

where k is the number of items, and $\overline{r}$ is the average correlation between the indicators, and values above 0.7 are satisfactory (Bland and Altman, 1997).

In the SEM model, the hyper-locality indicators will be grouped into custom factors based on theory and common sense. In addition, as accounted for in Section 4.2.1, EFA will be performed in advance of building the SEM model in Chapter 5 to examine if the nine hyper-locality variables cluster as expected. In this way, a high degree of concept validity can be ensured.

Chapter 5

Empirical Results

In this chapter, the aim is to examine the impact of location variables and factors on the sales price per square meter for apartments in Oslo. For estimating the hedonic models and SEM model presented in Chapter 4, the statistical software Stata and LISREL were used, respectively. In the following chapter, the model outputs and empirical results will be presented. Firstly, the output and results of the three hedonic models is examined, followed by a comparison of the models' fit. Subsequently, the complete SEM model is presented, and factor loadings, structural parameters, covariances, and model fit, are interpreted. In addition, assessments of reliability and validity of the SEM model are discussed.

5.1 Hedonic Models

5.1.1 Estimated Hedonic Models

An assessment of how the sales price per square meter is explained by the included independent variables is provided by Table 5.1. Here, the estimations of Model 1, Model 2 and Model 3, are presented. The hedonic variables size, sales year, build year, and dwelling facilities, stay fixed across the three models. Subsection 3.2.1 revealed non-normality in the dependent sales price variable. However, normal distribution in itself is not an assumption of OLS estimation, as discussed in Subsection 4.1.4, and deviation from normality will thus not affect the estimation of the models (Studenmund, 2017).

Table 5.1: Estimation of the dependent variable sales price per square meter in the three hedonic models. Unstandardized regression coefficients are presented. Robust standard errors $SE(\hat{\alpha})$ are reported in parentheses. Variables related to size, sales year, build year, and dwelling facilities are included in all three models, but not further presented. See Subsection 3.2 for variable definitions and Appendix C.1 for full model estimations.

	Model 1 Baseline model	**Model 2** Hyper-locality model	**Model 3** Socio-economic model
Grünerløkka	0.036***(0.002)	0.041***(0.002)	
Sagene	0.086***(0.002)	0.098***(0.003)	
St. Hanshaugen	0.153***(0.003)	0.064***(0.003)	
Frogner	0.247***(0.003)	0.158***(0.004)	
Ullern	0.106***(0.004)	0.141***(0.004)	
Vestre Aker	0.009**(0.004)	0.079***(0.004)	
Nordre Aker	0.051***(0.004)	0.087***(0.004)	
Bjerke	-0.174***(0.003)	-0.075***(0.004)	
Grorud	-0.380***(0.004)	-0.218***(0.005)	
Stovner	-0.445***(0.004)	-0.279***(0.005)	
Alna	-0.320***(0.003	-0.190***(0.004)	
Østensjø	-0.209***(0.003)	-0.103***(0.004)	
Nordstrand	-0.132***(0.004)	-0.028***(0.005)	
Søndre Nordstrand	-0.487***(0.005)	-0.363***(0.006)	
Distance primary school (ln)		0.012***(0.001)	0.006***(0.001)
Distance secondary school (ln)		0.010***(0.001)	0.020***(0.001)
Distance library (ln)		-0.093***(0.002)	-0.102***(0.001)
Distance museum (ln)		0.001(0.001)	-0.023***(0.001)
Distance theater (ln)		0.010***(0.001)	0.019***(0.001)
Distance gym (ln)		-0.013***(0.001)	-0.019***(0.001)
Distance retail (ln)		0.011***(0.001)	0.012***(0.001)
Distance subway and train (ln)		-0.012***(0.001)	-0.012***(0.001)
Distance industry (ln)		0.011***(0.001)	0.019***(0.001)
Socio-economic index			0.003*** (0.000)
Outer city			-0.113***(0.003)
Size	YES	YES	YES
Sales year	YES	YES	YES
Build year	YES	YES	YES
Dwelling facilities	YES	YES	YES
Constant	12.266	12.373	12.250
Adjusted R^2	0.745	0.766	0.745
RMSE	0.141	0.135	0.141
Number of observations	62,134	62,134	62,134

Note: ***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

With Gamle Oslo as the baseline district, Model 1 indicates that district Grünerløkka, Sagene, St. Hanshaugen, Frogner, Ullern, and Nordre Aker, are expected to have a higher sales price per square meter compared to Gamle Oslo, all other variables kept constant. If an apartment is located at Frogner, it is expected to increase sales price per square meter with 24.7% compared to Gamle Oslo. Further, the remaining districts are expected to give a lower sales price per square meter, as seen from their negative coefficient signs. Finally, each district is significant at the 1% level.

In Model 2, the nine hyper-locality features are also included. They describe the effect of hyper-locality on sales price per square meter. The inclusion of the variables has led to a small change in the district coefficients. However, the coefficient signs are consistent with Model 1, and all the districts are still significant at the 1% level. Table 5.1 shows that when the distance to the closest library increases with 1%, sales price per square meter is expected to be reduced with 0.093%, while holding all the other predictors constant. Thus, 0.093 is the elasticity of sales price per square meter with respect to distance to library. Distance to gym, and distance to subway and train, also have a negative coefficient, and are interpreted analogously with distance to library. However, when the distance to the closest industrial building increases with 1%, sales price per square meter increases with 0.011%, holding all the other predictors constant. This corresponds to an elasticity of sales price per square meter, with respect to distance to an industrial building, of 0.011. Further, distance to the closest primary school, secondary school, museum, theater, and retail store all have positive coefficients, and are interpreted the same way.

Generally, most of the coefficient signs of Model 2 seem reasonable, yet the coefficient signs of distance to primary school, secondary school, theater, and retail store can be questioned and will be discussed in Chapter 6. In addition, all of the hyper-locality variables in Model 2 are significant at the 1% level, except for distance to museum which has a coefficient equal to 0.001 and is non-significant. This implies that distance to museum cannot be claimed to have an effect on the dependent variable, and can be a result of chance.

In Model 3, the continuous socio-economic variable and the binary city area variable are included as a substitute for the original district variable, as seen in Table 5.1. These are included to more clearly distinguish between the effect of travel expenses, and the effect related to status and socio-economic conditions. The coefficient signs of the hyper-locality features in Model 3 are largely the same as in Model 2, and significant at the 1% level. However, the coefficient sign of distance to museum has changed. Nonetheless, in contrast to Model 2, the variable is now significant and can be said to have an effect on the dwelling's price per square meter. The city area variable has a significant, negative coefficient with a value of -0.113. The variable is binary with inner city as reference group. In other words, if the district is located in the outer part of the city, *ceteris paribus*, the apartment's sales price per square meter decreases with 11.3%.

Ultimately, the socio-economic variable has a positive coefficient sign, which is as expected, and is additionally significant at the 1% level. Table 5.2 presents an overview of the total marginal effects of the socio-economic index on each of the 15 districts. The sales price per square meter in Nordre Aker, Vestre Aker, and Ullern, are most positively affected by the districts' status, with an marginal increase of 29.4%, 29.1%, and 28.5%, in sales price per square meter, respectively. On the other hand, Stovner, Søndre Nordstrand, and Alna, have the lowest marginal increase in price per square meter based on socio-economic conditions, with an increase of 0.3%, 3.3%, and 4.8%, respectively. In general, the outer part of Oslo West has the highest ranked status, while the outer part of Oslo East has the lowest ranked socio-economic status, with an exception of Nordstrand and Østensjø. Further, the inner city districts surrounding Sentrum have a moderate marginal increase in sales price per square meter due to socio-economic status.

Table 5.2: The total marginal effect of the socio-economic index on each district of Oslo. Total effect equals the coefficient of the socio-economic index ($\hat{\alpha}_s = 0.003$) multiplied with the standardized individual socio-economic score S for each district, all other variables kept constant.

District	**Socio-economic score (S)**	**Total effect**
Gamle Oslo	37	0.111
Grünerløkka	56	0.168
Sagene	68	0.204
St. Hanshaugen	70	0.210
Frogner	67	0.201
Ullern	95	0.285
Vestre Aker	97	0.291
Nordre Aker	98	0.294
Bjerke	42	0.126
Grorud	18	0.054
Stovner	1	0.003
Alna	16	0.048
Østensjø	76	0.228
Nordstrand	92	0.276
Søndre Nordstrand	11	0.033

A comparison of the impact of the various hyper-locality variables on apartments' sales price per square meter in Model 2 and Model 3, is interesting to take a closer look at. To assess their estimated relative importance in the hedonic models, the unstandardized regression coefficients are appropriate for a comparison of the hyper-locality variables, as they have the same unit of measurement. They are therefore comparable without the use of the standardized beta which is often criticized in statistics literature (Kryvobokov and Wilhelmsson, 2007).

As seen in Table 5.1, distance to library was found to be the hyper-locality variable with the highest relative importance in both Model 2 and Model 3, followed by distance to gym in Model

2 and distance to museum in Model 3. Increased proximity to these three variables is expected to have a significant positive effect on sales price per square meter. The hyper-locality variables with the least effect on sales price per square meter are distance to theater and secondary school in Model 2 and distance to retail store in Model 3.

5.1.2 Model Fit

The inclusion of hyper-locality variables in Model 2, leads to an increase in adjusted R^2 with 2.1% compared to Model 1. The estimations give an adjusted R^2 of 0.745 and 0.766 in Model 1 and Model 2, respectively. 76.6% of the observed variance in the apartments' sales price per square meter can be explained by Model 2, which indicates that it adapts the data satisfyingly. Finally, Model 3 has an adjusted coefficient of determination of 74.5%. This indicates that the inclusion of the socio-economic index and city area still gives a good explanation of the variance in dwellings' sales prices per square meter. Model 1 and Model 3 have a RMSE of 0.141, compared to 0.135 in Model 2. Based on this measure, Model 2 relatively predicts the data most accurately, as RMSE is desired as low as possible.

5.1.3 Residual Analysis

Residual analysis is useful in order to further evaluate the model fit of a regression model. Detailed output of the residual analysis is provided in Appendix C.2. Firstly, assessments of multicollinearity indicate no multicollinearity for Model 1, with VIF scores below five for all explanatory variables. Further, the inclusion of hyper-locality variables in Model 2, leads to VIF scores above five for distance to library, distance to museum and, distance to gym. The estimated regression coefficient of distance to museum is non-significant in Model 2 and can perhaps be explained with the presence of multicollinearity in the model, which may overestimate the standard error of the regression coefficient. In Model 3, the dummy for city area has a VIF score equal to five, which indicates multicollinearity and that the standard errors of the regression coefficients may be overestimated. However, the effect of multicollinearity is not severe, because all explanatory variables in Model 3 are significant at the 1% level.

White's test for heteroscedasticity has an associated p-value equal to 0.000 for all three hedonic models, which indicates that heteroscedasticity has led to an underestimation of the standard errors of the regression coefficient. In order to take the discovered heteroscedasticity into account, robust standard errors are reported in Table 5.1. This ensures more reliable t-tests in relation to whether the variables can be claimed to be significant.

Further, the residual plots for all three hedonic models suggest a fan shape with increasing variance for increasing predicted values. This confirms the heteroscedasticity found with White's test. Standardized residuals show minimum and maximum values above +/-3, which can indicate outliers. However, outliers were considered in Subsection 3.1.2, and are in this case therefore not further inspected on the basis of standardized residuals. Minimum and maximum values of the standardized residuals, in addition to tests for skewness and kurtosis with associated p-values of 0.000, indicate significant non-normality for the standardized residuals. The characteristics of the standardized residuals are seen in the histograms as well. The P-P plots for all three models show curves close to the 45-degree line. This indicates that the standardized residuals have variance equal to that of the normal distribution.

Overall, the assessments of multicollinearity, heteroscedasticity, and standardized residuals, suggest that the estimation results are reliable, as violations of the OLS assumptions are taken into account.

5.2 Structual Equation Modeling

5.2.1 Test for Multivariate Normality

As discussed in Subsection 4.2.1, ML estimation requires multivariate normal distribution. To select the most suitable estimation method for estimating the SEM model, the data thus must be investigated for non-normality through tests for multivariate normal distribution.

Table 5.3: Tests for multivariate normality for the SEM model. The left column shows a test for multivariate skewness, the middle column a test for multivariate kurtosis, and the column to the right an overall test for multivariate skewness and kurtosis.

Skewness			**Kurtosis**			**Skewness and kurtosis**	
Value	Z-Score	P-Value	Value	Z-Score	P-value	Chi-Square	P-value
55.670	539.723	0.000	594.090	60.163	0.000	294,920.760	0.000

Table 5.3 shows a test for multivariate skewness, a test for multivariate kurtosis, and an overall test for multivariate skewness and kurtosis for the data. The tests for multivariate normal distribution conclude that the assumption of multivariate normality does not hold, as the p-values for all three tests are 0.000, and the tests reject the null hypothesis of normality. In other words, the model estimation should be corrected for non-normality. Therefore, as expected, RML is the preferred estimation method for estimating the SEM model.

5.2.2 Explanatory Factor Analysis

An EFA was conducted to achieve the recommended number of latent factors and data reduction for further SEM analysis. Additionally, multicollinearity can be mitigated. The factor analysis was performed using the principal-component factor method. The results of the EFA are shown in Table 5.4.

Table 5.4: Explanatory factor analysis for hyper-locality variables using the principal-component factors method. The table shows unrotated factor loadings. Three factors are retained with respect to the eigenvalue criterion.

Factor	**Eigenvalue**	**Difference**	**Proportion**	**Cumulative**
Factor 1	3.788	2.446	0.421	0.421
Factor 2	1.342	0.272	0.149	0.570
Factor 3	1.070	0.250	0.119	0.689
Factor 4	0.820	0.105	0.091	0.780
Factor 5	0.714	0.227	0.079	0.859
Factor 6	0.488	0.084	0.054	0.914
Factor 7	0.404	0.203	0.045	0.958
Factor 8	0.201	0.028	0.022	0.981
Factor 9	0.173	.	0.019	1.000

LR test: independent vs. saturated: chi2(36) = 2.4e+05 Prob>chi2 = 0.000

As seen from Table 5.4, three factors have an eigenvalue greater than one: Factor 1, Factor 2, and Factor 3. By extracting these factors, clusters of variables that explain more than each individual variable are formed (Hammervold, 2020). Thus, a data reduction from nine to three factors are implemented. Further, Table 5.4 highlights that the three factors explain 68.9% of the variance in the variables, which is relatively good. The LR test in Table 5.4 shows a p-value of 0.000, which is equivalent to a rejection of the null hypothesis of an orthogonal correlation matrix. This implies that the variables are correlated and a factor analysis with data reduction can be implemented.

Table 5.5: Promax rotated factor loadings and unique variances for the hyper-locality variables. The factor analysis of the nine hyper-locality variables resulted in three factors.

Variable	**Factor 1**	**Factor 2**	**Factor 3**	**Uniqueness**
Distance primary school	-0.184	0.037	0.918	0.207
Distance secondary school	0.189	-0.095	0.801	0.271
Distance library	0.970	-0.129	0.004	0.134
Distance museum	0.903	-0.020	-0.063	0.224
Distance theater	0.801	0.023	0.027	0.331
Distance gym	0.910	0.031	-0.037	0.168
Distance retail	-0.173	0.889	-0.041	0.301
Distance subway and train	0.279	0.337	0.139	0.682
Distance industry	0.198	0.622	0.003	0.482

Rotated factor loadings for each of the three factors are presented in Table 5.5. Here, an oblique rotation technique, more specifically the promax rotation, is used, which allows correlated factors (Hammervold, 2020). It is plausible to believe that the distance factors are correlated. Furthermore, Table 5.5 makes it clear that Factor 1 has high loadings for the variables associated with distance to leisure, Factor 2 for the distance variables related to commerce, and Factor 3 for the distance variables related to school. The expectation of three factors explained in Subsection 4.2.3 was thus confirmed by the EFA. Moreover, the explanatory power of each variable are represented by one minus uniqueness, where the latter is presented in Table 5.5. The explanatory power is over 0.5 for all variables, except distance to subway and train. Only 31.8% of the variance of distance to subway and train is explained by the factors in the model. Despite this, the remaining variables have high explanatory power, which indicates good reliability.

5.2.3 Estimated SEM Model

Figure 5.1 shows the estimated SEM model based on EFA and CFA, and the relationships between constructs derived from the nine hypotheses presented in Subsection 4.2.3. The model was estimated by applying RML to correct for non-normality, as discussed in Sebsection 5.2.1. As mentioned in Subsection 4.2.3, the SEM model coincides with hedonic Model 3.

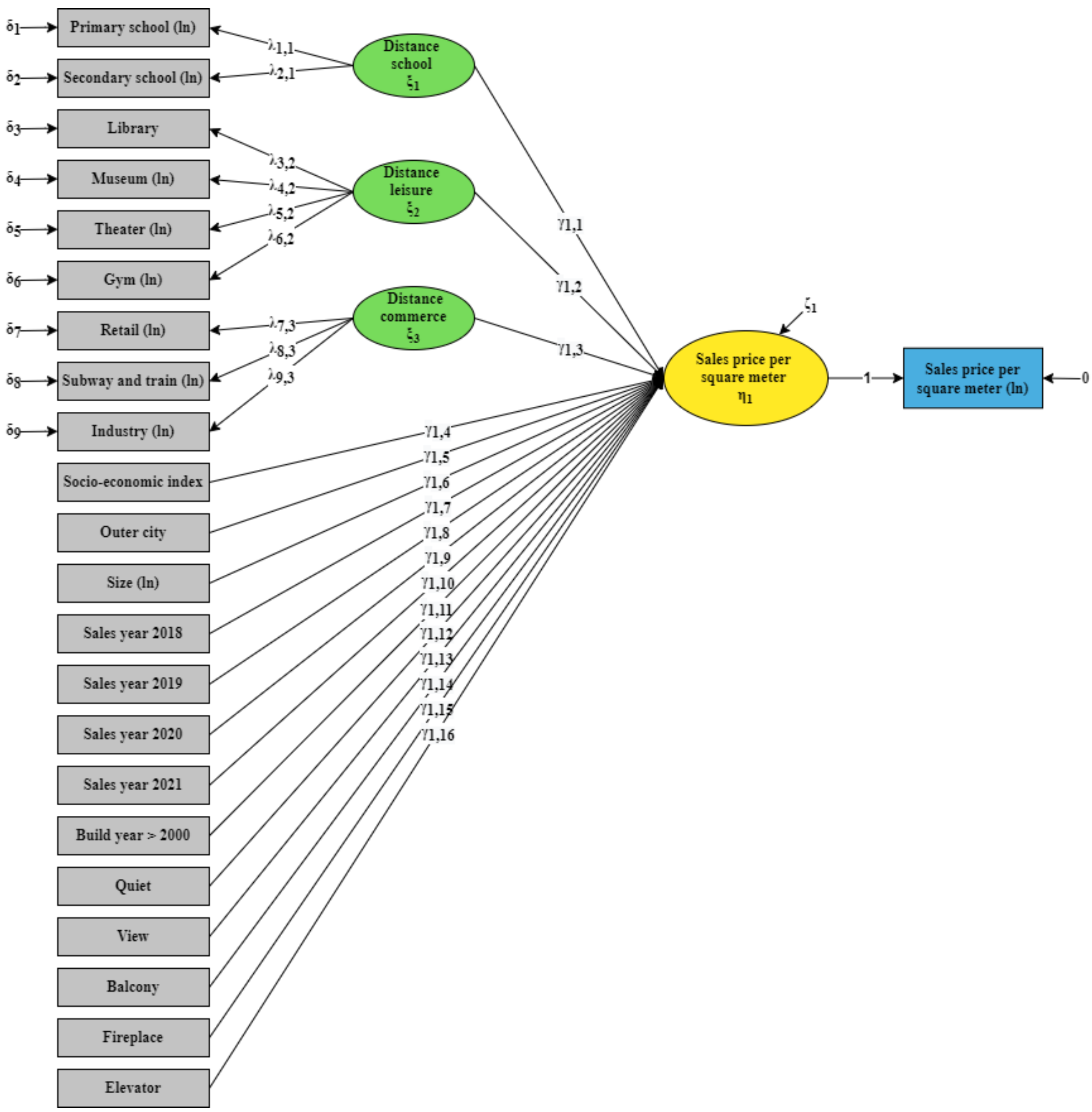


Figure 5.1: Estimated SEM model. The model is a recursive structural model with three latent variables and 23 observed variables. Green ovals denote hypothetical constructs ξ_i, and the yellow oval denotes the dependent variable in the model, η_1. The grey boxes denote empirical indicators X, while the blue box denotes the empirical dependent indicator Y. Measurement relations are represented with arrows from ovals to boxes, and arrows from the boxes express measurement error terms. Lower Greek letters express unstandardized parameters to be estimated. $\lambda_{i,j}$ describes factor loadings and refers to the correlation between the observed indicators and their belonging ξ_i. The structural parameters $\gamma_{i,j}$ show the effect from ξ_i to η_1. Measurement errors for each X-indicator is represented by δ_i. ζ_1 is the error term of the dependent variable.

The structure of a full LISREL model, in its general form, was explained in Section 4.2.1. The LISREL model consists of two measurement models, one related to the X's, and one for the Y's. See Appendix C.3.1 for the equations of the measurement models and the structural model. The observed X-indicators socio-economic index, outer city, primary room, sales year 2018 to

2021, build year > 2000, and the five dwelling facilities, are fixed. Furthermore, the observed Y-indicator sales price per square meter (ln) is also fixed. This implies that the factor loading is set equal to 1, as they alone are measuring their respective factor. For this reason, the standard errors and t-values are not given for these indicators. Estimation results, excluded the fixed indicators, are presented in Table 5.6.

Table 5.6: Estimation of the SEM model, presented with estimated unstandardized parameters and associated standard errors and t-values. R^2 is presented for the unstandardized factor loadings $\lambda_{i,j}$.

Parameter	**LISREL estimate**	**Standard error**	**t-value**	**R^2**	**Parameter**	**LISREL estimate**	**Standard error**	**t-value**
$\lambda_{1,1}$	0.308	0.003	101.828***	0.235	$\gamma_{1,1}$	0.018	0.001	21.814***
$\lambda_{2,1}$	0.719	0.002	315.759***	0.981	$\gamma_{1,2}$	-0.173	0.002	-80.262***
$\lambda_{3,2}$	0.831	0.002	363.540***	0.827	$\gamma_{1,3}$	0.059	0.003	18.853***
$\lambda_{4,2}$	0.863	0.003	340.439***	0.709	$\gamma_{1,4}$	0.002	0.000	48.846***
$\lambda_{5,2}$	0.794	0.003	286.560***	0.575	$\gamma_{1,5}$	-0.083	0.006	-14.702***
$\lambda_{6,2}$	0.882	0.002	364.798***	0.807	$\gamma_{1,6}$	-0.311	0.003	-107.254***
$\lambda_{7,3}$	0.187	0.003	56.972***	0.068	$\gamma_{1,7}$	0.000	0.002	-0.230
$\lambda_{8,3}$	0.439	0.004	124.485***	0.273	$\gamma_{1,8}$	0.034	0.002	17.009***
$\lambda_{9,3}$	0.470	0.004	132.130***	0.343	$\gamma_{1,9}$	0.092	0.002	44.844***
					$\gamma_{1,10}$	0.190	0.002	91.620***
δ_1	0.309	0.002	127.872***		$\gamma_{1,11}$	0.148	0.002	70.319***
δ_2	0.010	-	-		$\gamma_{1,12}$	0.020	0.002	11.293***
δ_3	0.145	0.002	90.620***		$\gamma_{1,13}$	0.037	0.001	27.907***
δ_4	0.306	0.003	113.819***		$\gamma_{1,14}$	0.035	0.002	15.645***
δ_5	0.466	0.004	131.873***		$\gamma_{1,15}$	0.046	0.002	27.176***
δ_6	0.186	0.002	74.303***		$\gamma_{1,16}$	0.004	0.002	2.240**
δ_7	0.478	0.003	137.027***					
δ_8	0.514	0.004	131.901***		ζ_1	0.019	0.000	109.404***
δ_9	0.422	0.003	120.832***					

Note: ***Significant at the 1% level, ** Significant at the 5% level, *Significant at the 10% level.

Measurement Model

Table 5.6 shows that all of the unstandardized factor loadings, $\lambda_{i,j}$, are higher than the minimum requirement of 0.4 (Ringdal, 2001), except $\lambda_{1,1}$ and $\lambda_{7,3}$. This indicates that distance to primary school is a weak indicator for the latent factor distance to school, and retail a weak indicator for the latent factor distance to commerce, in the model. On the other hand, distance to secondary school is a good indicator for the school factor ($\lambda_{2,1} > 0.4$), and additionally distance to subway and train and distance to industry are good indicators for the commerce factor ($\lambda_{8,3} > 0.4$, $\lambda_{9,3} > 0.4$). As mentioned in Subsection 4.2.1, high factor loading is equivalent to a strong correlation between the observed indicators X and associated latent factor, ξ_i. Next, the factor loadings $\lambda_{3,2}$, $\lambda_{4,2}$, $\lambda_{5,2}$, and $\lambda_{6,2}$ are particularly high, with a factor loading of over 0.75 with ξ_2. This indicates that the observed indicators for distance to libraries, museums, theaters, and gyms, measure what they intend to measure, and thus are good indicators for measuring distance to leisure. Overall, there are small standard errors and high t-values for the indicators in the measurement model of the X's, and all of the parameters are significant.

The measurement errors, δ_i, are seen in 5.6, and are desired to be as low as possible. If the variance of the error terms are high, this corresponds to low R^2 for the associated X-indicator, and thus lower reliability. Reliability is further discussed in Subsection 5.2.5. δ_2 was fixed to be 0.01, because the estimation of the model gave a small negative value for this parameter, which is a sign of illness in the model. A small correction like this did not affect the model in general, but since the parameter is fixed, standard error and t-value were not computed.

Structural Model

The structural parameters, $\gamma_{i,j}$, are also presented in Table 5.6. The parameters represent the relationship between sales price per square meter of an apartment, and all of the latent variables. In the structural model, the estimated structural parameters $\gamma_{i,j}$ are directly linked to the developed research hypotheses, see Subsection 4.2.3.

The expected negative relationship between the apartment's sales price per square meter and increased distance to schools (H_1) was not supported ($\gamma_{1,1}$ = 0.018). H_2 stated that the expected relationship between sales price per square meter and increased distance to leisure was negative, and was strongly supported ($\gamma_{1,2}$ = -0.713). Further, the relationship between the price of an apartment and increased distance to commerce also was proposed to be negative (H_3), which was not supported ($\gamma_{1,3}$ = 0.059). Overall, all of the estimated structural parameters related to the three latent location factors were found significant at the 1% level.

The next hypothesis, H_4, implied that increased socio-economic status of the district would be positively related to the apartment's sales price, and was supported ($\gamma_{1,4}$ = 0.002). H_5 stated that if the apartment was located in the outer city, the sales price per square meter would be negatively affected, and the hypothesis was supported ($\gamma_{1,5}$ = -0.083). Next, the relationship between increased living area in square meters and sales price of the apartment was expected to be negative (H_6). The estimated structural parameter ($\gamma_{1,6}$ = -0.311) is in support of the stated hypothesis. H_7 proposed that all of the sales years 2018 to 2021 had an increasing positive effect on sales price per square meter per year, relative to reference year 2017. The hypothesis is supported for the sales years 2019-2021 ($\gamma_{1,8}$ = 0.034, $\gamma_{1,9}$ = 0.092, $\gamma_{1,10}$ = 0.190). However, the structural parameter of sales year 2018 ($\gamma_{1,7}$ = 0.000) indicates that the apartment's sales price per square meter is not affected by this year. However, sales year 2018 is not a significant structural parameter in the model. Further, the structural parameter for hypotheses H_8 ($\gamma_{1,12}$ = 0.020) is in support of the hypotheses that apartments built after year 2000 have a higher sales price per square meter, compared to apartments built before year 2000. Lastly, hypothesis (H_9) proposed that the presence of dwelling facilities would have a positive effect on sales price per square meter. The statement is supported by the belonging structural parameters of the dwelling facilities ($\gamma_{1,13} = 0.037$, $\gamma_{1,14} = 0.035$, $\gamma_{1,15} = 0.046$, $\gamma_{1,16} = 0.004$).

All in all, the standard errors of the structural model are small for all of the parameters, and associated t-values are mostly high. All the structural parameters are significant at the 1% level, except for $\gamma_{1,7}$ which is non-significant.

Table 5.7: Coefficient of determination for the dependent variable sales price per square meter (η_1).

	R^2
η_1	0.758

Table 5.7 shows that 75.8% of the variance in sales price per square meter (η_1) is explained by the independent variables in the model. The high coefficient of determination indicates that relevant variables are included in the model. This could also be seen by low the error term of the dependent variable ζ_1, in Table 5.6.

5.2.4 Correlation Matrix for Latent and Observed Variables

Table 5.8: Correlation matrix for sales price per square meter, latent location variables, socio-economic index, and outer city.

	Sales price per square meter	**Distance school**	**Distance leisure**	**Distance commerce**	**Socio-economic index**	**Outer city**
Sales price per square meter	1.000					
Distance school	-0.046***	1.000				
Distance leisure	-0.195***	0.350***	1.000			
Distance commerce	-0.124***	0.292***	0.673***	1.000		
Socio-economic index	0.112***	0.081***	-0.303***	0.111***	1.000	
Outer city	-0.159***	0.442***	0.863***	0.794***	-0.007*	1.000

Note: ***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

The SEM model and the γ parameters give the effect of the latent independent variables on the sales price per square meter, separately. On the other hand, the correlation matrix presented in Table 5.8 expresses the correlation between the latent variables, and the observed variables socio-economic index and outer city. The bivariate correlation between two variables, is controlled for the effect of all other variables combined. See Appendix C.4 for the full correlation matrix. The observed variables socio-economic index and outer city are included due to their interesting interpretation. All correlations are significant at the 1% level, except the correlation between socio-economic index and outer city, which is significant at the 10% level.

As seen in Table 5.8, sales price per square meter is negatively correlated with all three latent location variables. This means that an increase in distance, either to school, leisure, or commerce, results in a reduction in sales price per square meter, controlled for the effect of all other variables combined. Distance to leisure is the latent variable with the strongest correlation with sales price per square meter, with a correlation of -0.195, followed by distance to commerce which has a correlation of -0.124 with sales price per square meter. Distance to school has the lowest correlation with sales price per square meter, with a correlation of -0.046.

Furthermore, the effect of distance to school and distance to commerce on the sales price per square meter, separately, is small and positive, meaning $\gamma_{1,1}$ and $\gamma_{1,3}$ are small and positive. Consequently, the correlation sign can easily change to negative when all other variables are controlled for. This can explain why distance to school and distance to commerce are negatively correlated with sales price per square meter, although the associated γ parameters are positive in the estimated SEM model. Here, some variables can dominate with negative covariances, and thus contribute to negative correlation in the correlation matrix.

Next, the correlations between the latent location factors are all positive. The correlation between distance to leisure and distance to commerce stands out as stronger than the other correlations, with a correlation of 0.673. Distance to leisure and distance to commerce are both positively correlated with distance to school, with correlations of respectively 0.350 and 0.292.

Further, an increase in socio-economic status in a given district is associated with an increase in sales price per square meter, due to the positive correlation of 0.112. Next, socio-economic status is negatively correlated with distance to leisure, with a correlation of -0.303. This gives that the higher the status in a given district, the closer the apartments will be located to leisure activities. Lastly, socio-economic status is weakly positively correlated with distance to school and commerce, with correlations of 0.081 and 0.111, respectively.

Finally, outer city is negatively correlated with both sales price per square meter and socio-economic status. Moreover, outer city is positively correlated with the three latent location variables. In other words, if the apartment is located in the outer districts of Oslo, it is associated with increased distance to schools, leisure and, commerce. Outer city and distance to leisure and distance to commerce have a strong positive correlation of 0.863 and 0.794, respectively. The correlation between outer city and distance to school is weaker than outer city and the two aforementioned latent variables, with a correlation of 0.442, controlled for the effect of all other variables combined. From Table 5.8 it is observed that the correlation between outer city and socio-economic status is low and negative, with a value of -0.007.

5.2.5 Reliability and Validity

To enhance the quality of the study, assessments of reliability and validity are beneficial. One sees from Table 5.6 that the computed coefficients of determination of the X-indicators are diverse. A high R^2 indicates good reliability for the measurement model of X. Distance to retail has the lowest R^2 with a value of 0.068, and distance to secondary school has the highest with a value of 0.981. In other words, 6.8% of the variance in distance to retail is explained by the latent variable distance school, while 98.1% of the variance in distance to secondary school is explained by distance commerce. Even though they vary, most of the coefficients of determination are moderate or high, which indicates good reliability in the measurement model of X.

Table 5.9: Tests of reliability for the latent factors. The tests presented are Composite Reliability, Average Variance Extracted, and Cronbach's Alpha.

Latent factors	**CR**	**AVE**	**Cronbach's Alpha**
Distance school	0.768	0.657	0.655
Distance leisure	0.911	0.721	0.910
Distance commerce	0.459	0.241	0.447

For a further assessment of the SEM model's reliability, the reliability measures mentioned in Subsection 4.2.5 are calculated for each of the three latent variables and presented in Table 5.9. Both latent factors distance school and distance leisure have CR values above 0.6, AVE values above 0.5, and Cronbach's Alpha above 0.7. This indicates that the factors have good conceptual reliability, as well as they capture a high amount of variance and have high internal consistency. However, the latent factor distance commerce performs unsatisfactory across the three reliability tests, not managing to meet the requirements. This can be seen in connection with the results from the EFA in Table 5.5. Distance to retail had a rotated factor loading below the requirement of 0.4, in addition to an explanatory power beneath the requirement of 0.5. In general, the three variables that clustered into distance to commerce have relatively lower factor loadings than the other clustered variables. In spite of this, the distance to commerce factor is consistent with theory and common sense. Thus, a high degree of concept validity still are insured. Overall, the squared multiple correlations, CR, AVE, and Cronbach's Alpha, indicate sufficient reliability in the model. In addition, the latent factors disclosed from the EFA have good validity, as the clustering is in line with theory and what common sense would indicate.

5.2.6 Model Fit

Table 5.10: Goodness-of-Fit indices for the SEM model. The GoF indices for chi-squares C2NNT and C3 with related conclusions are presented.

GoF index	**C2NNT**	**Conclusion**	**C3**	**Conclusion**
χ^2	28,123.716	Bad fit	27,796.425	Bad fit
RMSEA	0.064	Acceptable fit	0.064	Acceptable fit
Close fit test (p-value)	1.000	Good fit	1.000	Good fit
SRMR	0.027	Good fit	0.027	Good fit
NFI	0.998	Good fit	0.951	Good fit
NNFI	0.996	Good fit	0.887	Bad fit
CFI	0.998	Good fit	0.951	Acceptable fit
GFI	0.962	Good fit	0.962	Good fit
AGFI	0.903	Good fit	0.903	Good fit

The C2NNT chi-square statistics of the model show bad fit, with a C2NNT χ^2 value of 28,123.716. However, RMSEA of 0.064 indicates acceptable fit, and the Close-fit test indicates good fit with its associated p-value of 1.000. A SRMR of 0.027 is close to zero, and suggests good fit. Further, all three incremental GoF indices, NFI, NNFI, and CFI, indicate a good model fit, with values of 0.998, 0.996, and 0.998, respectively. A GFI equal to 0.962 and AGFI equal to 0.903 both support the model and illustrate good fit. Moreover, the difference between GFI and NFI is not appreciable large, which indicates absence of significant noise in the model. All in all, the GoF indices presented support the estimated SEM model and suggest that the model adapts data well. The strict chi-square test is the only test to reject the model, which can be explained with the chi-square's sensitivity to large sample sizes.

The C3 chi-square statistics give a C3 χ^2 value of 27,796.425, which indicates bad fit similar to the C2NNT χ^2. Furthermore, C3 shows identical conclusions for the remaining GoF indices, with an exception of NNFI and CFI. NNFI now concludes with bad model fit, and CFI indicates an acceptable fit for the model. All in all, the C2NNT chi-square statistics indicate better model fit than the C3 chi-square statistics. This can be seen in connection with the fact that C2NNT is customized for large samples, which is the case here. In conclusion, the GoF indices, especially for C2NNT, in combination with reliability and validity assessments, indicate that the model has a good fit and is correctly specified.

Chapter 6

Discussion

This thesis has examined both direct and underlying effects of different location variables and factors on apartments' sales price in Oslo. A thorough review of the data led an extension of the baseline model (Model 1), through an inclusion of nine distance-measured hyper-locality variables. This was done in order to build a hedonic hyper-locality model (Model 2) with increased attention to the direct effect of alternative location explanatory variables. Furthermore, the study sought to achieve a deeper understanding of the underlying effects of location, through compiling a proxy variable for socio-economic status of the district, in addition to a travel cost variable. Therefore, a comprehensive socio-economic hedonic model (Model 3) included the two aforementioned variables, in addition to the hyper-locality variables. Eventually, a SEM model based on hedonic Model 3, was built for the purpose of uncovering potential location factors.

The hedonic modeling resulted in interesting findings. In Model 1, the coefficient signs associated with the districts of Oslo were as expected. As seen in Table 3.8, Gamle Oslo is the district with an average sales price per square meter closest to the average of Oslo as a whole on the upper side. With Gamle Oslo as the reference district, the coefficients in Model 1 were as expected considering the map shown in Figure 3.2. The results of Model 1 are therefore consistent with the average sales price per square meter for apartments in Oslo in the time period 2017 to 2021. In addition, the results illustrate the east-west division discussed in Chapter 1 to a large extent. Not unexpectedly, the districts in Oslo West are expected to have a higher sales price per square meter compared to Gamle Oslo. In contrast to the other districts in Oslo East, Grünerløkka and Sagene are expected to have a positive effect on the sales price per square meter, with Gamle Oslo as reference. Whether this is an effect of centrality in terms of distance to Sentrum, travel costs, status, or distance to services, is difficult to decide in Model 1, and it appears advantageous that Model 2 and Model 3 were introduced to assess underlying effects of location.

In Model 2, the many significant hyper-locality variables are consistent with the expectation that distance to key amenities and services affects willingness to pay and transaction prices. Further, the substitution of districts with the socio-economic index and the city area variable in

Model 3, led to a small decrease in explanatory power. However, Model 3 enables a discussion of the effect of the socio-economic status of the districts and travel costs on the apartment sales price per square meter. The coefficient sign of the socio-economic variable was as expected, and confirmed that an increased district socio-economic score gives a higher apartment sales price per square meter. It turned out that the socio-economic index captured the underlying effect of the socio-economic status of the districts, in contrast to the district variable in itself. The historical east-west division was confirmed, as the square meter prices in the outer Oslo West increased the most due to socio-economic status, while outer Oslo East had the smallest increase, with an exception of Nordstrand and Østensjø. These two districts differ from the other outer eastern districts in socio-economic conditions.

Further, the inner-outer district division of Oslo was positively correlated with the hyper-locality variables, as seen in Table 5.8. However, this division captures travel costs in a different way than the hyper-locality variables. The city area dummy variable illustrated that apartments located in the outer part of Oslo, are expected to have a lower sales price per square meter. This is consistent with the expectation that people are willing to pay more to live closer to the city center, due to lower travel costs. The city area variable can perhaps also capture other advantages related to centrality. In summary, Model 3 disclosed valuable insights about the underlying effects of location.

Surprisingly, the coefficient signs of primary school and secondary school were positive in both Model 2 and Model 3. A closer investigation was therefore needed, although the results were consistent with the findings of Dziauddin and Idris (2017). A sensitivity analysis was performed to strengthen the robustness of the results. By estimating Model 2 and Model 3 for subsets of data dividing the city into east and west, changes in the coefficient signs were detected. In Oslo East reduced distance to primary school now increased the apartment sales price, while reduced distance to secondary school increased the sales price in Oslo West. In essence, this indicates unstable coefficient signs, and suggests that variables related to school capture other effects in addition to distance. In this case it would be interesting to take a closer look at the characteristics of the apartment buyers. If households with children in primary school age are overrepresented in the eastern districts, and households with children in secondary school age are overrepresented in the western districts, this is an interesting finding about Oslo as a whole.

Another finding was that the coefficient signs of distance to theater and distance to retail were positive in both Model 2 and 3, which indicates surprising relationships. It can be argued that the effect of the two aforementioned hyper-locality variables is drawn out by some of the other variables. Theaters and retail stores often have a central location, and thus may be capturing centrality effects. This effect could be drawn out by the district variables in Model 2, and by the city area variable in Model 3. Moreover, library stood out as the hyper-locality variable with the highest relative effect in both models, which also can be due to the centrality aspect.

The majority of the hypotheses presented in Subsection 4.2.3, were supported by the estimated

SEM model. Increased distance to leisure was expected to lower the sales price per square meter, which was strongly confirmed by the LISREL estimation. Distance to leisure played out as the most important factor when valuing the location of a home. However, the latent factors distance to school and commerce got the opposite parameter signs than expected. Surprisingly, increased distance to schools rose the sales price of an apartment. Anyhow, this finding is in line with the results of the hedonic Model 2 and Model 3, and the study of Dziauddin and Idris (2017). Despite this, the east-west divided data subsets also here were run to investigate the robustness of the school factor in the SEM model. In contrast to the Oslo West SEM model, the model for Oslo East found that closeness to schools was appreciated, as the sales price per square meter increased in Oslo East. This can be due to an unstable parameter sign, or significant differences in demographic composition related to households. Furthermore, the latent commerce factor unfolded to have a relatively strong positive effect on the sales price per square meter, in the opposite way than expected. This finding indicates that an apartment located with some distance to commerce, is appreciated and reflected in a higher apartment sales price. Seen in conjunction with the variable quiet, which concludes that a quiet neighborhood gives a higher sales price, the result is plausible. This argumentation is further confirmed by the positive correlation between the distance to commerce factor and quiet, see Appendix C.4.

In the correlation matrix in Table 5.8, all of the correlation signs are as expected. The correlations between the latent location factors are relatively low, except for the correlation between distance to leisure and distance to commerce. This may be due to the fact that both of these latent variables can be associated with centrality, and the fact that the hyper-locality variables included in the two latent variables tend to locate near each other. The relatively low correlations between the other latent factors indicate that the data reduction of hyper-locality variables from nine to three factors in Subsection 5.2.2 was appropriate. In essence, the hyper-locality variables clustered as they were expected a priori. Overall, the correlations give useful information about the relationships between the latent location factors and their relation to the apartment sales price, controlled for the effect of all other variables combined.

All of the models in the thesis, both hedonic models and the SEM model, have high and approximately equal explanatory powers. In other words, the dependent variable sales price per square meter has high and similar explained variance in both approaches, which is desirable. Nevertheless, Model 2 with hyper-locality variables had the highest explanatory power and methodically outperformed the other models. Furthermore, the residual analysis for the hedonic models showed higher VIF scores for the hyper-locality variables related to leisure compared to the other variables, see Appendix C.2. By conducting an EFA and clustering the hyper-locality variables, the SEM model mitigated potential multicollinearity. Anyhow, as discussed in Subsection 4.1.4, multicollinearity is not a severe problem in the estimated hedonic models due to the large sample size. However, reducing problems with multicollinearity in general, in addition to uncovering latent location factors, makes SEM a useful modeling tool for examining location in depth.

The findings reported in this thesis must be seen in the light of limitations of the models. A trade-off between the time-consuming process of reviewing observations in depth versus managing time properly was encountered. In terms of cleaning outliers, the process was based on the suggestions of Pollestad and Helgaker (2021), as their process was deemed sensible and realistic. Even after a thorough cleaning process, the final dataset was still deemed sufficiently large for estimating models. Further, limitations due to selection of variables and the model's functional form should also be considered. It is recommended to rely as much as possible on theory rather than statistical fit when choosing variables (Studenmund, 2017). In the thesis, variables and factors were chosen on the basis of earlier theory, in addition to own assessments and possibilities due to the dataset. Lastly, the log-linear functional form of the models in the analyses was chosen on the basis of suggestions in earlier theory, as presented in Chapter 4, and seems to be fitting. As mentioned, the coefficients of determination in both the hedonic models and the SEM model were high. For the SEM model, the GoF indices indicated good model fit. GoF indices may be affected by small-sample bias, model mis-specification, violation of normality assumptions, and the estimation method. This was taken into account by maintaining a large sample size, as well as using robust model estimation that corrected for non-normality. The GoF indices, in combination with reliability and validity assessments, indicated that the SEM model had a good fit and was correctly specified.

In general, earlier literature within the field of property value estimation contributes with context-specific studies where generalizability is limited. Here, the socio-economic index is made continuous. The index is thereby generalizable to other cities, and additionally makes it easier to distinguish between the various components that location consists of. The components included in the socio-economic index were carefully considered based on the case of Oslo, and the thought of generalizability to other cities. Oslo is characterized by a high immigration rate, similar to other major cities, and thus the share of immigrants with long residence were sensible to include as an additional component compared to the study of Heyman and Sommervoll (2019). Despite this, a more detailed socio-economic index may be compiled to capture an even more accurate socio-economic ranking of the city districts. Additionally, studies regarding other cities have the opportunity to replace or supplement the index with different, more relevant socio-economic components. Further, the proxy for traveling costs also contributes to more generalizability as the vast majority of cities consist of an inner core surrounded by outer districts, similar to the city of Oslo. On that basis, results of hedonic Model 3 and the SEM model are generalizable to a greater extent, compared to earlier literature.

Chapter 7

Conclusion and Further Work

Heterogenity in sales price for apartments fundamentally depend on location, which is a broad concept and not clearly observable. Thus, looking at several aspects of location is relevant. A deeper dive into hyper-locality features and socio-economic considerations gives a broader understanding of the concept of location, and highlight aspects of location valued by apartment buyers. This thesis sought to examine the impact of these aspects in the real estate market of Oslo, in the time period 2017 to 2021.

The study was conducted by creating three different hedonic models, before a SEM model was estimated based on the socio-economic hedonic model, which included a variable capturing travel costs and a self-compiled socio-economic index. In advance of the structural modeling, an EFA was run to identify potential latent location factors. The latent location factors in the SEM model were assembled to identify which type of hyper-locality features that had the greatest impact on apartment sales prices. Additionally, the predictive performance of each of the included models in the thesis was examined.

As expected by previous literature, the thesis finds that there are differences within the Oslo districts in apartment prices, depending on which hyper-locality features that are nearby. Distance to library stood out as the hyper-locality feature with the highest relative importance on the sales price per square meter. The EFA disclosed three location factors: distance to school, leisure, and commerce. The SEM analysis further concluded that distance to leisure activities is the most important location factor. The latent factor consisted of the distance-measured hyper-locality variables library, museum, theater, and gym. Further, the sales price of apartments was found negatively correlated with all three latent location factors, controlled for the effect of all other variables combined. Additionally, it is found that the socio-economic status of the district plays a significant role for people's willingness to pay when purchasing an apartment in Oslo. Moreover, apartments located in the outer parts of Oslo are found to have a significantly lower sales price per square meter than apartments in the inner city.

The hedonic hyper-locality model had the highest explanatory power and outperformed the other models methodically. This indicates that hyper-locality features give additional explanation of variations in the sales price of apartments. By additionally analyzing location variables which cannot be directly observed, the thesis contributes with a more complete picture of what location entails. The estimations of the hedonic socio-economic model and the SEM model made this possible. Through the inclusion of a socio-economic index for measuring status, and an inner-outer division for capturing travel costs, underlying effects of location were assessed. In addition, the SEM analysis uncovered latent location factors, and provided the correlation between them and the apartment sales price, which makes SEM a useful modeling tool for examining location in depth.

While location variables are often included in property valuation models, the inclusion of hyper-locality features in such models has generally been modest. This thesis examines more hyper-locality features compared to earlier studies. Additionally, previous literature within the field of property value estimation has contributed with context-specific studies where generalizability is limited. By substituting the district variable with a socio-economic index for measuring status and an inner-outer division for capturing travel costs, its easier to generalize the results to other cities. Further, the findings of the underlying effects of location will be of great importance with regards to urban issues, planning, and site selection, especially from a real estate developer's perspective.

Today, the presence of well-developed Business Intelligence systems enables greater data access and lower extraction costs. The dataset used in this thesis was extensive and contained detailed location variables. Nevertheless, some variables that were considered relevant were absent. In example, distance to restaurants was not included in the dataset. In addition, there was no information on distances to bus and tram, which are two of the most commonly used means of transport in Oslo. By including this, the latent factor distance to commerce might be strengthened. A suggestion for further research is to take these variables into account when estimating apartment prices. Another suggestion for further research is to study the quadratic polynomial of the hyper-locality variables, based on the idea that people want to live close to certain amenities, but not too close. Ultimately, it would be interesting for future studies to look more closely at the characteristics of the buyers. A further differentiation of household types could give valuable information, as the need for proximity to various hyper-locality features changes over a life span and is influenced by different life-cycle events. This would also further explain the differences in the socio-economic status of the districts on a deeper level. Using SEM methodology for discovering latent socio-economic status factors can be of relevance.

Bibliography

Bagozzi, R. P. and Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science*, 16(1):74-94. Retrieved 2022-04-28 from https://doi.org/10.1007/BF02723327.

Balchin, P. N., Bull, G. H., and Kieve, J. L. (1995). *Urban land economics and public policy.* New York: Palgrave Macmillan.

Bland, J. M. and Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj*, 314(7080):572. Retrieved 2022-04-28 from https://doi.org/10.1136/bmj.314.7080.572.

Bourassa, S. C., Hoesli, M., and Peng, V. S. (2003). Do housing submarkets really matter? *Journal of Housing Economics*, 12(1):12-28. Retrieved 2022-05-02 from https://doi.org/10.1016/S1051-1377(03)00003-2.

Bowen, N. K. and Guo, S. (2012). *Structural Equation Modeling.* Oxford: Oxford University Press.

Bydelsfakta (2022a). Boligpriser: Boligpris for blokkleiligheter. Retrieved 2022-03-11 from https://bydelsfakta.oslo.kommune.no/bydel/alle/boligpriser.

Bydelsfakta (2022b). Bydelsfakta - statistikk om befolkning, levekår og boforhold. Retrieved 2022-03-10 from https://bydelsfakta.oslo.kommune.no.

Bydelsfakta (2022c). Bygningstyper: Boliger etter bygningstype i Oslo. Retrieved 2022-03-17 from https://bydelsfakta.oslo.kommune.no/bydel/01-02-03-04-05-06-07-08-09-10-11-12-13-14-15/bygningstyper.

Bydelsfakta (2022d). Levekårsindikatorer. Retrieved 2022-03-11 from https://bydelsfakta.oslo.kommune.no/bydel/alle/levekaar.

Cheng, W. (2020). Location Selection Method of Real Estate Developers Based on Big Data Technology. In *2020 International Conference on Robots Intelligent System (ICRIS)*, pages 426-429. Retrieved 2022-01-22 from https://doi.org/10.1109/ICRIS52159.2020.00110.

De Bruyne, K. and Van Hove, J. (2013). Explaining the spatial variation in housing prices: an economic geography approach. *Applied Economics*, 45(13):1673-1689. Retrieved 2022-03-21 from https://doi.org/10.1080/00036846.2011.636021.

Dziauddin, M. F. and Idris, Z. (2017). Use of Geographically Weighted Regression (GWR) Method to Estimate the Effects of Location Attributes on the Residential Property Values. *The Indonesian Journal of Geography*, 49(1):97–110. Retrieved 2022-03-03 from https://www.proquest.com/scholarly-journals/use-geographically-weighted-regression-gwr-method/docview/1941338962/se-2?accountid=12870.

Eurostat (2021). Owning or renting? What is the EU's housing situation? Retrieved 2022-03-09 from https://ec.europa.eu/eurostat/web/products-eurostat-news/-/wdn-20211230-1.

Fallan, L., Hammervold, R., and Grønhaug, K. (1995). Adoption of tax planning instruments in business organizations: A structural equation modelling approach. *Scandinavian Journal of Management*, 11(2):177-190. Retrieved 2022-05-10 from https://doi.org/10.1016/0956-5221(95)00007-I.

Finn.no (2022). Eiendom / Bolig til salgs. Retrieved 2022-03-20 from https://www.finn.no/realestate/homes/search.html?sort=PUBLISHED_DESC.

Forbrukertilsynet (2022). Prisopplysninger: Prisantydning: Forbrukertilsynets veiledning for markedsføring av bolig. Retrieved 2022-03-10 from https://www.forbrukertilsynet.no/lov-og-rett/veiledninger-og-retningslinjer/forbrukertilsynets-veiledning-om-markedsforing-av-bolig.

Fornell, C. and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 18(1):39-50. Retrieved 2022-04-22 from https://doi.org/10.1177/002224378101800104.

Hammervold, R. (2020). *Multivariate analyser med Stata - En kort innføring*. Bergen: Fagbokforlaget.

Heyman, A. V. and Sommervoll, D. E. (2019). House prices and relative location. *Cities*, 95:1. Retrieved 2022-02-15 from https://doi.org/10.1016/j.cities.2019.06.004.

Hui, E. C., Chau, C. K., Pun, L., and Law, M. (2007). Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix. *Building and environment*, 42(6):2333-2343. Retrieved 2022-05-02 from https://doi.org/10.1016/j.buildenv.2006.05.004.

Huseierne (2022). Klager Finn.no inn for Konkurransetilsynet. Retrieved 2022-03-10 from https://www.huseierne.no/nyheter/klager-finn.no-inn-for-konkurransetilsynet/.

Iversen, K. O. (2020). Hva betyr primærrom og bruttoareal? Retrieved 2022-03-16 from https://dnbeiendom.no/altombolig/kjop-og-salg/tips-til-kjopere/hva-inngar-i-boligens-areal.

Javid, N., Khalili-Damghani, K., Makui, A., and Abdi, F. (2019). Multi-dimensional flexibility-complexity trade-off modeling in manufacturing systems: Structural equation modeling approach. *Kybernetes*, 48(8):1757–1781. Retrieved 2022-03-21 from https://doi.org/10.1108/K-05-2018-0259.

Jordaan, A. C., Drost, B., and Makgata, M. A. (2004). Land value as a function of distance from the CBD: The case of the eastern suburbs of Pretoria. *South African journal of economic and management sciences*, 7(3):532-541. Retrieved 2022-05-02 from https://hdl.handle.net/10520/EJC31446.

Jöreskog, K. G., Olsson, U. H., and Wallentin, F. Y. (2016). *Multivariate analysis with LISREL*. Switzerland: Springer International Publishing.

Kauko, T. (2003). Residential property value and locational externalities: On the complementarity and substitutability of approaches. *Journal of Property Investment & Finance*. Retrieved 2022-05-02 from https://doi.org/10.1108/14635780310481676.

Krogsveen (2022). Boligprisstatistikk: Prisutvikling for Oslo. Retrieved 2022-03-10 from https://www.krogsveen.no/prisstatistikk/oslo.

Kryvobokov, M. and Wilhelmsson, M. (2007). Analysing location attributes with a hedonic model for apartment prices in Donetsk, Ukraine. *International journal of strategic property management*, 11:157-178. Retrieved 2022-05-14 from https://doi.org/10.3846/1648715X.2007.9637567.

Lindquist, K.-G., Mæhlum, S., and Helge, B. (2021). Boligmarkedet i pandemiåret 2020. *Norges Bank*. Retrieved 2022-02-15 from http://hdl.handle.net/10419/246161.

Liu, Y. and Wu, Y.-x. (2009). Analysis of residential product's value based on Structural Equation Model and Hedonic Price Theory. In *2009 International Conference on Management Science and Engineering*, pages 1950-1956. Retrieved 2022-01-22 from https://doi.org/10.1109/ICMSE.2009.5317673.

Malpezzi, S. et al. (2003). Hedonic pricing models: a selective and applied review. *Housing economics and public policy*, 1:67–89. Retrieved 2022-02-26 from https://www.researchgate.net/publication/285369028_Hedonic_Pricing_Models_A_Selective_and_Applied_Review.

McLeod, P. B. (1984). The demand for local amenity: an hedonic price analysis. *Environment and Planning A*, 16(3):389–400. Retrieved 2022-10-02 from https://doi.org/10.1068/a160389.

Mesthrige, J. W. and Ka, C. W. (2017). Do Buyers Value a Balcony as a Green Feature? An Empirical Analysis of the Hong Kong Residential Property Market. *Journal of Real Estate Practice and Education*, 20(1):27-49. Retrieved 2022-05-02 from https:

//www.proquest.com/trade-journals/do-buyers-value-balcony-as-green-feature/docview/1933850502/se-2?accountid=12870.

National Geographic (2012). Location. Retrieved 2022-02-16 from https://www.nationalgeographic.org/encyclopedia/location/.

Nesset, I. Q. and Oust, A. (2020). The impact of historic preservation policies on housing values. *International Journal of Housing Policy*, 20(2):252–272. Retrieved 2022-03-19 from https://doi.org/10.1080/19491247.2019.1688633.

Oslo Byleksikon (2022). Bydeler. Retrieved 2022-03-19 from https://www.oslobyleksikon.no/side/Bydeler.

Oslo Kommune (2018). Boligpriser. Retrieved 2022-03-10 from https://www.oslo.kommune.no/statistikk/boliger-byggevirksomhet-arbeids-og-naringsliv/boligpriser/#gref.

Oslo Kommune (2021). Befolkningens landbakgrunn. Retrieved 2022-05-19 https://www.oslo.kommune.no/statistikk/befolkning/landbakgrunn/#gref.

Oslo Kommune (2022). Folkemengde og endringer. Retrieved 2022-03-09 from https://www.oslo.kommune.no/statistikk/befolkning/folkemengde-og-endringer/#gref.

Oust, A. (2012). House price indices Oslo 1970-2011. Retrieved 2022-03-18 from https://doi.org/10.13140/RG.2.2.12212.68482.

Oust, A., Hansen, S. N., and Pettrem, T. R. (2020). Combining property price predictions from repeat sales and spatially enhanced hedonic regressions. *The Journal of Real Estate Finance and Economics*, 61(2):183-207. Retrieved 2022-01-25 from https://doi.org/10.1007/s11146-019-09723-x.

Palmquist, R. B. (1984). Estimating the Demand for the Characteristics of Housing. *The Review of Economics and Statistics*, pages 394–404. Retrieved 2022-03-23 from https://doi.org/10.2307/1924995.

Pollestad, A. J. and Helgaker, E. (2021). Adverse selection in iBuyer business models: A study of adverse selection in the use of automated valuation models for iBuyers. Retrieved 2022-01-19 from https://hdl.handle.net/11250/2986105.

Rehák, Š. and Síbert, A. (2017). Public transportation and apartment prices: a hedonic model of Bratislava. In *Proceedings from 6th Central European Conference in Regional ScienceCERS, 2017. Banská Bystrica.*

Ringdal, K. (2001). *Enhet og mangfold: samfunnsvitenskapelig forskning og kvantitativ metode.* Bergen: Fagbokforlaget.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34-55. Retrieved 2022-03-23 from https://doi.org/10.1086/260169.

Satorra, A. and Bentler, P. (1988). Scaling corrections for statistics in covariance structure analysis.

Schermelleh-Engel, K., Moosbrugger, H., Müller, H., et al. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2):23-74.

Schirmer, P. M., Van Eggermond, M. A., and Axhausen, K. W. (2014). The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, 7(2):3–21. Retrieved 2022-03-02 from https://www.researchgate.net/publication/272643866_The_role_of_location_in_residential_location_choice_models_A_review_of_literature.

Segal, D. (1979). *The economics of neighborhood*, pages 193–217. New York: Academic Press.

Sirmans, S., Macpherson, D., and Zietz, E. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, 13(1):1-44. Retrieved 2022-05-02 from https://doi.org/10.1080/10835547.2005.12090154.

SSB (2022). Fakta om bolig. Retrieved 2022-03-09 from https://www.ssb.no/bygg-bolig-og-eiendom/faktaside/bolig.

Statistics Norway (2022). House prices and house price indices: Price index for existing dwellings. Retrieved 2022-05-05 from https://www.ssb.no/en/priser-og-prisindekser/boligpriser-og-boligprisindekser/statistikk/prisindeks-for-brukte-boliger.

Studenmund, A. H. (2017). *A Practical Guide to Using Econometrics*. Essex: Pearson Education Limited.

Titman, S. and Wessels, R. (1988). The determinants of capital structure choice. *The Journal of finance*, 43(1):1–19. Retrieved 2022-03-21 from https://doi.org/10.1111/j.1540-6261.1988.tb02585.x.

Twycross, A. and Shields, L. (2004). Validity and reliability-What's it all about? Part 2 Reliability in quantitative. *Paediatric Nursing*, 16(10):36. Retrieved 2022-04-15 from https://www.researchgate.net/publication/305253817_Validity_and_reliability_-_What's_it_all_about_Part_2_Reliability_in_quantitative_studies.

UCLA Statistical Consulting Group (2021). What Does Cronbach's Alpha Mean? Retrieved 2022-04-28 from https://stats.oarc.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean.

Venerandi, A., Fusco, G., Tettamanzi, A., and Emsellem, D. (2019). A machine learning approach to study the relationship between features of the urban environment and street value. *Urban Science*, 3(3):100. Retrieved 2022-02-15 from https://doi.org/10.3390/urbansci3030100.

Öye, B.-E. (2019). Sentralt i Oslo øker mest. *Prognosesenteret*. Retrieved 2022-02-15 from https://blogg.prognosesenteret.no/sentralt-i-oslo-oeker-mest.

Appendix A

Data Analysis

A.1 Socio-Economic Variable

Table A.1: The four socio-economic components retrieved from Bydelsfakta (2022b), followed by the calculated socio-economic index S at district level.

	Unemployment (Em)	**Low level of education (Ed)**	**Low level of income (I)**	**Share of immigrants (Im)**	**Socio-economic score (S)**
Gamle Oslo	63	40	93	56	37
Grünerløkka	42	27	66	42	56
Sagene	34	21	49	24	68
St. Hanshaugen	37	16	45	22	70
Frogner	55	15	37	24	67
Ullern	8	2	1	10	95
Vestre Aker	7	0	0	4	97
Nordre Aker	0	1	6	0	98
Bjerke	49	53	66	66	42
Grorud	80	83	74	91	18
Stovner	100	100	98	100	1
Alna	84	80	74	96	16
Østensjø	16	31	22	27	76
Nordstrand	6	14	7	6	92
Søndre Nordstrand	90	79	89	96	11

Appendix B

Methodology

B.1 Price Index for Existing Dwellings, 2017-2021

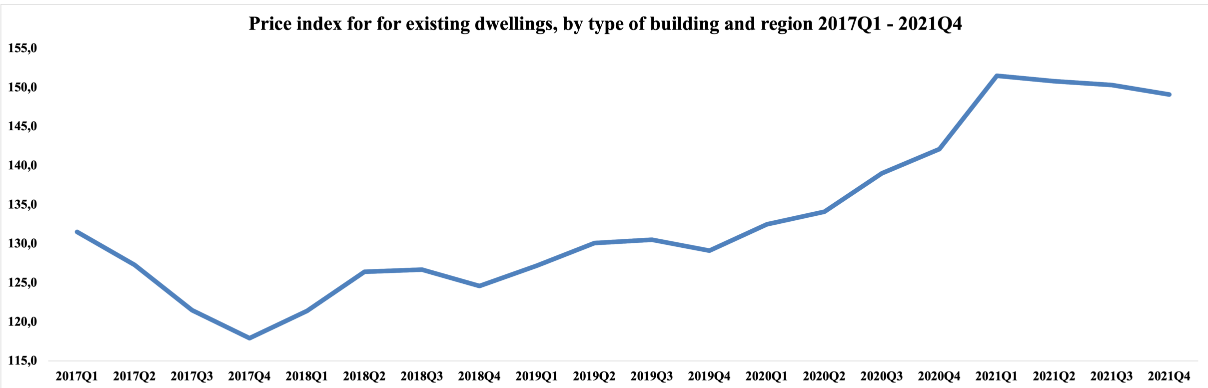


Figure B.1: Price index for existing dwellings, multi-dwelling, Oslo including Bærum, 2017Q1-2021Q4. *"The statistics measure the value of the stock of existing dwellings, based on current price information of existing dwellings sold on the free market. The figures are seasonally adjusted."* (Statistics Norway, 2022).

Appendix C

Results

C.1 Hedonic Regressions

C.1.1 Model 1

Table C.1: Full model estimation for Model 1, with unstandardized regression coefficients, robust standard errors $SE(\hat{\alpha})$, t-values, p-values, and standardized regression coefficients, beta.

Sales price per square meter (ln)	Coefficient	Robust standard error	t	P > | t |	Beta
District					
Grünerløkka	0.036	0.002	17.390	0.000	0.046
Sagene	0.086	0.002	40.770	0.000	0.101
St. Hanshaugen	0.153	0.003	60.230	0.000	0.141
Frogner	0.247	0.003	95.890	0.000	0.260
Ullern	0.106	0.004	27.980	0.000	0.071
Vestre Aker	0.009	0.004	2.360	0.018	0.006
Nordre Aker	0.051	0.004	14.460	0.000	0.038
Bjerke	-0.174	0.003	-52.210	0.000	-0.134
Grorud	-0.380	0.004	-103.620	0.000	-0.258
Stovner	-0.445	0.004	-108.190	0.000	-0.241
Alna	-0.320	0.003	-114.980	0.000	-0.301
Østensjø	-0.209	0.003	-81.780	0.000	-0.198
Nordstrand	-0.132	0.004	-36.950	0.000	-0.100
Søndre Nordstrand	-0.487	0.005	-108.010	0.000	-0.233
Size (ln)	-0.292	0.002	-130.290	0.000	-0.375
Sales year					
2018	0.002	0.002	1.040	0.298	0.003
2019	0.037	0.002	20.400	0.000	0.054
2020	0.095	0.002	52.200	0.000	0.139
2021	0.192	0.002	103.350	0.000	0.266
Build year > 2000	0.130	0.002	75.380	0.000	0.195
Quiet	0.018	0.001	12.680	0.000	0.026
View	0.035	0.001	29.720	0.000	0.063
Balcony	0.022	0.002	11.450	0.000	0.026
Fireplace	0.052	0.002	34.270	0.000	0.083
Elevator	0.006	0.001	4.310	0.000	0.011
_cons	12.266	0.009	1356.280	0.000	.

C.1.2 Model 2

Table C.2: Full model estimation for Model 2, with unstandardized regression coefficients, robust standard errors $SE(\hat{\alpha})$, t-values, p-values, and standardized regression coefficients, beta.

Sales price per square meter (ln)	Coefficient	Robust standard error	t	P > | t |	Beta
District					
Grünerløkka	0.041	0.002	18.000	0.000	0.052
Sagene	0.098	0.003	38.430	0.000	0.115
St. Hanshaugen	0.064	0.003	19.320	0.000	0.059
Frogner	0.158	0.003	50.260	0.000	0.167
Ullern	0.141	0.004	35.680	0.000	0.094
Vestre Aker	0.079	0.004	17.810	0.000	0.052
Nordre Aker	0.087	0.004	24.460	0.000	0.064
Bjerke	-0.075	0.004	-19.480	0.000	-0.058
Grorud	-0.218	0.005	-42.670	0.000	-0.148
Stovner	-0.279	0.005	-51.980	0.000	-0.151
Alna	-0.190	0.004	-49.350	0.000	-0.178
Østensjø	-0.103	0.004	-27.680	0.000	-0.097
Nordstrand	-0.028	0.005	-5.850	0.000	-0.022
Søndre Nordstrand	-0.363	0.006	-65.390	0.000	-0.174
Distance primary school (ln)	0.012	0.001	11.260	0.000	0.028
Distance secondary school (ln)	0.010	0.001	8.710	0.000	0.025
Distance library (ln)	-0.093	0.002	-46.640	0.000	-0.305
Distance museum (ln)	0.001	0.001	0.480	0.632	0.002
Distance theater (ln)	0.010	0.001	9.640	0.000	0.038
Distance gym (ln)	-0.013	0.001	-9.250	0.000	-0.047
Distance retail (ln)	0.011	0.001	12.760	0.000	0.029
Distance subway and train (ln)	-0.012	0.001	-14.080	0.000	-0.035
Distance industry (ln)	0.011	0.001	12.980	0.000	0.033
Size (ln)	-0.290	0.002	-134.340	0.000	-0.373
Sales year					
2018	0.003	0.002	1.840	0.066	0.005
2019	0.039	0.002	22.200	0.000	0.056
2020	0.096	0.002	54.500	0.000	0.139
2021	0.194	0.002	108.500	0.000	0.269
Build year > 2000	0.134	0.002	79.080	0.000	0.201
Quiet	0.023	0.001	16.490	0.000	0.033
View	0.036	0.001	31.440	0.000	0.064
Balcony	0.030	0.002	15.930	0.000	0.036
Fireplace	0.045	0.001	30.180	0.000	0.071
Elevator	0.006	0.001	4.240	0.000	0.010
_cons	12.373	0.009	1347.520	0.000	.

C.1.3 Model 3

Table C.3: Full model estimation for Model 3, with unstandardized regression coefficients, robust standard errors $SE(\hat{\alpha})$, t-values, p-values, and standardized regression coefficients, beta.

Sales price per square meter (ln)	Coefficient	Robust standard error	t	P > | t |	Beta
Distance primary school (ln)	0.006	0.001	5.560	0.000	0.014
Distance secondary school (ln)	0.020	0.001	19.270	0.000	0.051
Distance library (ln)	-0.102	0.001	-69.180	0.000	-0.334
Distance museum (ln)	-0.023	0.001	-22.620	0.000	-0.086
Distance theater (ln)	0.019	0.001	21.970	0.000	0.070
Distance gym (ln)	-0.019	0.001	-14.340	0.000	-0.067
Distance retail (ln)	0.012	0.001	12.960	0.000	0.030
Distance subway and train (ln)	-0.012	0.001	-15.190	0.000	-0.037
Distance industry (ln)	0.019	0.001	22.390	0.000	0.055
Socio-economic index	0.003	0.000	107.740	0.000	0.284
Outer city	-0.113	0.003	-41.600	0.000	-0.201
Size (ln)	-0.284	0.002	-129.620	0.000	-0.365
Sales year					
2018	0.000	0.002	0.260	0.793	0.001
2019	0.037	0.002	20.150	0.000	0.054
2020	0.093	0.002	50.720	0.000	0.136
2021	0.192	0.002	103.110	0.000	0.266
Build year > 2000	0.133	0.002	76.690	0.000	0.200
Quiet	0.026	0.001	17.690	0.000	0.037
View	0.038	0.001	31.880	0.000	0.067
Balcony	0.030	0.002	15.650	0.000	0.036
Fireplace	0.049	0.002	32.340	0.000	0.077
Elevator	0.014	0.001	9.620	0.000	0.024
_cons	12.250	0.010	1287.860	0.000	.

C.2 Residual Analysis for Hedonic Models

C.2.1 Variance Inflation Factors

Table C.4: VIF scores for the independent variables in the three hedonic models.

	Model 1	Model 2	Model 3
Grünerløkka	1.88	2.41	
Sagene	1.76	2.53	
St. Hanshaugen	1.52	2.13	
Frogner	1.74	2.39	
Ullern	1.27	1.56	
Vestre Aker	1.29	1.89	
Nordre Aker	1.31	1.54	
Bjerke	1.36	2.33	
Grorud	1.32	2.84	
Stovner	1.22	1.98	
Alna	1.60	3.48	
Østensjø	1.58	3.29	
Nordstrand	1.39	2.52	
Søndre Nordstrand	1.16	1.73	
Distance primary school (ln)		1.52	1.37
Distance secondary school (ln)		1.94	1.60
Distance library (ln)		7.48	4.61
Distance museum (ln)		5.09	3.31
Distance theater (ln)		3.63	2.53
Distance gym (ln)		5.95	4.40
Distance retail (ln)		1.20	1.14
Distance subway and train (ln)		1.59	1.36
Distance industry (ln)		1.58	1.42
Size (ln)	1.27	1.29	1.25
2018	1.68	1.68	1.68
2019	1.69	1.70	1.69
2020	1.71	1.71	1.70
2021	1.64	1.64	1.64
Build year > 2000	1.49	1.54	1.44
Quiet	1.03	1.05	1.04
View	1.08	1.08	1.08
Balcony	1.27	1.28	1.27
Fireplace	1.35	1.37	1.34
Elevator	1.41	1.42	1.37
Socio-economic index			1.52
Outer city			5.00
Mean VIF	1.44	2.30	1.99

C.2.2 White's Test

White's test for H_0 =	homoscedasticity
against H_1 =	unrestricted heteroscedasticity

Model 1:	chi2(228) = 9,192.43	Prob >chi2 = 0.000
Model 2:	chi2(507) = 11,703.43	Prob >chi2 = 0.000
Model 3:	chi2(258) = 9,283.43	Prob >chi2 = 0.000

C.2.3 Standardized Residuals

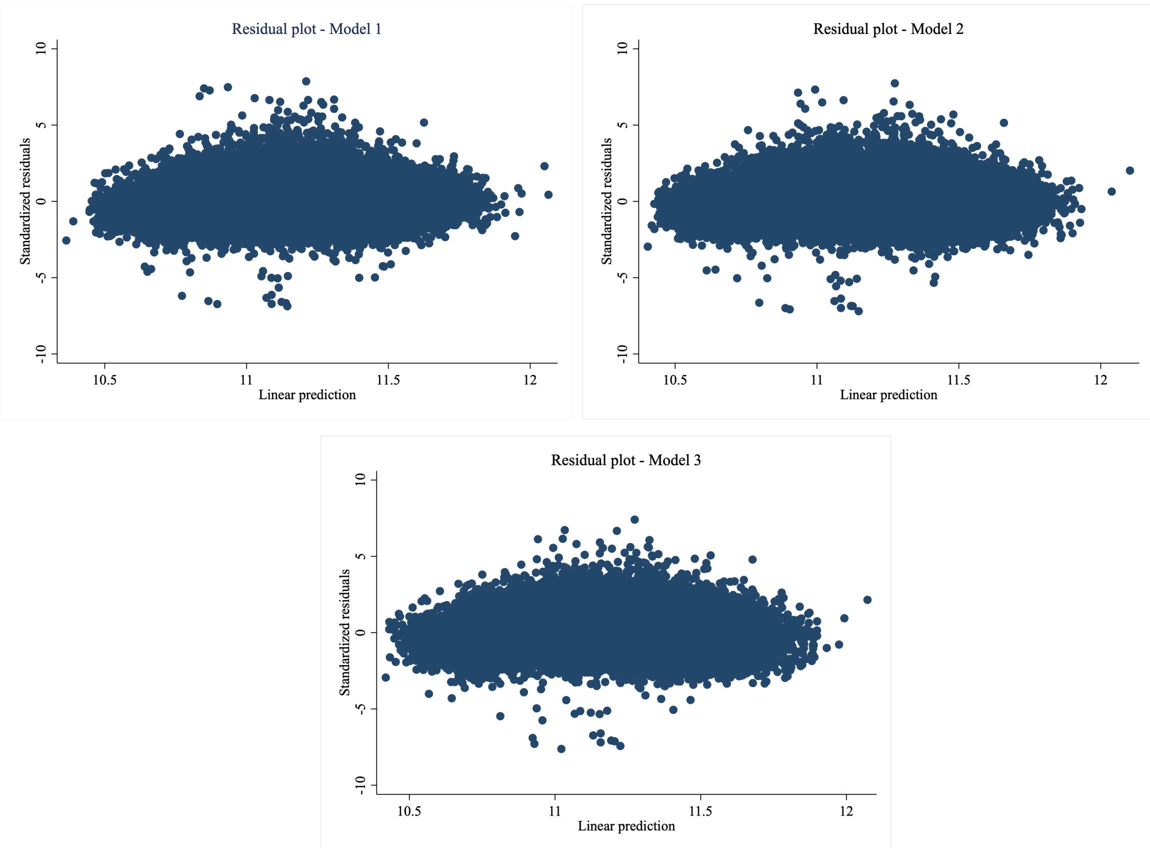
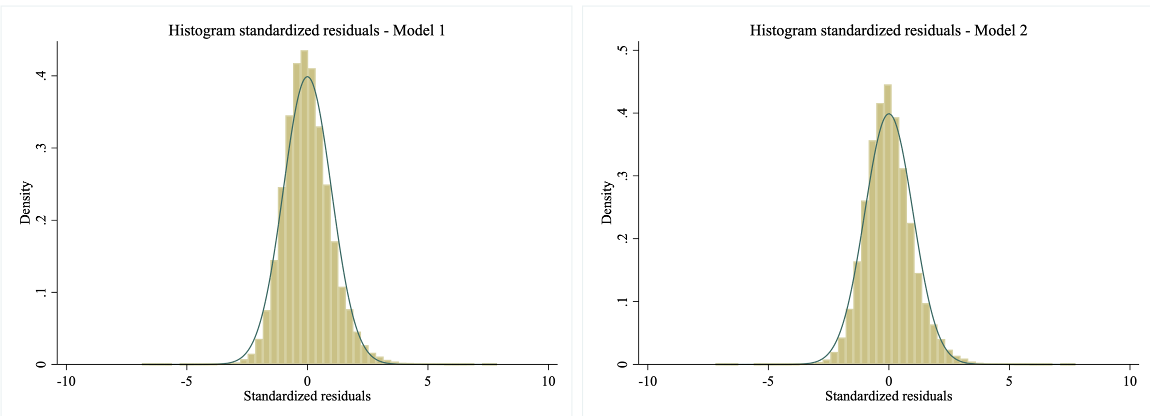


Figure C.1: Residual plot for hedonic models.



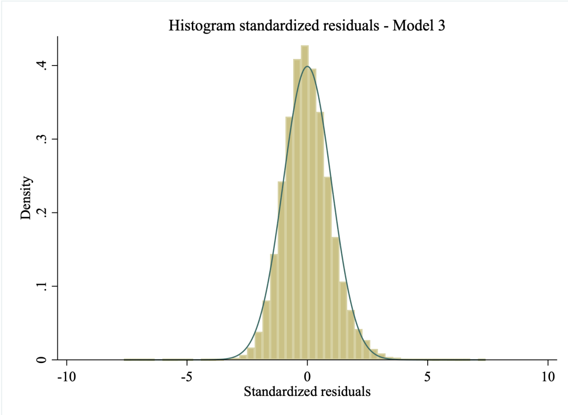


Figure C.2: Histogram for standardized residuals for hedonic models.

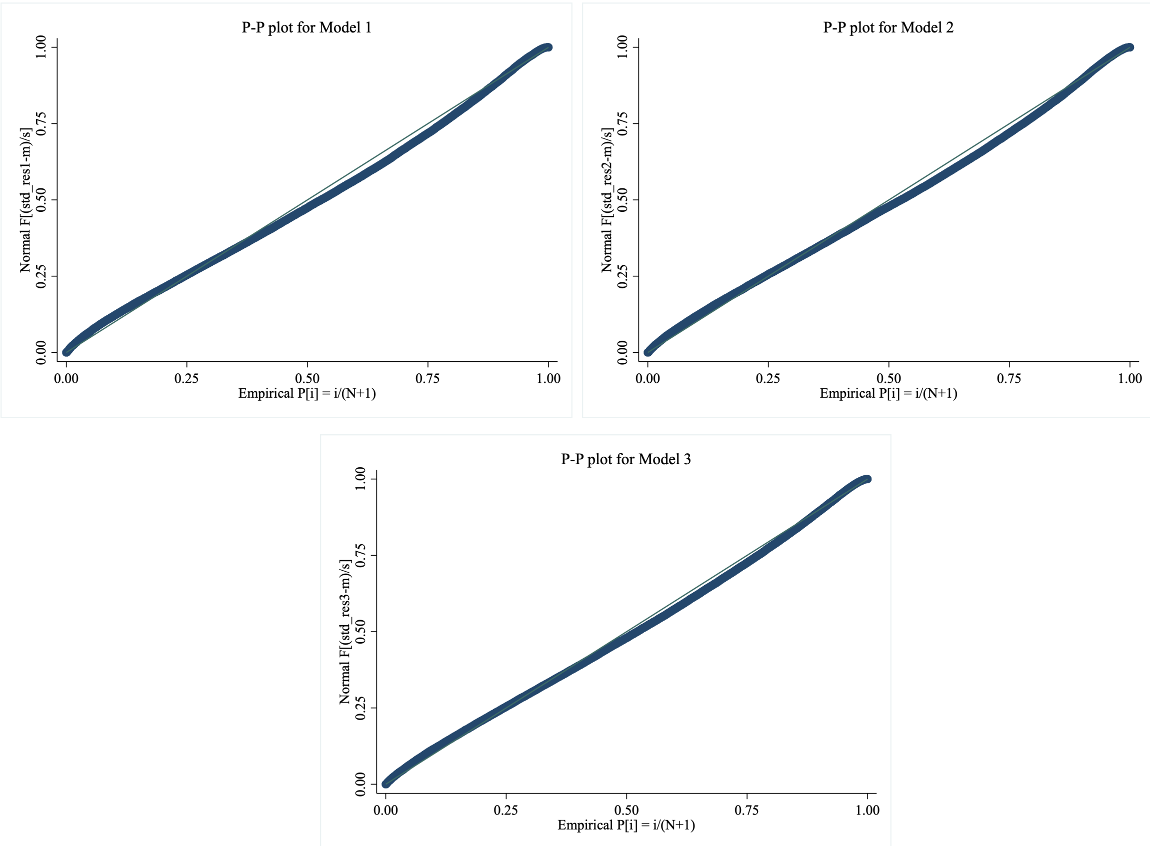


Figure C.3: P-P plot for hedonic models.

Table C.5: Skewness and kurtosis test for normality in standardized residuals, in addition to minimum and maximum values.

			Joint-test			
	Pr(skewness)	Pr(kurtosis)	chi2(2)	Prob>chi2	Minimum	Maximum
Standardized residual Model 1	0.000	5,314.20	0.000	0.000	-6.869	7.868
Standardized residual Model 2	0.000	4,249.13	0.000	0.000	-7.197	7.740
Standardized residual Model 3	0.000	3,145.47	0.000	0.000	-7.623	7.406

C.3 SEM Equations

C.3.1 Measurement Model

$X = \lambda \xi + \delta$

$Primary\ school\ (ln) = \lambda_{1,1}\xi_1 + \delta_1$

$Secondary\ school\ (ln) = \lambda_{2,1}\xi_1 + \delta_2$

$Library\ (ln) = \lambda_{3,2}\xi_2 + \delta_3$

$Museum\ (ln) = \lambda_{4,2}\xi_2 + \delta_4$

$Theater\ (ln) = \lambda_{5,2}\xi_2 + \delta_5$

$Gym\ (ln) = \lambda_{6,2}\xi_2 + \delta_6$

$Retail\ (ln) = \lambda_{7,3}\xi_3 + \delta_7$

$Subway\ and\ train\ (ln) = \lambda_{8,3}\xi_3 + \delta_8$

$Industry\ (ln) = \lambda_{9,3}\xi_3 + \delta_9$

C.3.2 Structural Model

$\eta = \Gamma \xi + \zeta$

$$Sales\ price\ per\ square\ meter\ (ln) = \gamma_{1,1}\xi_1 + \gamma_{1,2}\xi_2 + \gamma_{1,3}\xi_3 + \gamma_{1,4}\xi_4 + \gamma_{1,5}\xi_5 + \gamma_{1,6}\xi_6 + \gamma_{1,7}\xi_7$$
$$+ \gamma_{1,8}\xi_8 + \gamma_{1,9}\xi_9 + \gamma_{1,10}\xi_{10} + \gamma_{1,11}\xi_{11} + \gamma_{1,12}\xi_{12} + \gamma_{1,13}\xi_{13} + \gamma_{1,14}\xi_{14} + \gamma_{1,15}\xi_{15} + \gamma_{1,16}\xi_{16} + \zeta_1$$

C.4 Full Correlation Matrix for Latent and Observed Variables

Table C.6: Full correlation matrix for all latent and observed variables included in the SEM analysis.

	Sales price per square meter	Distance school	Distance leisure	Distance commerce	Status	Outer city	Size (ln)	Sales year 2018	Sales year 2019
Sales price per square meter	1.000								
Distance school	-0.046***	1.000							
Distance leisure	-0.195***	0.350***	1.000						
Distance commerce	-0.124***	0.292***	0.673***	1.000					
Status	0.112***	0.081***	-0.303***	0.111***	1.000				
Outer city	-0.159***	0.442***	0.863***	0.794***	-0.007*	1.000			
Size (ln)	-0.380***	0.081***	0.131***	0.352***	0.027***	0.175***	1.000		
Sales year 2018	-0.106***	-0.025***	-0.030***	-0.020***	0.004	-0.030***	0.000	1.000	
Sales year 2019	-0.053***	0.007	0.005	0.012**	-0.007*	0.005	0.014***	-0.266***	1.000
Sales year 2020	0.044***	0.017***	0.027***	0.022***	-0.007	0.030***	0.014***	-0.266***	-0.265***
Sales year 2021	0.213***	0.010**	0.021***	0.008	-0.012***	0.016***	-0.007	-0.242***	-0.247***
Build year >2000	0.282***	-0.029***	-0.131***	-0.255***	0.056	-0.178***	0.007**	-0.023***	0.018***
Quiet	-0.009**	0.030***	0.111***	0.166***	0.012***	0.086***	0.049***	-0.025***	-0.012***
View	-0.036***	0.070***	0.143***	0.122***	-0.080***	0.117***	0.095***	-0.020***	0.000
Balcony	-0.140***	0.154***	0.280***	0.255***	-0.042***	0.267***	0.253***	-0.089***	0.015***
Fireplace	0.065***	-0.036***	-0.156***	0.109***	0.217***	-0.068***	0.279***	0.022***	-0.006
Elevator	0.193***	-0.017***	-0.120***	-0.158***	-0.026***	-0.121***	-0.087***	-0.015***	0.010**

	Sales year 2020	Sales year 2021	Build year > 2000	Quiet	View	Balcony	Fireplace	Elevator
Sales year 2020	1.000							
Sales year 2021	-0.247***	1.000						
Build year > 2000	0.023***	0.023***	1.000					
Quiet	0.025***	0.039***	0.006	1.000				
View	0.025***	0.005*	0.053***	0.111***	1.000			
Balcony	0.096***	0.070***	0.172***	0.068***	0.194***	1.000		
Fireplace	-0.011***	-0.023***	-0.232***	0.040***	-0.077***	-0.082***	1.000	
Elevator	0.015***	0.011***	0.450***	-0.037***	0.104***	0.138***	-0.277***	1.000

Note: ***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

NTNU

Norwegian University of
Science and Technology