Aalto University
School of Engineering

The difference between the development value of residential land and its sales price in the Helsinki metropolitan area

Master's Thesis
Department of Built Environment
School of Engineering
Aalto University

Espoo, 27th May 2019

Pyry Haahtela

Supervisor: Professor Kauko Viitanen
Advisor: D.Sc. (Tech) Ari Pennanen

Aalto University
School of Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Pyry Haahtela

Title of thesis The difference between the development value of residential land and its sales price in the Helsinki metropolitan area

Master programme Degree Programme in Real Estate Economics **Code** ENG24

Thesis supervisor Professor Kauko Viitanen

Thesis advisor D.Sc. (Tech) Ari Pennanen

Date 27.5.2019 **Number of pages** 81 **Language** English

Abstract

The estimation of land values has been a topic in the real estate economics since the beginning of the field of study. Many governing theories of land value assume that the value of the location is capitalized in land value to the landowner and this should be reflected in the sales prices of land. In development projects, land value for the developer can be estimated using the residual valuation method, where the total development costs and the required development return are subtracted from the end-product value. The result is the highest price that a developer can pay for the land and stay profitable. The result of the residual valuation is often referred as the development value of the land.

The purpose of this study was to use the residual valuation method to find out if there is a difference in the development value and sales price of residential land in the Helsinki metropolitan area (HMA). There are many previous studies about residential land values in the HMA, but a comparison of residual values and lot sales prices regionally is a new approach where there were no previous studies from Finland.

The research data consisted of three independent datasets. Vacant lot sales data was provided by the National Land Survey of Finland, apartment price data was provided by Oikotie.fi and construction cost data was provided by Haahtela-kehitys Oy. The research would not have been possible without these key contributors. The apartment price data and construction cost data were used to calculate the residual land values that were compared to the lot sales prices.

The results of the study show that there are significant differences in the residual land values and lot sales prices in the HMA. Residual land values have significantly higher differentiation between areas than lot sales prices. The residual land values ranged from 143 to 3,180 €/net apartment area while the lot sales prices ranged from 297 to 2,198 €/net apartment area. The larger differentiation of residual land values between the areas shows that the development value of residential land is not entirely represented in the land sales prices. This leads to a situation where it is possible for the developer to achieve higher than the minimum required development return by concentrating development on the areas that have the highest difference between the development value and sales price of land. The expected development returns correlate heavily with apartment price levels. Based on the results the highest expected development returns with the current land price levels can be achieved from inner city areas with the highest apartment prices in the HMA.

Keywords Residual value, Land value, Land price, Development profitability, Housing market, Residential development, , Investment value

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Diplomityön tiivistelmä

Tekijä Pyry Haahtela

Työn nimi Asuntotonttien kehitysarvojen ja myyntihintojen erot Helsingin seudulla

Maisteriohjelma Kiinteistötalous **Koodi** ENG24

Työn valvoja Professori Kauko Viitanen

Työn ohjaaja TkT Ari Pennanen

Päivämäärä 27.5.2019 **Sivumäärä** 81 **Kieli** Englanti

Tiivistelmä

Maan arvon määrittäminen on yksi kiinteistötalouden tutkituimmista aiheista ja sitä on tutkittu monin tavoin vuosien saatossa. Useat maan arvoon liittyvät teoriat lähtevät siitä, että kiinteistön sijainnin merkitys näkyy ensisijaisesti maan arvossa ja maan arvo on perustana tonttien myyntihinnoille. Kiinteistökehittäjän näkökulmasta kehityskelpoisen tontin arvo voidaan määrittää arvioimalla kehityksen lopputuotteen arvo ja vähentämällä tästä rakennuskustannukset ja kiinteistökehittäjän kate. Tätä menetelmää kutsutaan varsinkin englanninkielisessä kirjallisuudessa nimellä ”Residual valuation method”, eli suoraan suomennettuna residuaalimenetelmäksi. Residuaaliarvo on korkein hinta, jonka kiinteistökehittäjä voi maksaa tontista niin, että kehitysprojekti pysyy riittävän kannattavana. Residuaaliarvoa voidaan kutsua myös kehitysarvoksi ”development value”.

Tämän tutkimuksen tarkoituksena on residuaalimenetelmää käyttämällä arvioida poikkeavatko tonttien myyntihinnat niiden kehitysarvoista Helsingin seudulla. Asuntotonttien arvoa on tutkittu Helsingin seudulla laajasti erilaisilla menetelmillä, mutta residuaaliarvojen ja tonttien myyntihintojen vertailusta ei löydy aiempaa Suomessa tehtyä tutkimusta.

Tutkimusaineisto koostui kolmesta erillisestä tietolähteestä. Tonttien myyntihinta-aineisto perustuu Maanmittauslaitoksen keräämiin tietoihin, asuntokauppa-aineisto Oikotie.fi palvelun toimittamiin tietoihin ja rakennuskustannusaineisto Haahtela-kehitys Oy:n toimittamiin tietoihin. Tutkimuksen tekeminen ei olisi ollut mahdollista ilman näiden yhteistyökumppanien toimittamaa aineistoa. Asuntojen hinta-aineistoa ja rakennuskustannusaineistoa käytettiin residuaaliarvojen laskentaan, joita verrattiin tutkimuksessa tonttien myyntihintoihin.

Tutkimustulokset osoittavat, että residuaaliarvot ja tonttien myyntihinnat poikkeavat merkittävästi toisistaan tutkimusalueilla. Residuaaliarvoissa on huomattavasti enemmän vaihtelua eri kaupunginosien välillä kuin tonttien myyntihinnoissa. Residuaaliarvot vaihtelivat välillä 143 – 3180 € / huoneiston nettoala, kun tonttien myyntihintojen vaihteluväli oli 297–2198 € / huoneiston nettoala. Residuaaliarvojen suurempi vaihtelu osoittaa, että asuntotonttien kehitysarvot eivät näy täysimääräisesti tonttien hinnoissa. Tämä johtaa tilanteeseen, jossa kiinteistökehittäjän on mahdollista saada minimivaatimuksia suurempia tuottoja kehityskohteista, jos kehitysalueet valikoidaan sen perusteella, missä tonttien myyntihinnat ja kehitysarvot poikkeavat eniten toisistaan. Laskennallisesti arvioidut kiinteistökehityksen tuotot korreloivat vahvasti asuntojen hintatason kanssa. Tutkimustulosten perusteella korkeimmat kiinteistökehityksen tuotot ovat saatavilla keskeisiltä kantakaupungin alueilta Helsingistä, joissa asuntojen hinnat ovat korkeimmillaan.

Avainsanat Residuaaliarvo, Maan arvo, Tontin hinta, Kiinteistökehitys, Kannattavuus, Asuinrakentaminen, Asuntomarkkinat, Investointiarvo

Preface

Urban economics and especially the value of land has always been of a particular interest to me. I have wanted to understand how the real estate market works spatially, why housing prices change and how the developers and other actors in the market define how a city develops. These are the main reasons why I chose the program of real estate economics and I have learned a lot about these subjects during the program.

However, the topic of this thesis came from one of the problems in real estate valuation that has bothered me from the very start of my studies. It seems that there are a lot of different valuation methods and approaches that are used to solve valuation problems from different angles, but these perspectives are somewhat isolated. For example, the market-, income-, and cost approaches are all considered important and used in valuation but there are few methods that try to combine these in a broader context. The residual valuation method is one of the few methods that combines these approaches, but it has most commonly been used in development project profitability analysis. If the method can be used in the valuation of individual projects, then why not try to use it in a broader regional analysis of land values?

The goal of this thesis was quite ambitious, and it required a lot of work to put this together. It would not have been possible without the inspiration and support of all my colleagues at Haahtela-kehitys Oy. The construction modeling at Haahtela is unique in the world and it has been a true inspiration to my own work as well. If the construction markets can be modeled in the finest detail and accuracy, then we have to be able to do the same for the real estate market as well. This thesis is not the solution yet, but it is a good start.

I want to give special thanks to my advisor Ari Pennanen for all the insight and help especially in the early stages of this thesis. We had many good conversations that helped me formulate this research properly. I want to thank my supervisor Kauko Viitanen for being genuinely interested on the topic and for all the good feedback that helped me along the way. Warm thanks go to my friends and family and my girlfriend Maiju for all the emotional support. Also, I want to thank my father Yrjänä Haahtela for mentoring me during all my studies. You have always helped me believe in myself, showed me the value of critical thinking and made me understand that anything is possible.

It is often said that the journey is more important than the destination and I couldn’t agree more! Even though my studies took a while, I could not have spent my time better. The student community at Aalto University has taught me so much that it is hard to even believe it. I have learned valuable skills in organizing and leadership, but most importantly I have had a great time with good friends and learned the importance of community. I want to thank everyone at Aalto University Student Union, Aalto Cocktail and most importantly Maanmittarikilta. Without you guys these years wouldn’t have felt the same. MK 4 life!

Helsinki, 27.5.2019

Pyry Haahtela

Table of Contents

1 Introduction

1.1 Study purpose and background

The purpose of this study is to find out if there is a difference between the development value of residential land and its sales price in the Helsinki metropolitan area (HMA). The estimation of land values has been a topic in the real estate economics since the beginning of the field of study. Many governing theories of land value assume that the value of the location is capitalized in land value to the landowner and this should be reflected in the sales prices of land. The most widely known theories include those of Ricardo (1809), von Thünen (1826), Alonso (1964), Muth (1969), Mills (1972) and Fujita (1989). These theories are sometimes contradicted with real world evidence of the land market behaving differently than expected. For example, in a study by Oikarinen (2014) it is shown that residential land sales prices do not react to changes in apartment prices as fast as expected. This raises a question if land is being sold under its development value in some areas, particularly those where apartment prices have raised recently.

The idea of the research is to approach this matter by using the reasoning of the property developer to estimate the land values in the HMA and to compare this to the actual vacant lot sales prices in the area. The developer is mainly concerned in the transformation of real estate from one state to another, while making the best possible profit on the endeavor. These endeavors include the development of undeveloped land as well as the transformation of land from one use to another. Often the developer does this by acquiring underutilized property that is developed and sold to investors at a profit. The developer can only be able to make a profit if he can estimate the value of the developed property, the construction costs and the right price to pay for the land. Often the developer is not the only participant in the market and he has to make a high enough offer for the owner to sell the property. If the developer's bid is too low the owner might not sell and if the bid is too high the developer might not make enough profit from the project. This problem can only be solved by estimating the development value of the land for the project.

The value that results from this kind of valuation is focused on the development project and is subjective to the developer and the project at hand. However, if the developers are the main actors in the land market then on average these values should affect the market value of the site. If the development projects follow similar assumptions and business models, this should lead to an inter-subjective development value to the group of developers that make up the market. Of course, each development project can be different and the land value from this perspective is different with different kind of development. The differences in development however are smaller in the housing market than in many other parts of the real estate sector as housing is highly regulated with zoning and building regulations. Also, the apartments in urban areas are somewhat standardized and it can be expected that the developers pursue the local standards to avoid unnecessary market risks.

When these factors are considered it should be possible to use the developer's income-based land valuation to find out if there are differences in development values and lot sales prices for residential land in urban areas. The aim of this study is to test this idea and to estimate the residential land development values in different districts in the HMA using the developers reversed income approach that is often referred to as the residual valuation method.

1.2 Research problem

To find out the best ways to approach the subject, we must first formulate the appropriate research problem that can be studied empirically. The purpose of the study can be formulated in to the following research problem: Is it possible to apply the reasoning behind the developer's residual land valuation method to estimate the differences in the value of residential land in an urban area and compare this to land sales prices? To help the formulation of the analysis the problem is divided into research questions that can be answered separately from each other. The research questions are:

1. Are there significant differences in predicted housing development profitability or in residual land values in different districts and apartment types in the Helsinki metropolitan area, when the residual land valuation method is used in the estimation?

2. Is there a correlation between the residual valuation parameters (apartment prices, construction costs, vacant lot sales prices) and the valuation results (expected development returns & residual land values)?

3. Can the residual land valuation method be used to predict vacant lot sales prices in the Helsinki metropolitan area?

The first of these questions aims to find out whether there are any differences in development profitability or land prices in different areas at all using this method. If the differences are not significant, then it is possible that the developers should not be overly concerned of local characteristics and should buy the land that is best suited for the development with the best price available. This can also be the case if housing prices and construction costs have a very high correlation.

The second question aims to find out if the most important factors related to the residual land valuation method have a correlation with each other. If for example the housing prices and construction costs have a very high correlation, it can lead to smaller differences in land values in the model.

The third question aims to find out if the estimates made with the residual valuation method fall anywhere close to actual lot sales prices in the area. If they do not, then the model might have significant flaws, or the method is not very applicable to the estimation of residential land sales prices.

With the research questions we have a good understanding of what this study tries to find answers to but the formulation of the actual empirical part of the research requires hypotheses. Hypotheses help the interpretation of the results, so each research question is also followed with a corresponding hypothesis. The hypotheses are:

1. There are significant differences in development profitability and land value estimates in different districts of the Helsinki metropolitan area using the residual valuation method.

2. There is a correlation between housing prices, construction costs and lot prices. Also, all these parameters of the residual valuation function have a correlation with the estimated development returns and residual land values on individual parameter level.

3. The residual land valuation method can be used to predict vacant lot sales prices in the Helsinki metropolitan area. It provides better estimates in the valuation than any of the parameters alone.

The first hypothesis is the basis of this entire research. If there was no reason to believe that the estimates received from this type of valuation have any differences between areas, then it would make little sense to even consider this as a possible approach to the valuation of vacant lot development values. There is significant data that the vacant lot sales prices of residential land do differ between areas and districts in the Helsinki area and this is an indication that development values of the lots are different as well.

The second hypothesis assumes that the current valuation methods used in the market do not provide specific estimates and this results in less differentiation in sales prices. Also, a very large landowner in Helsinki area is the city of Helsinki that uses set administrative practices to decide the sales prices of land. These practices can lead to undervaluation in high value areas if they are not based entirely on market value.

The third hypothesis assumes that the value of the land is related to the value housing and as the housing market is a derivative of the residential real estate market the differences of housing prices should be reflected on land value. This is indicated by previous theories such as the four-quadrant model of real estate markets that attributes the inherent value of a residential property to the demand of space in the first quadrant (space market), that is represented by the housing market with residential properties. (Dipasquale & Wheaton 1992, p.186). Also some more recent studies have implied that there is a causal relationship between housing and land prices and that land prices follow the changes in housing prices (Ooi & Lee. 2006. p. 1-2).

If the developers value the real estate using the income approach the value of housing should have a significant correlation with the value estimates. This can be contradicted if construction costs have a correlation with housing prices. If the higher housing price can only be achieved through higher costs of construction, then the correlation between housing prices and land values can be smaller. This would however have a big contradiction to the value of location and as such it is considered, while possible, to be a less likely result.

1.3 Limitations

The study is focused on the estimation of residential land values in the urban districts of the Helsinki metropolitan area. With this focus it is important to limit the scope of the study to residential properties only. Residential markets differ from other real estate markets in a number of significant ways. First the residential property market is partly a business to consumer market. In Finland and in Helsinki a very large part of residential properties is owned by private individuals through condominiums and traded as housing stocks that is a very different method of ownership than for example commercial real estate. As a result, these properties are rarely valued as a whole in the market but are valued as individual apartments. There are also investors in the market, but these investors compete with the homeowners which makes the market different from commercial property market. It is thus important to acknowledge that the results of this study cannot be applied to any other form of real estate market without further study related to that market.

Besides the limitation to residential property market another important limitation is that the study focuses on urban areas that consist mainly of multi-story apartment buildings. The dynamics of detached or semi-detached houses can be entirely different based on the literature. One of the most important differences is that large apartment buildings are almost entirely constructed by developers and not the individual homeowners. This is different with smaller units of housing as for example single-family residences are not always constructed by developers but by the individual homeowners. As the study is focused in the behavioral logic of the real estate developers it is evident that these dynamics are most likely not true in areas where the market is not dominated by real estate developers. If the developers do not make up most of the market, then their logic of valuation is most probably not applicable to the estimation of land values in the area.

The third important limitation is the geographical area that the study focuses on. All the data in the study is gathered from the Helsinki area and it is not applicable to any other market area without further research. The dynamics of real estate markets are at least somewhat local, and this results in the fact that any dynamics studied in this research can be different in other market areas.

The final limitation to this study is the focus on land values only. The structure of the research is designed in a way to limit the effects of the buildings in the results. Although apartments are a relatively standardized commodity there are different types of apartments and the age and quality of the building can have a large impact on the apartment prices (Mäkinen 2017). To limit this effect the data is collected only from new residential construction and not from older apartments. This is an effective way to limit the effect of depreciation of the buildings in the empirical results but leads to the fact that the study cannot represent all areas or districts in the Helsinki area. The data cannot be collected comprehensively from all geographical areas as it is limited to the areas that currently have new residential construction. This is a large but necessary limitation to the study that must be considered when analyzing the results.

1.4 Research structure

The research consists of two major parts, a literature review and an empirical analysis. The purpose of the literature review is to find out what previous research has been conducted on residential land valuation and how the residual land valuation method can be used in different types of valuation problems. The literature review is divided into three parts that make up the chapters 2-4 of this study with each chapter focused on different aspects of the subject. The first part of the literature review (chapter 2) consists of a brief overview of the general valuation concepts and approaches used in property valuation. The introduction of these concepts gives a good framework of valuation that can be reflected upon when the different land valuation methods are analyzed in the next chapters. The second part (chapter 3) focuses on different land valuation methods and provides examples on the use of these methods. Although our focus is on the residual land valuation method the other valuation methods are used as a comparison to the residual valuation method to better understand its advantages and limitations. The third part of the literature is entirely focused on the residual valuation method and here we will go into more detail on the application of this method. The residual valuation method is somewhat complex compared to some other methods and it requires a large amount of different types of data as well as some significant assumptions. In this part we examine the previous use of the method and different ways that the construction of the method can be approached based on literature. At the end of chapter 4 we will go through some theoretical implications and assumptions that relate to the residual valuation. These theoretical frameworks are studied so that we can understand the possible problems of the method that need to be considered when analyzing the empirical results of the study. The literature was gathered from the electronic archive of Aalto Finna and is based on the material that was available from this source. The search of the material was based on keywords as well as the references used in other related articles. Following keywords were predominantly used in the search of the material: Residual method, Residual valuation, Land value, Residential land value, Land valuation, Land valuation methods, Land price dynamics, Housing price, Housing price dynamics, Housing supply, Property markets, Property valuation, Development return, Development profit, Development value, Real option analysis. These keywords provided most of the initial material that was used to gather the literature content. Most of the other literature was gathered by using the references of the articles that were gathered using the keywords.

The second major part of this research is the empirical analysis that makes up the chapters 5-6. The empirical part of the research aims to answer the research questions by testing the hypotheses with statistical data. The empirical part is a quantitative research based on statistical data of housing prices, construction costs and vacant lot sales. As a quantitative research the empirical part uses statistical analysis methods to test the hypotheses. The statistical analysis is done in three parts that each try to answer a different research question. The first part uses the residual land valuation method to estimate the development returns and residual land values for each of the observation areas. The aim of this part is to answer the first research question: “Are there significant differences in predicted housing development profitability or in residual land values in different districts and apartment types in the Helsinki metropolitan area, when the residual land valuation method is used in the estimation?”. In addition to answering the first research question, the first part also provides us with the required estimates that are used in the second and third part of the research. The second part of the research uses correlation analysis to analyze the relationships between the different datasets and the residual valuation results. The aim of this part is to answer the second research question: “Is there a correlation between the residual valuation parameters (apartment prices, construction costs, vacant lot sales prices) and the valuation results (expected development returns & residual land values)?” The results of the second part are also used in the third part of the research to evaluate the possible limitations and problems that may result from these relationships between the datasets. The third part of the research uses regression analysis to evaluate the relationship of the residual land values and vacant lot sales prices. The aim of this part is to answer the third and final research question: “Can the residual land valuation method be used to predict vacant lot sales prices in the Helsinki metropolitan area?”

The data that was used in the empirical part was gathered from multiple different sources and the data requirements, gathered data and the sources of the data are all described in detail in chapter 5.

2 Valuation concepts and approaches

2.1 Value concepts

The goal of this chapter is to find out the requirements for land valuation with the residual valuation method. These requirements are approached by introducing the value concepts and valuation approaches that are generally used in the field of property valuation. These concepts are then further analyzed in the context of land valuation. Finally, the requirements for the application of the residual valuation method are briefly summarized at the end of this chapter. The following chapters then focus on the analysis of these requirements along with the residual valuation method to provide us insight on how the empirical part of this research should be constructed.

2.1.1 Market value

The International Valuation Standards (IVS) define the concept of Market value by being "the estimated amount for which an asset or liability should exchange on the valuation date between a willing buyer and a willing seller in an arm's length transaction, after proper marketing and where the parties had each acted knowledgeably, prudently and without compulsion" (IVS 2017, p.18).

IVS regulates the conceptual framework that is required to fulfill the definition of market value. Market value is the most probable price for that can reasonably be obtained in the market. It represents the highest price that is reasonably obtainable by the seller and the lowest price for which the buyer can reasonably acquire the property. It excludes any special contract terms, atypical financing, sale and leaseback arrangements or any element of value available only to a specific owner or purchaser. When these circumstances are present the price of the asset does not represent market value. (IVS 2017, p.18)

The definition of market value presumes that the price is negotiated on an open and competitive market. There is no restriction however to the number of market participants. There can be large amount of market participants or only a few of them as long as the market where the asset is being notationally exchanged is the market where the asset is normally exchanged. (IVS 2017, p.19). The definition does not restrict the approaches that can be used to estimate a market value. It can be estimated using market approach, income approach or cost approach as long as the valuer can show that the inputs of these methods are used by the market participants and they give them significant weight in the transaction. (IVS 2017, p.19). Also, it is important to note that the value of an asset is an estimated amount and it does not refer to the actual historical sale price of the property. However, market value is the price in a transaction that fulfills all the requirements of the market value definition at the valuation date. (IVS 2017, p.19)

2.1.2 Investment value

The IVS definition of investment value is "the value of an asset to a particular owner or prospective owner for individual investment or operational objects" (IVS 2017, p.22). IVS further specifies that investment value is an entity-based value: "the (investment) value of an asset may be the same as the amount that could be realized from its sale to another party, this basis of value reflects the benefits received by an entity from holding the asset and, therefore, does not involve a presumed exchange." The main differentiation from market value is the fact that investment value does not deal with the concept of selling the asset and is a subjective value that may only be received by a particular owner. (IVS. 2017 p.22)

Investment value for a particular owner or a potential owner can be calculated for an asset that does not have an existing market or available market information, but that value does not necessarily reflect the value that the property may have for a different owner. As such, the investment value can be higher or lower of the market value and can exist even when market value cannot be reliably calculated. Investment value is always based in the benefit for the owner that may be received in holding the asset. (IVS. 2017 p.22)

2.1.3 Fair value

In addition to market value and investment value the concept of fair value is often used in property valuation context. The exact definition of the word seems to rely on the context of the valuation. International financial reporting standards define fair value as "the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date" (IVS 2017, p.23). It seems that this definition does not substantially differentiate from the IVS definition of market value and is more or less a synonymous expression for the relatively same definition. However, the concept is also used in a legal context in many national legislations where the definitions can vary significantly and may be the result of legislative action (IVS 2017, p.23).

In the context of this study the legislative framework of value is not a main topic of research and for this reason the study focuses mainly in market value and investment value and the concept of fair value is not discussed any further in the study.

2.2 Valuation approaches

Valuation approaches are the ways that a valuer can address a valuation problem. There are three main approaches: Market approach, Income approach and Cost approach (IVS 2017, p.29). In this chapter our goal is to describe these approaches so that we can utilize this categorization in the further analysis of the land valuation methods. Valuation can be done for multiple purposes. Common purposes include, but are not limited to financial reporting, tax reporting, litigation support, transaction support, and to support secured lending decisions. (IVS 2017, p.4). It is important to understand the purpose of valuation to ensure that the valuation is not used out of context and it will typically influence or determine the bases of value to be used (IVS 2017, p. 10).

The purpose of the valuation however does not define the approach of the valuation, but rather IVS emphasizes the responsibility of the valuer in choosing the most appropriate method (IVS 2017, pp. 29-30). The value concept has an essential effect in the valuation as they all provide unique definitions and limitations to the value of a certain property. The chosen methods should be consistent with these views and limitations and they should utilize the best available market information. The goal in selecting valuation methods is to find the most appropriate method under the circumstances. These circumstances include the bases of value (value concepts), strengths and weaknesses of the methods, nature of the asset (property), methods used by participants in the market, and the availability of market information. (IVS 2017, p.29). The next subchapters briefly summarize the valuation approaches defined by the IVS (2017)

2.2.1 Market approach

Market approach is defined by the IVS as an approach that "provides an indication of value by comparing asset with identical or comparable (that is similar) assets for which price information is available" (IVS 2017, p.30). As stated by the definition market approach requires information on the recent transaction price of the subject asset or substantially similar assets or that similar assets are actively publicly traded in the market. (IVS 2017, p.20). The heterogeneous nature of assets means that it is often not possible to find market information involving identical or even similar assets (IVS 2017, p.31). Even in these cases where market approach cannot be used in this sense, it is best to maximize the market-based inputs in the application of other approaches (IVS 2017, p.31). Most typical application of the market approach is comparable transactions method, but there are also numerous other methods that should be judged based on whether the market participants give substantial weight on them or not (IVS 2017, pp.31-35).

2.2.2 Income approach

Income approach is defined by the IVS as an approach that "provides an indication of value by converting future cash flow to a single current value" (IVS 2017, p.36).
Under this definition the property is only as valuable as the income it can produce. In detail the value is determined by the value of income, cash flow or cost savings that can be achieved by holding the asset. (IVS 2017, p.36). It is recommended to use the income approach when the income from the asset is the main creator of value in the participant's perspective and the projections of future income can be made with reasonable accuracy and suitable information of comparable assets is not available to use the market approach (IVS 2017, p.36). The income approach does not require that the property is producing any income at the current state as long as reasonable future projections can be made (IVS 2017, p.37).

2.2.3 Cost approach

The predominant definition of the cost approach is defined by IVS (2017) an approach that "provides an indication of value using the economic principle that a buyer will pay no more for an asset than the cost to obtain an asset of equal utility, whether by purchase or by construction, unless undue time, inconvenience, risk or other factors are involved" (IVS 2017, p.42). Thus, cost approach has the predefined assumption of replacement or reproduction of a particular asset with a perfect substitute considering utility, not necessarily other characteristics. IVS (2017) suggests that the cost approach should be given significant weight when "the participants would be able to recreate an asset with substantially the same utility as the subject asset, without regulatory or legal restrictions, and the asset could be recreated quickly enough that a participant would not be willing to pay a significant premium for the ability to use the subject asset immediately" (IVS 2017, p.43). Also, weight should be given to the approach if: "the asset is not directly income generating and the unique nature of the asset makes using an income approach or market approach unfeasible, or the basis of value used is based on replacement cost such as replacement value" (IVS 2017, p.43). Deteriorating assets such as buildings also need deductions for physical deterioration and other relevant forms of obsolescence. (IVS 2017, p.43)

The cost approach can be broadly divided into three methods: replacement cost method, reproduction cost method and summation method (IVS 2017 p.43). Replacement cost method seeks to substitute the utility that the current asset provides, and the cost are calculated to a similar (but not necessarily identical) asset with equivalent utility.

Reproduction cost method is used to calculate the costs of reproducing an identical replica of the asset. Summation method is somewhat different from these with a goal to calculate the value of an asset from separate values of its component parts. (IVS 2017, p.43)

2.3 Discussion of value concepts and approaches

In the context of this study it is important to understand how these valuation approaches relate to the estimation of the market value of land. There are some guidelines for the application of the methods set by the IVS (2017). There is no priority between the different approaches, but they have practical limitations for example facing the market information required. The valuer should maximize the use of relevant observable market information that relates to the base of value that is being analyzed. (IVS 2017, s.30). The availability of relevant and observable market information can sometimes prove to be a problem as the property market is not a very efficient market compared to for example the stock market (Evans 2004, s.60).

The market approach is one possibility in estimating land value. When the market value of land is estimated it is important that the valuation reflects the market circumstances (IVS 2017, pp.18-19). The observation of transaction via the market approach can be useful as the observation of transaction prices reflects the value perceived in the market and not the value for a particular investor or a particular business case. However, in the land market transactions are infrequent and dispersed so that the gathering of relevant market data from transaction can prove to be a problem and this limits the applicability of the approach.

Another possibility would be the use of income approach. As stated above, IVS recommends the use of income approach when the income from the asset is the main creator of value and suitable information of comparable assets is not available to use the market approach (IVS 2017, p.36). It is hard to say whether land falls into this category. Land can be a part of an income producing asset in a developed property or it can be rented separately from the buildings. If the land produces income by itself, the income approach can be utilized in land valuation by analyzing the income streams the asset produces (IVS 2017, p.36). In many cases land is owned by the same owner that owns the buildings. In these cases, the land is often not rented separately from the buildings and does not produce income by itself but rather produces income together with the buildings. This is problematic when the income approach is used as the income streams come from the entire property and not specifically from the land. In these situations, separation of the income streams to the land component and the building component is needed so that the income approach can be used in land valuation. The income streams need to be related to the asset that is been valued so that income approach can be used (IVS 2017, p.36). Another problem related to this approach is that income streams are often related to a certain business logic that is tied to the investor that is holding the asset. This leads to the fact that the value is measured in relation to a particular owner or business logic and as such reflects the investment value rather than market value of the asset (IVS 2017, p.36). The utilization of this approach in the evaluation of market value requires the use of inputs and assumptions that would be adopted by the market participants (IVS 2017, p.18).

The third possible approach is the cost approach. As stated previously the cost approach is usually relevant in valuation when the participants would be able to recreate the asset, the asset is not directly income generating, or the basis of value used is based on replacement cost such as replacement value (IVS 2017, p.43). The cost approach would require the valuer to estimate a cost for which the asset can be replaced or recreated. In land valuation this can be problematic when it is considered that each site has a unique location and thus at least partly different characteristics.

The different value concepts and valuation approaches give a good framework for further analysis of land valuation. First part of this chapter described the value concepts that determine the boundaries for valuation as for example the market value and investment value have different meanings (IVS 2017, pp. 18-22). The second part of this chapter introduced the valuation approaches that are used to categorize different valuation methods that use different kinds of data in the determination of value (IVS 2017, pp. 30-43). The value concepts have different meanings and they have different requirements for the valuation. The requirements of market value are that all the assumptions reflect market conditions and that the asset should be exchanged in a transaction with an amount that reflects the market value (IVS 2017, p.18). The investment value does not require the assumption of a transaction but rather focuses on the asset's ability to generate income that can be reasonably estimated (IVS. 2017 p.22). The main difference between these concepts is that the investment value can be subjective to a particular investor and business logic whereas the market value must be inter-subjective and shared by the market participants. The value concepts and valuation approaches that were presented in this chapter are used in the following chapters of the literature review, where we discuss the theories and empirical evidence that relate to these the application of the residual valuation method in land valuation.

3 Land valuation methods

The aim of this chapter is to list and describe the most common methods for land valuation. One of these methods is the residual land valuation method that has been chosen as the method for the empirical analysis of this study. It is important however to describe the other common methods of valuation as well so that we can better understand the nature of the residual valuation method and how it is different from other common methods. Also, we aim to classify the land valuation methods according to the IVS (2017) principles that were introduced in the previous chapter. This will help us understand the requirements and limitations that relate to the use of these methods in land valuation and to build a bridge between the value concepts, valuation approaches and the valuation methods.

The chapter is divided into three parts. The first part introduces a brief overview of different property valuation methods and their classification described by Pagourzi et al. (2003). The second part describes in more detail some of the methods that are most commonly used in land valuation. The third part contains a discussion of the contents of this chapter where the different methods are compared and analyzed in relation to this study.

3.1 Overview of land valuation methods

A review of the valuation methods considering both developed property and land was conducted in a study by Pagourzi et al. (2003). The study listed valuation methods as traditional valuation methods and advanced valuation methods. Traditional methods include: comparable method, income method, profit method, residual method, cost method, multiple regression method and stepwise regression method. (Pagourzi et al. 2003 p.4). The comparable sales method is the most widely used and it focuses on analyzing comparable sales of the subject property or land parcel. The income method and the profits method focused in analyzing the income streams related to the property via direct capitalization or capitalized cash flows. (Pagourzi et al. 2003 pp.6-7)

Pagourzi et al. (2003) also list several non-traditional methods that are categorized as advanced methods. These include artificial neural networks, hedonic pricing models, spatial analysis methods, fuzzy logic and autoregressive integrated moving average. The basis of artificial neural networks is in the self-learning algorithm that combines the input data layer that can include property prices and characteristics and hidden layers that combine these to produce the price estimates. Hedonic pricing models define the property or the land component as a vector of different characteristics that can be analyzed independently to produce an understanding of value components that create the value of the property. The spatial analysis method also focuses on value components but uses spatial tools to analyze these aspects in a more detailed manner. (Pagourzi et al. 2003. pp. 12-15). Pagourzi et al. (2003) also introduce the use of fuzzy logic and autoregressive integrated moving average models that are more mathematically focused tools but have similar limitations to the other models regarding the use of observable market information (Pagourzi et al. 2003. pp. 15-17).

As listed by Pagourzi et al. (2003) there are several different methods considering property valuation. In the first chapter we introduced the valuation approaches defined by the IVS (2017). If we follow these definitions we should be able to categorize these valuation methods into either the market approach, the income approach or the cost approach.

Methods that focus on the asset transactions can be categorized to the market approach (IVS 2017, p.20). Pagourzi et al (2003) describe that the comparable sales method, multiple regression method and stepwise regression method deal with sales data of properties. For this reason, they can be categorized to market approach (IVS 2017, p.20). The Income method and profits method attribute value to the income that can be generated to the owner. In the cases that income is the main creator of value the methods should be categorized to income approach. (IVS 2017, p.36). The cost method described by Pagourzi et.al (2003) focuses on the recreation of the asset and does not give any weight to the income or transaction data of the asset and should be categorized to cost approach. (IVS 2017, p.43)

The residual land valuation method however is not as easy to categorize. It has three essential components: developed property value, development costs and the resulting residual land value. The development costs can be calculated for example with a method categorized to the cost approach. The developed property value can then be calculated using either a method belonging to the market approach or the income approach. (Pagourzi et al. 2003, p.7). This leads to a situation where the residual land valuation method is a mix of traditional approaches and focuses on more than one approach to estimate a land value. This view is supported by Skarzyński (2006) who classifies the method as a mixed approach.

The focus of this study is in the valuation of the market value of land with the residual valuation method. For this reason, the residual valuation method must be explained in more detail and this is done in the next subchapter. From the methods listed above the hedonic pricing method is one of the most popular methods used in land valuation and we will also describe this method in more detail to compare it to the residual valuation method. One method that was not described by Pagourzi et al. (2003) is the real option analysis that has been used in land valuation (Vimpari 2014, p. 13). This method is different from both the residual valuation method and the hedonic pricing method and we will use this method in the comparison as well. The method is further described in a subchapter below.

3.1.1 Residual land valuation method

There is a good amount of research about the residual valuation method considering its use in the investment analysis of development projects. In this chapter we will go through the research of (Pagourzi et al. 2003), Skarzyński (2006) and Greenhalgh & Bendel (2015) who describe the method and provide examples of its use. We will also introduce some parts of the method through the studies of Darlow (1982), Newell (1989) and (Havard 2014). There are also some studies that apply the residual valuation method to the valuation of developed real estate. The studies of Wolverton (1993) and Boyd & Boyd (2012) deal with these applications of the method.

Pagourzi et al. (2003) define the residual valuation method as one of the property valuation methods that is specifically focused in land valuation. In residual valuation method the developed form of the land (for example residential real estate) is valued with another method and all the costs of developing the land are then subtracted from the value resulting in a form of residue that represents the maximum capital expenditure for buying the land. (Pagourzi et al. 2003, p.7). This method has been studied further in the development process at least by Skarzyński (2006), Greenhalgh & Bendel (2015) but also as a method for the valuation of the land component of developed real estate for example by Wolverton (1993) and Boyd & Boyd (2012). Some of the more detailed aspects of the method such as the developer’s profit and the effect of the development process have been studied at least by Crosby, Devaney & Wyatt (2018) and Ogunbayo et al. (2018).

In his study Skarzyński (2006) presented the theoretical basis of the method and analyzed its different applications. According to Skarzyński (2006) the residual method is predominantly used for performing profitability analyses of development and redevelopment projects or either to define the value of the land component or the value of the buildings as a part of developed property (Skarzyński 2006, p.1). All this considered the residual valuation method falls into the category of mixed approaches with an assumption that development must be profitable for the product to exist (Skarzyński 2006, p.1).

Table 1. Approaches and methods used in Poland for property valuation (Skarzyński 2006, p.2)

Market value			Reproduction Value		Mixed approach	
1. Appraisal by comparison or comparable approach	2. Appraisal by capitalization or revenue generating approach		3. Cost approach		4. Mixed approach	
METHODS	METHODS	TECHNIQUES	METHODS	TECHNIQUES	METHODS	TECHNIQUES
1	2	3	4	5	6	7
1.1. The method of comparing pairs	2.1. The investment method	2.1.1. The simple capitalization 2.1.2. The discounted cash flow	3.1. The method of reproduction costs	3.1.1. Detailed 3.1.2. Integrated elements 3.1.3. Indexes	4.1. THE RESIDUAL METHOD	The mixed approach can contain the elements of the approaches: – comparable – revenue generating – cost
1.2. The method of average price correction	2.2. The profits method	2.2.1. The simple capitalization 2.2.2. The discounted cash flow	3.2. The method of replacement costs	3.2.1. Detailed 3.2.2. Integrated elements 3.2.3. Indexes	4.2. The method of liquidation costs	
1.3. The method of statistical analysis of the market					4.3. The method of estimaited farmland indexes	

Pagourzi et al. (2003) describe that the residual valuation method is particularly useful when land is redeveloped from its existing use toward the theoretical “best and highest” use. In these cases, the redevelopment is a way to release latent value that is a result from the increase of land value due to more profitable land use. (Pagourzi et al. 2003, p.7). Skarzyński (2006) further describes that the residual valuation method is an important tool while taking action in any real estate development (Skarzyński 2006, p.2). The residual valuation method is also listed as one of the five principal property valuation methods in the UK and much of the developed world (Greenhalgh & Bendel 2015, p.3). Darlow (1982) argues that the main essential purposes of the residual method are the calculation of the maximum acquisition price of a land parcel, the calculation of the expected profit from the development project and the calculation of the cost ceiling for the construction in the case that land has already been acquired (Greenhalgh & Bendel 2015, p.3). Also, Newell (1989) argued that the residual method is the key in calculating land bid prices for the developers when land is acquired (Greenhalgh & Bendel 2015, p.3).

The residual valuation is widely used and recognized in the property development field and it is particularly used in determining the land bid prices of the developers. The problems relating to application of the method are closely related to its mixed approach that requires the utilization of other valuation methods. Havard (2014) describes some of the varying techniques that are: traditional residual method, residual cash flow approach and discounted cash flow approach. All the techniques listed are focused on the analysis of development profitability with variations in the determination of profits or cash flow. (Greenhalgh & Bendel 2015, p.4)

There are also techniques that focus more on the valuation of the land component of developed property. In Australia the residual method is used to split the value of developed property into land and building components for land taxation purposes. This method is used in conjunction with the comparable sales method for valuing the developed property from where the depreciated construction cost is then subtracted. This however presents a problem for the valuation of the developed property especially if the property is complex and there is not enough comparable evidence. (Boyd & Boyd 2012, pp. 1-2)

3.1.2 Hedonic pricing models

The hedonic pricing model is based on the fact that housing as well as residential land are heterogeneous goods and to be able to compare them one should understand the underlying attributes that affect the price of the good. The theoretical framework of hedonic price models is often attributed to Rosen (1974) who provided the foundation of non-linear hedonic models. (Owusu-Ansah 2011, p.3). However, there are even earlier examples of the use of hedonic pricing in developed property and land valuation. Renshaw (1958) used multiple regression analysis to demonstrate the relationship of agricultural land and four different attributes and Pendelton (1965) demonstrated that average sales prices of houses could be predicted with a statistical model. (Adair & McGreal 1987, p.2). In Finland for example Lönnroth (2017) introduced a study of both housing prices and residential land rents in the Helsinki area using the hedonic pricing method. Peltola & Väänänen (2006) introduced a study that used hedonic pricing method to determine both residential land prices regionally as well as the characteristics that had the most impact on the prices statistically.

A hedonic regression curve shows the relationship between a dependent variable that represents the price of the property and an independent variable that represents the value of some characteristic for which there is an assumed price relation (Owusu-Ansah 2011, p.3). These price relations are usually first tested for correlation then tested via multiple regression analysis. There are variations in the functional forms used in the regression analysis, but they all include this basic formulation. These functional forms can be divided into parametric and nonparametric approaches but there are also applications that use a hybrid of these methods. (Owusu-Ansah 2011, p.3)

The parametric form assumes that the regression curve has some pre-specified functional form while the non-parametric form does not have this assumption and is deemed more flexible but more complicated as well (Owusu-Ansah 2011, pp.4-7). The parametric form includes for example log-linear ordinary least squares, box-cox OLS and weighted least square models. These parametric models are the most widely used models and if the assumptions are right they can be accurate and relatively easy to apply in modeling. However, the assumptions can be restricting, and these methods have been criticized for strong assumptions such as linearity between dependent and independent variables. There are also problems when the assumptions are violated for example when there is an assumption of homoscedasticity and heteroscedasticity is present in the model. (Owusu-Ansah 2011, pp.6-7).

The non-parametric forms include for example Kernel regression method, nearest neighbor method and locally weighted regression. These forms provide adaptability in the exploration of relationships between dependent and independent variables and work with smaller datasets and with missing values. However, since these methods are based on local averaging the observations become sparsely distributed even for large sample sizes and it can lead to inaccurate estimations. (Owusu-Ansah 2011, p.7)

While there are problems with the hedonic pricing method it is still widely used, and it can provide good practical answers if its conditions are met. In practical application it is most essential that the valuer is aware of the limitations of the model before jumping into conclusions. (Adair & McGreal 1987, p.4). The biggest problem with the method in the valuation perspective is data quality. The data must be suitable for regression analysis and if there is an assumption of linearity the data should be normally distributed and on a continuous scale of measurement. These problems can be approached with mathematical tools that enable the valuer to transform skewed data and outliers can be removed from the model. Also, the models can include dummy variables (using values from 1 to 0) to control some of the circumstances where data is not continuously measurable. (Adair & McGreal 1987, p.4)

Still none of these corrections can set aside the fact that regression analysis needs a large dataset representing both the independent and the dependent variables and its use is limited to the situations where sufficient data is available. There is a lot of discussion on what amount of data is enough but the more there are independent variables in the model the more data is usually needed to account for the variation of the variables. Also, studies indicate that multiple regression analysis works best only in well-defined market areas and thus their application can be problematic to wider areas. (Adair & McGreal 1987, p.4). There is also a major problem caused by multicollinearity where the different independent variables influence each other, and this cannot be eliminated in the creation of variables. Multicollinearity is quite common at least in housing research. (Adair & McGreal 1987, p.4). These problems can be controlled in regression models technically using for example stepwise regression and excluding statistically insignificant variables from the model. The overall statistical significance can also be tested using F-test to the final regression model. These statistical requirements however further limit the application of such models into a few key independent variables and this can lead to problems in determining models that can truly predict the measured data any more than partly. (Adair & McGreal 1987, p.5)

Gallimore, Fletcher & Carter (1996) presented a study that focused in modeling the influence of location on housing prices where they studied the different methods that can be used to isolate location components from housing value. In a study by Isaksson (1997) land value components were modeled in detail and even some unconventional attributes were analyzed such as the impact of buyer and seller characteristics to vacant land prices. Ahlfeldt (2008) on the other hand used hedonic regression with data from Berlin to test the hypothesis of Alonso (1964) that land values are reliant on the distance of the central business district. In his study Ahlfeldt (2008) concluded that this relation has been shown in previous studies, it can be diluted by the effect of cheap and efficient transportation in a city such as Berlin where transportation network is well developed. (Ahlfeldt 2008, p.25). In Finland Peltola & Väänänen (2006) constructed hedonic models where they combined vacant lot sales data with housing price data and made an extensive list of characteristics that influence land value. These characteristics included many locational factors like the distances to central city areas as well as micro-locational factors such as the distance to waterfront, highways and public transit stop locations (Peltola & Väänänen 2006, p. 16-20).

3.1.3 Real options analysis

The Hedonic pricing model and the residual valuation method were introduced in the previous chapters. There are however other methods that can be used to land valuation such as the real options pricing method that has lately gained popularity. Foundations for the method have been provided by Myers (1984), Kester (1984), McDonald & Siegel (1986) and Pindyck (1991) who highlighted the importance of options in investment decision making (Vimpari & Junnila 2015, pp.1-2). The method has been used in varying cases for example by Yao & Pretorius (2014) and Cunningham (2006). The subject has been extensively studied in Finland by Vimpari (2014), Vimpari & Junnila (2014), Vimpari et al. (2014) and Vimpari & Junnila (2015).

The real option analysis is used in the field of real estate to explain market phenomena such as market behavior, development cycles, role of competition and risks related to development decisions (Vimpari 2014, p.12). One of the original applications of real options analysis is in land valuation. In land valuation context the method is used in cases where there is uncertainty of future conditions that affect the development and when there is an option to wait for later development instead of developing the property right away. (Vimpari 2014, p. 13). The method is argued by Capozza & Sick (1994) to be useful in the valuation of land that is not in its best and highest use and could be developed to a more profitable use (Vimpari 2014, p.13). Grissom et al. (2010) argue that this is also the case in many development projects as the projects have different development options and the method can be used to evaluate which of the options is most profitable in short or long term (Vimpari 2014, p.13). Cunningham (2006) described that in land valuation uncertainty considering for example future housing prices raises land prices as the owners of the land hold a call option that gives them the right but not obligation to develop an optimal building in the future by paying the price of construction cost (Cunningham 2006, p.3).

The real options analysis is conducted by considering these different scenarios that the landowner has as options and estimating a value for the options. In the case where the options relate to future development, the finance costs of holding the land must be considered in the analysis. (Vimpari 2014). In practice, these options can vary between projects and include strategic decisions about the development process as well. Supporting evidence from the Helsinki area was gathered for example in a case-study by Virenius (2014) who studied the application of the method in a large development project in Kruunuvuorenranta, Helsinki. In the study the author discovered that the options to phase the project in different phases and the possibility to delay the phases if needed raised the project value by 22%. Virenius (2014).

3.2 Discussion of land valuation methods

During the previous chapters we introduced an overview of property valuation methods and described in more detail the residual valuation method and in comparison, to this method two completely different methods: hedonic pricing models and real options analysis. The method that we focus on, the residual land valuation method, has been used both for the analysis of development projects as described by Pagourzi et al. (2003), Skarzyński (2006) and Greenhalgh & Bendel (2015) and in the valuation of land components of developed properties as described by Wolverton (1993) and Boyd & Boyd (2012).

Regardless of its use, the land residual method is not a standalone method for valuation but rather combines other methods to extract the value of the land and building components. The method requires other methods to calculate the total development costs and the end-product value. There are no existing limitations to the use of different methods to value these components. The development costs could be valued for example with methods belonging to the cost approach, such as replacement cost or depreciated replacement cost. The end-product value could be valued using methods from the market approach, such as comparable sales or with methods from the income approach such as discounted cash flow analysis. (Pagourzi et al. 2003, p.7). The residual valuation method is often utilized to estimate investment values for investors or developers. To use this method in the estimation of market value it is required that the business logic and the goals of the investment are shared between the market participants. The evaluation of market value requires that the definitions set by the IVS (2017) are met by the data and the valuation methods. There must be adequate evidence that the assumptions used reflect the market conditions so that the estimate can be considered as market value (IVS 2017, p.18).

The hedonic pricing method approaches land valuation from a different angle compared to the residual valuation method. Where the residual valuation method bases its reasoning to the analysis of developed property values and development costs it is possible with the hedonic pricing method to focus on land prices only via the analysis of transaction data. The hedonic pricing method follows the assumption that value can be separated to value components that can be analyzed separately and used to construct a model that can predict transaction prices (Owusu-Ansah 2011, p.3). Compared to for example the comparable sales method the hedonic pricing method can provide information that can be much more widely applied than comparable sales as individual characteristics are statistically defined from a larger dataset instead of comparing a few case examples. When the statistical data requirements are met the method can be very useful in determining value components related to land value and many of the results are in line with the governing theories of the field. (Adair & McGreal 1987, p.4).

The main problem with the hedonic pricing models that focus on land transactions is the availability of transaction data. This data can be hard to find especially in areas that are fully developed and where there has been no interest by the market participants to sell the properties. In these areas there can be much more transaction data available from the sales of the end products, either the residential property or parts of it such as apartments. In these cases, the residual valuation method can be easier to apply than a hedonic model of land transactions. However, if there is not enough data of the end-product sales, then the residual valuation method faces the same problem (Boyd & Boyd 2012, pp. 1-2). Also, the hedonic pricing method is not limited to the analysis of land values. It has also been used extensively to predict the developed property values (Adair & McGreal 1987, p.2). As the residual valuation method is mainly the framework of combining the valuation of the developed property value and development costs there is nothing preventing the use of a hedonic model in the estimation of the developed property value as a part of this analysis.

The third method that was described in the previous chapter is the real options analysis. This method describes that the options related to the development of a property can have a large impact on the property value and the land component value. (Vimpari 2014, pp.12-13). Although this study focuses on the value of land and not on development projects it is fair to assume that the developers buying the land consider these different options of land use and they can have a significant effect on the acquisition prices that the developers are willing to pay. The real options analysis is a method that could be tested together with the residual valuation method, since it enables considering multiple scenarios as options that can lead to more accurate estimates in land values. Although interesting, this analysis would require a much wider and deeper study that can be included into this paper and thus these possibilities are included into the suggestion for further research that is discussed in chapter 8.

4 Residual land valuation

The aim of this chapter is to further analyze the residual land valuation method and its components. A brief introduction of the method was presented in the previous chapter and in this chapter, we will go through the details of applying the method to land valuation problems. This chapter is divided into three parts.

In the first part we list the components of the residual valuation method and then analyze each component separately through the research that been conducted on the subject.

The second part of the chapter tries to answer some unanswered questions about the limitations of this method through the theories that describe the dynamics of property and land markets. The aim of this part is to highlight some theoretical assumptions that are used in the residual valuation method and to find out how the main governing theories of property and land markets support or contradict these assumptions.

In the third and final part we will discuss the details of the method and its theoretical implications in relation to what research has been made on the subjects. This will provide us with a framework to the empirical part of this study and help us understand the limitations of this method in the empirical research.

4.1 Residual land valuation components

The residual valuation method breaks down the valuation of real estate into the components of the end-product price (housing price), construction costs and developer's profit (Pagourzi et al. 2003, p.7). As described by Boyd & Boyd (2012) the application of the residual land valuation method for developed property also requires the valuation of the developed property itself along with the construction costs or depreciated construction costs of the property. Also, the price of capital investment that reflects the value of capital and the risk taken by the developer must be evaluated when this method is used (Pagourzi et al. 2003, p.7). These three components of the method are further discussed in the next three subchapters that focus on the methods that can be used to estimate these components. In the first part, we analyze the studies regarding housing prices through the research of Dipasquale & Wheaton (1992), Archer & Ling (1997), Evans (2004), Oikarinen (2007), Laakso (1997), Zahirovich-Herbert & Gibler (2014) and Oikarinen, Peltola & Valtonen (2015). In the second part, construction costs are analyzed in the Finnish market through the studies of Haahtela & Kiiras (2014) and Haahtela 2015. In the third part, the developer's profit is analyzed through the studies of Crosby, Devaney & Wyatt (2018) and Ogunbayo et al. (2018).

4.1.1 Housing prices

One of the main components of the residual valuation method is the price of the developed residential real estate. In the English literature the residential property market is often referred as the housing market. This term is used widely in the references of this part for example by Jud & Winkler (2002) and Capozza et al. (2002) and in the Finnish studies that are written in English Oikarinen (2007), Oikarinen and Peltola & Valtonen (2015). The reason for the use of these different terms to describe the same phenomenon is the fact that in many countries residential properties are also traded in the form of apartments.

In legal terms the apartment ownership is treated and defined somewhat differently in different countries. Apartment ownership can be divided into direct and indirect ownership where direct ownership means that the legal status is equivalent with property ownership (Lujanen 2004 p. 58). In Finland, multi-story apartment buildings are usually legally defined as housing companies where apartments are being traded as stocks of the company. This system has some of the characteristics of indirect ownership, where a separate legal entity owns the property and the residents are members or joint owners in this legal entity (Lujanen 2004 p. 62). However due to the substantial commitment of the shareholder to the housing company the share ownership is treated more as equivalent to direct property ownership (Lujanen 2004 p. 58). The Finnish system has some unique traits compared to the ownership methods in many other countries and has some of the characteristics of indirect ownership, but in terms of legal commitments it seems that it is best to be compared to direct ownership practices in other countries (Lujanen 2004 p. 58).

The price dynamics of the housing market have been studied for example by Jud & Winkler (2002) and Capozza et al. (2002). The division of property markets into different submarkets and their dynamics and relations have been studied for example by Archer & Ling (1997), Evans (2004) and most notably Dipasquale & Wheaton (1992). In the Finnish context this topic has been examined extensively by Oikarinen (2007) and Laakso (1997). The effect of supply factors, new construction and available stock to apartment prices have been studied for example by Zahirovich-Herbert & Gibler (2014) and Evans (2004) and in the Finnish context Oikarinen, Peltola & Valtonen (2015).

According to Dipasquale & Wheaton (1992) the real estate market can be divided into four submarkets: Market for space, investments, construction and available stock. This framework defines the dynamics that shape the interaction between these markets and it has had a long-standing effect in real estate market analysis. Dipasquale & Wheaton model or the four-quadrant real estate market model starts with the demand of space in the space market which is the first quadrant. In residential context this is the number of households that compete for the same apartments along with the existing stock of such apartments. The result of the supply and demand in the space market is the market rent of the apartment. (Dipasquale & Wheaton 1992, p.186). The second quadrant, called the asset market for ownership, is where the investors compete for the ownership of rent-producing properties for which the market rent is defined by the space market. The result of the asset market is the price for which the properties are traded. It is defined by capitalizing the market rent with the investors required rate of return. (Dipasquale & Wheaton 1992, p.188).

The property prices defined by the asset market follow to the third quadrant: the construction sector. Construction sector is where the property developers decide on new development projects. If the prices are high enough to generate profit from additional construction, then new construction will occur. It is assumed that the developers are more eager to construct with higher profits and thus a high price for properties will lead to more additional development. The final quadrant in the model is the available stock of apartments that is present at the market. In this final quadrant the additional development defined by the third quadrant along with property depreciation and obsolescence creates the new stock available in the space market and thus defines the supply present in the next cycle of the space market. The new supply affects the equilibrium in the market and along with space demand continues to define the market rent level. (Dipasquale & Wheaton 1992, p.188). In general, the first two quadrants of the model (space and asset markets) are often referred to as property market and the division of the property market to these two markets is sometimes referred to as a two-market-model (Archer & Ling 1997. p.7).

There is critique for the two-market model especially in relation to the fact that the model assumes that the capitalization rates or yields that define the asset value are exogenously determined and as such do not reflect the properties and riskiness of a single investment (Archer & Ling 1997. p.7). Archer & Ling argue that this can only be fixed by adding a third component, the general capital market, to the equation. In their model the property market is categorized in three dimensions: space, property and general asset markets. The distinction between this model and the four-quadrant model is that in Dipasquale & Wheaton's model the asset market defines the asset value of the property by capitalizing the rent from space market with an exogenous capitalization rate whereas the model proposed by Archer & Ling argues that the role of the asset market is to combine the property specific risk with the general yield of the capital market. The end results of this analysis are the risk adjusted yields of a specific properties in the market that do not have an equal standing in the minds of the investors. (Archer & Ling 1997. p.7). The Archer & Ling model defines the role of the developer with the question "Is the asset value higher than the total construction costs". If the value is higher the project should continue and if not, it should be aborted. (Archer & Ling 1997. p.7)

There are several methods that can be used in the valuation of residential properties. These methods were reviewed in chapter 3 of this study. The traditional methods include: comparable method, income method, profit method, residual method, cost method, multiple regression method and stepwise regression method. (Pagourzi et al. 2003. p.4). There are many examples of the hedonic pricing method from the Finnish market that has been used to calculate factors that influence housing prices for example by Laakso (1997) and Oikarinen (2007). The method is useful when the focus of the study is in the differentiation of products such as housing caused by differences in the physical product or in the location of the property (Laakso 1997 p.25).

In this study the differentiation of the product is not the focus and these differences are controlled in the collection of the data. The differences in the location are also controlled by collecting the data from predefined districts in the greater Helsinki area and the differences caused by location are studied with the residual valuation method. For these reasons the aim in the collection of housing price data is to construct local averages of housing prices that are used in the residual valuation method instead of modeling the price factors of individual apartments. The construction of the local averages is further described in the empirical part of the study that starts from chapter 5.

4.1.2 Construction costs

This part is focused on the analysis of the construction costs of properties in Finland. As this study is focused on the Finnish property market and there is enough evidence of construction cost estimation from Finland we prefer to limit the discussion to these methods. There are of course numerous other methods used in other markets, but as the focus of the study is on the Finnish market, the further analysis of these methods would not have a large benefit for this study.

The costs related to residential development projects in the Finnish market have been studied widely by Haahtela & Kiiras (2014). Haahtela-kehitys Oy has published a software in target costing (TAKU) that is widely used all over Finland in construction cost estimation and has achieved a status of a standard in the field. Although the methods are rather complicated to use without the software their theoretical basis is described more closely by Haahtela & Kiiras (2014). In their paper Haahtela & Kiiras (2014) divide their cost theory into three different areas of cost estimation that have their own price concepts. The price concepts are new construction price, present price and renovation price. (Haahtela & Kiiras 2014. pp.47-48).

New construction price represents the amount capital required to construct a new building with characteristics that are similar to the property being valued but with present methods, constructed at the time of valuation and has no relation to the sales or market value of the property (Haahtela 2015. p. 5). This approach can be used to value development projects as well as the replacement values of existing property (Haahtela & Kiiras 2014. p.47). The present price represents the amount that results when the technical decrease of value that results from depreciation of the property is subtracted from the new construction price. In their method Haahtela & Kiiras (2014) describe that the depreciation is calculated proportional to the new construction price and has standardized methods that are applied. The final construction price concept is the renovation price that represents the amount that is required for the predefined renovation actions take are estimated to take place on the property (Haahtela & Kiiras 2014, p.48).

The present price represents the depreciated construction cost, that is a more widely used term in the field, and it can be utilized to extract the building component of a developed but older property to value the land component. The last-mentioned method is at least in use in Australia where it is used to calculate the value of land for land taxation. (Boyd & Boyd 2012, p. 2). In this study the focus is on the valuation of residential land and not the valuation of development projects and for this reason the new construction price and the present price (or depreciated construction cost) are the most viable methods to construction cost estimation as renovation prices do not represent a property as a whole. (Haahtela & Kiiras 2014. pp.47-48)

Although the depreciated construction costs would probably be the best method for cost estimation for developed properties as they account for the decrease of value of the building component in relation to technical depreciation the study is limited to the examination of new construction (see chapter 1.4) that could take place on the property and this is used as a basis for land value. In this respect, although interesting as a method, the present price is not suitable for the aim of the study and the new construction price is better for this examination. Haahtela & Kiiras (2014) and Haahtela (2015) provide the means to analyze the new construction price that can be used in the residual valuation method to estimate the construction costs that are required as an input value for the model. The further description of how this method is used in this study is described in more detail in chapter 5.

4.1.3 Developer's profit

The final essential component that is required in for the residual valuation method is the developer's required return or required gross profit of the development project. This component is relatively hard to estimate as the developers are not exactly eager to provide this information, but it is also the result of developer heterogeneity, development project heterogeneity and the fact that there are no existing benchmarking practices for development schemes compared to for example real estate holdings (Crosby, Devaney & Wyatt 2018, p.2).

Development profit requirements have not been studied extensively or at least there is little evidence to be found in the Finnish context. There are some studies from other countries such as Crosby, Devaney & Wyatt (2018) and Curry (2013). Also, the effect of developer heterogeneity has been studied in this respect by Dong & Sing (2014).
Crosby, Devaney & Wyatt (2018) describe that there is extensive literature on the subject in terms of corporate finance that focuses on the methods that can be used to calculate the required return. These include but are not limited to: Discounted cash flow method (DCF), weighted average cost of capital (WACC) and capital asset pricing model (CAPM). DCF is often used reversely for this objective by determining the price of the development as well as the cash flows and using internal rate of return (IRR) to calculate the return of the investment. (Crosby, Devaney & Wyatt 2018, p.3).

This method however requires the total development price as an input value and therefore would require benchmarks from other projects if it is used in conjunction with the residual valuation method. The use of CAPM requires that the valuer can identify an expected market return rate and then estimate how sensitive the cash flows from the project will be to shifts in market return rate. (Crosby, Devaney & Wyatt 2018, p.3). The most significant problem in the estimation of the required returns is that each development scheme can be unique, and it results in the creation of a new asset for which there is no prior cash flows available. Geltner & Miller (2000) suggest that this could be addressed in several ways such as using the historical data from listed property development companies, real option pricing or a "reinterpreted" WACC. (Crosby, Devaney & Wyatt 2018, p.3)

In the context of the residual method the most common way of including the developer's profit is a cash sum proportional to either development costs or scheme value. Coleman et al. (2012) argue that this is somewhat inconsistent in terms of capital budgeting principles, but it can provide approximate outcomes if the valuer is able to adjust the figure according to the features of different development projects. (Crosby, Devaney & Wyatt 2018, p.4). The estimation of developer's profit as a proportion to development costs has some major problems to the estimation if the development costs are not certain (which they rarely are) and it can lead to huge differences from the method where the developer's profit is calculated proportional to the development project value. In their study the authors show that if the developer's profit is calculated as a percentage of the development cost, it can seem that some projects (that have an equal profit on cost) would be equally profitable although they have completely different profits on development value or IRR. (Crosby, Devaney & Wyatt 2018, p.7)

4.2 Theoretical questions considering residual land valuation

In this chapter we discuss some of the theoretical assumptions and implications of the residual valuation method. As we discussed in the previous chapter the residual land valuation method has been used both for the analysis of development projects as described by Pagourzi et al. (2003), Skarzyński (2006) and Greenhalgh & Bendel (2015) and in the valuation of land components of developed properties as described by Wolverton (1993) and Boyd & Boyd (2012).

From the perspective of this study it is important to understand if the method is applicable to the valuation of the market value of land. Although Wolverton (1993) and Boyd & Boyd (2012) described the use of the residual method in valuation of the land component of developed properties they did not fully describe the assumptions that support the use of this method in this kind of valuation. The method focuses on the idea development profitability, as the land value is calculated by subtracting total development costs from the end-product value as described by Pagourzi et al. (2003). The application of such a method that focuses on development profitability requires additional reasoning to support its use in the valuation of developed properties.

There are some questions that have been left unanswered by the literature that discusses the applications of this method. First if the method is used in the valuation of developed properties, do we have evidence that the land values are determined by the end-product prices as the residual method suggests? Second if this is true on the macroeconomic scale, are there factors in the local markets that affect this relationship? In this chapter we try to find answers to these questions from the more theoretical part of the literature that focus on the land market and the residential property market (often referred as the housing market). We hope that these studies can provide answers to the questions that the research focused on the practical applications leaves unanswered.

4.2.1 Relationship of land and property markets

There are a multiple of different urban models that describe the dynamics of land value. Some of these go back at least for a couple of centuries and the most widely known theories include those of Ricardo (1809), von Thünen (1826), Alonso (1964), Muth (1969), Mills (1972) and Fujita (1989). The classic monocentric urban model and the bid rent theory of land was described by Alonso (1964). There are some theoretical problems that limit the usefulness of these models to land valuation. A review of these problems was made by Özdilek (2011). In the study the author explains that classical urban models may help to understand land value patterns, but they do not provide satisfactory answers to land valuation. Özdilek (2011) criticizes the assumptions about spatial equilibrium, homogeneity and continuity that in his opinion are the main problems in the models. He also states that the biggest problem is that the value of the land parcel is lost under the buildings and merged with them as the capital known as "real estate". In the author's opinion every land valuation should focus on seven major questions that define the valuation: Type and use of land, time of valuation, location, valuation method, the importance of land value, the actors in the land market and the comprehensiveness of the explanations that result from the valuation. (Özdilek 2011. p.31)

In the Finnish context Oikarinen (2007) argues that the entire residential property can be divided into the components of land and structures. The price of the structure is typically measured as a replacement cost of the physical building considering depreciation and the price of the land is the market value of the location and the site characteristics. These add up to a sum total of the price of the property. (Oikarinen 2007). These studies suggest that every property can be divided to land and building components. However, they do not answer the question: is land value dependent on housing prices? One answer is provided by Ooi & Lee (2006) who analyzed the causal relationship of land and housing prices. In their article Ooi & Lee (2006) argue that it is not at all clear from the start which way the relationship goes. The neoclassical theory of land rent supports the view that high property prices are the result of high land prices whereas the rent theory proposed by Ricardo (1809) implies that high property prices lead to high land prices. (Ooi & Lee. 2006. p. 1). In their research Ooi & Lee (2006) find out that the residential property prices and the urban land prices are integrated in the long term which implies a relationship between the two. Further in the analysis the authors find out that the results modeled in an error-correction framework indicate that there is a Granger causality that runs from housing market to the land market and not the other way around. This empirical result supports the view of the Ricardian theory which claims that land price is dependent on property prices and not the other way around. (Ooi & Lee. 2006. p. 1-2)

There is some critique to this assumption regarding the Ricardian theory, since the theory assumes that the supply of land is completely fixed and there is only one use for land, agriculture. However, in the long term the supply of land is not fixed as land can be transformed to other uses for example from agriculture to housing. (Evans. 2004). Also, in one of the governing theories of property prices, the four-quadrant model, the supply of housing is not fixed as new construction can result in a higher supply of space that in turn shifts the market equilibrium toward lower property values. (Dipasquale & Wheaton. 1992)

The macroeconomic level cannot provide definite answers to this question as locations are not equal and some locations are preferred more than the others. These location specific factors where first modeled by Alonso (1964) and then continued by Muth (1969), Mills (1972 and Fujita (1989). The resulting modern neoclassical household's location theory is often referred to as Alonso-Muth-Mills theory. (Laakso. 1997. p. 14). Alonso-Muth-Mills theory suggest that the households prefer more centrally located areas and are willing to pay a premium for a central location (Laakso. 1997. p. 15). There are several other locational factors that the households consider as well, such as local services, quality of environment and social structure of the area (Laakso 1997. p. 28). This view is supported by some of the studies in the supply of housing. Oikarinen, Peltola & Valtonen (2015) studied the supply elasticity of housing in different regions in Finland and compared their results to similar studies in the US. Their results indicated that even though Finland is very sparsely populated and has an abundance of undeveloped vacant land the supply elasticity of housing follows quite closely the same dynamics than in the US. In both countries the supply elasticity (sensitiveness of supply to price changes) is significantly higher in large cities that already have high housing prices. (Oikarinen, Peltola, Valtonen. 2015. p. 28)

Even though the study by Oikarinen, Peltola & Valtonen (2015) was a regional study and did not focus in the dynamics between the districts in a city it is largely in line with some studies focused on micro location. Micro-locational factors were studied by Zahirovich-Herbert & Gibler (2014) who concluded that new construction only influences housing prices in a very limited spatial area within the city. In their study the authors compared a baseline hedonic model of housing prices considering panel data from existing properties with a model with spillover effects from new construction. The result was that new construction of similar sized houses from a half mile radius of the subject property had a negative effect on existing property prices. The construction of larger houses however had a positive effect on smaller house prices in the same area. (Zahirovich-Herbert & Gibler. 2014. p. 10). The results on the effect of infill development where somewhat inconclusive in the study, as infill development seemed to either lower or raise the prices depending on apartment sizes and had a very limited geographical effect radius (Zahirovich-Herbert & Gibler. 2014. p. 10). There are comparable results available from studies focused on Finnish cities. For example, Ahvenniemi et al. (2018) used a difference-in-difference hedonic regression to study the effects of infill development to existing property values in the Helsinki metropolitan area. In their study, the authors discovered that infill development had no statistically significant effect on the prices of existing properties in the different districts of the Helsinki metropolitan area (Ahvenniemi et al. 2018. p. 164).

4.3 Discussion of residual land valuation

In chapter 4 we described the residual valuation method in more detail by focusing on the different components that the method requires. Residual valuation method focuses in analyzing land values with the assumption that residential property prices can be divided into land price, construction costs and development profit (Pagourzi et al. 2003, p. 7). In multi-story apartment buildings, the residential property is not always traded as a whole but rather there is a submarket, housing market, where parts of this property are traded as apartments. If an apartment is valued with the residual valuation method, it requires the valuer either to divide the costs of the entire building to the apartments as well as the land parcel to each apartment or to calculate the sum total of the apartment prices that can be compared to the total construction costs and the land price. The residual valuation method focuses on the maximum capital expenditure, in the perspective of an investor, for buying a piece of land follows the definition of investment value (IVS. 2017 p. 22). There would have to be an assumption that the business logic and the perspective of value is shared by the market participants for it to reflect the market value (IVS 2017, pp. 18-19).

We know that there are other participants in the land market than property developers and they can have a significantly different business logic compared to the developers. This leads to uncertainties in the estimation of land market values using the residual method. Also, the planning controls can affect the possibilities that are available to developers and in turn lead to different land prices that the classical urban models suggest. For example, the bid rent theory of Alonso (1964) described a situation where the different land uses have differing utilities as factors of production in a given location and (without planning restrictions) the market price of the land is based on the highest bid that reflects the "best and most profitable use" of the select land parcel. This theoretical assumption is contradicted by zoning in cities that can restrict land to a specific use.

There are also some theoretical questions that relate to the application of the residual valuation method. If the method is used in the valuation of developed properties, do we have evidence that the land values are determined by the end-product prices? Also, if this is true on the macroeconomic scale, are there factors in the local markets that affect this relationship? There is evidence that the land prices are determined by housing prices at least in macroeconomic scale (Ooi & Lee. 2006. p. 1-2). These results are understandable at least when property developers are considered. The residual valuation method is used to calculate the purchase prices of land that the developers are willing to pay and if the asking price of land is higher than any developer would pay it would simply lead to no development. (Skarzyński 2006, p. 2)

When individual sites are compared in different regions or with different types of housing products, the locational and product-specific factors must be accounted for in the analysis as housing is a heterogeneous good and different sized apartments and different locations do not compete completely in the same market (Oikarinen, Peltola & Valtonen 2015) & (Zahirovich-Herbert & Gibler 2014). There are some theoretical implications to these factors. If we consider that some locations are preferable to others, then it might result that new construction does not necessarily lead to a decrease in all land value as it might not increase the supply of the most preferable locations that are already fully developed. These areas might in fact enjoy the monopolistic nature suggested in the Ricardian rent theory. This can lead to a situation where developers are most willing to initiate new construction in areas that already have minimal construction opportunities in the form of vacant land. According to the four-quadrant model few development opportunities would most likely lead to a smaller likelihood in the fall of housing prices because of new development. (Dipasquale & Wheaton 1992, p. 188). As the developer's profit requirement is based on the risk of the development this should lead to a situation where dense urban areas with less development opportunities are considered to have a lower risk for the developer. This hypothesis is tested in the empirical part of this research that starts in chapter 5. The developer's profit can be calculated either as proportional to the developed property value or as proportional to development costs. The profit proportional to the developed property value is preferred in the discussion presented by Crosby, Devaney & Wyatt (2018). The development timing however does matter on the value of the profits if the profits are discounted and this can only be considered by using IRR or DCF in the estimation (Crosby, Devaney & Wyatt 2018, p. 7).

In this study the focus is not in the analysis of a particular development project but in the analysis of land values regionally and for this reason the IRR or DCF does not bring any additional benefits compared to the profit proportional to the development value. The aim of this study is to analyze the land as it is most probably seen by the developers in general acting in the market and thus it should apply information that is accepted on the market by average rather than by any particular developer in any particular project. From the point of view the most viable method could be the analysis of historical data from listed property development companies as suggested by Geltner & Miller (2000).

5 Empirical research

In this chapter we describe the empirical analysis of this study. The empirical analysis was done using both correlation analysis and regression analysis. The data for the empirical part consists of apartment price data, construction cost data and vacant lot sales data. Other supporting data is also used to combine the datasets to the residual valuation formula. The formula and the supporting data is described in more detail in the following subchapters 5.1-5.3.

The aim of the empirical part is to find answers to the research questions:

1. Are there significant differences in predicted housing development profitability or in residual land values in different districts and apartment types in the Helsinki region, when the residual land valuation method is used in the estimation?

2. Is there a correlation between the residual valuation parameters (apartment prices, construction costs, vacant lot sales prices) and the valuation results (expected development profits & residual land values)?

3. Can the residual land valuation method be used to predict vacant lot sales prices in the Helsinki region?

This chapter is divided into three parts. The first part uses the literature findings of chapter 4 to establish a functional form of the residual valuation method. A functional form of the valuation method is vital so that we understand the model that is being tested with the data.

The second part focuses on the research data. First we use the established functional form of the residual valuation model to set the requirements for the research data that must be gathered for the research. The requirements are set by the conditions of the residual valuation method and are discussed through the findings of chapter 4 of the literature review. After the data requirements are set, we move forward to describe the actual research data that was gathered from multiple different sources for the purposes of the research. Here the statistical information of the data as well as the filtering of the data is described in more detail. At the end of the first part the problems and limitations of the research data are discussed in relation to the set data requirements.

In the third and final part of this chapter we describe the research methods that are used in order to answer the research questions. We start by describing the research situation and what we aim to achieve by analyzing the research data. From here we describe the chosen research methods for the analysis along with the justifications for the use of these methods. At the end of the third part we summarize what is being tested, what are the roles of the different datasets and briefly discuss the expectations for the results.

5.1 Residual valuation formula

In this part we describe the residual valuation formula that is being used in the empirical research. In chapter 4 we described in detail the application of the residual valuation method to land valuation as it is presented in the literature. Here the main components of the model were identified, and they are the end-product prices (apartment prices), construction costs and required development returns. The residual function combines these components to calculate a residue that can be used as a measurement for the highest possible price that the developer can pay for vacant land to ensure the required return for the development project (Pagourzi et al. 2003, p. 7). The residual valuation function can be expressed in the following form:

End-product (apartment) price - Total development costs - Required development return = Residual land value

This form can be used by the developer when all the components can be reliably estimated. This can be the case when the developer has a set profit requirement and is willing to pay no more than the amount for the vacant land that results in the fulfillment of the profit requirement. The calculation of this profit requirement can prove to be problematic when the function is used for research purposes as there is little public data for real estate development profit requirements and they can be actor- and project-specific at least to some extent. However, the mathematical form for the residual valuation function does not restrict the use of the function to the calculation of the residual land value specifically. It can be used to calculate the maximum acquisition price of a land parcel, the calculation of the expected profit from the development project and the calculation of the cost ceiling for the construction in the case that land has already been acquired (Greenhalgh & Bendel 2015, p. 3). If we use the function to calculate the expected development profit the other variables are transferred to independent variables and the function is used in the following form:

End-product (apartment) price - Total development costs - Lot price = Expected development profit

In this formula the estimate for the development profit can be calculated when there is enough data available for the end-product (apartment) prices, construction costs and vacant lot prices. This data should be more easily obtained as there is more public statistical data available for vacant lot sales prices than development profit requirements. This formula nevertheless presents another problem with the research situation: if the expected development profits can be calculated, then how can we test if these estimates represent the reality of real estate development? This requires observations for the development profits, either from expectations or the actual results of development projects.

As there are problems related to the available data using either of the functional forms, we have decided to conduct this research using both so that we can have more diversity in the results. First, we will use the second functional form of the residual valuation function to combine the gathered data from apartment prices, construction costs and vacant lot sales to calculate the expected development profits for each of the observation areas. Then we will make appropriate corrections to the assumptions of the model if needed and continue with a second analysis that is aimed in the estimation of residual land values. In the second analysis we will use the first presented functional form of the residual valuation method where housing prices, construction costs and required development profits are used to calculate the residual land values for the observation areas. The average estimated development returns are used as a representation of the required development returns. Although the data that is used in both analysis methods remains the same, this will allow us to see the results both as a representation of the expected development returns and as residual land values. This enables us to discuss the results together with a broader literature content and it can provide us with more insight on the benefits and problems of the residual valuation method.

5.2 Data requirements

In this chapter we will set the data requirements for the research according to the research situation and the residual valuation formula that was introduced in the previous chapter. The residual valuation method breaks down the valuation of real estate into the components of the end-product (housing) price, construction costs and developer's profit (Pagourzi et al. 2003, p. 7). This breakdown, however leaves a lot of room for interpretation as to how these components should be calculated. In this study these requirements are set to support the research purpose which is to evaluate the regional differences in land values and development returns using the residual valuation method and to test these applications with real world data considering vacant lot sales prices. This analysis requires multiple different datasets that represent the valuation components as well as the observations considering vacant lot sales. Next, we will go through the different sets of data that are needed for the research and set the requirements for the data that must be gathered.

The first dataset that is required is the housing price data. As this study is focused on the analysis of housing development the end-product price should reflect the price that can most likely be achieved from selling the developed residential real estate. This could be achieved by collecting data of residential property transactions where the entire residential property is sold. This kind of data is available for single-family houses that have a relatively large consumer market. However, this study is focused on multi-story apartment buildings for which there is fewer open data available considering property transactions and entire apartment buildings are traded less frequently. For apartment buildings it is more typical in Finland that the apartments are traded separately and there is a different market for individual apartments. The apartment market is a consumer market in which apartments are traded frequently and there is a very large amount of data available with representation among different districts and apartment types. It is important that the locational and product-specific factors are represented in the data as housing is a heterogeneous good and different sized apartments and different locations do not compete completely in the same market (Oikarinen, Peltola & Valtonen 2015) & (Zahirovich-Herbert & Gibler 2014).

Data considering apartment sales can be obtained from either public or private databases and the apartment market has the most frequent trading, so it should be possible to obtain data from specific apartment types that is confined to the observation areas. It is vital to gain enough statistical representation of the observation areas so that the results can be compared between the districts. Also, as the apartment types do not compete completely in the same market it is important to include the differences in the products and analyze these separately. These differences are most obvious considering the apartment size and the amount of rooms in the apartment. The data must include these factors as well as district-level locational information so that the analysis can be reliable. Other differences in the product must also be considered and if these cannot be included in the data it will lead into inaccuracies in the results. Also, the aim in this study is not to analyze any specific construction projects but the development profits and land values in the different districts on average. For this reason, the housing data should represent the average prices of apartments in the observation area rather than the absolute prices of individual apartments.

The second required dataset for the analysis is the construction cost data. This data should include both the costs for the construction as well as any other development costs that the developer must account for to develop the property. Ideally this data would include all the costs that are related to the development project and the costs related to the sales and marketing of the end-product.

The regional differences of construction costs of individual spaces (for example apartments) have been studied widely by Haahtela-kehitys Oy. Haahtela-kehitys Oy has published and maintains an index (Haahtela-index) that represents the regional differences in construction costs as well as statistics of the construction costs of individual space types (Haahtela & Kiiras 2014). However, there are some key limitations to this data. First the regional differences represented by the index represent the pricing of macro-level economic zones (such as Helsinki metropolitan area) and do not consider differences in the micro location (such as different districts). This is justified by the fact that there is little pricing difference in unit prices of materials and labor within the index areas (Teittinen 2019). However, we know that statistically there are differences in construction costs between different districts even within the same index areas that affect each construction project. The main causes of these differences are not in the unit prices of labor and materials but in the differences of the product that is being constructed. The differences in the product are the result of both market factors and external factors such as zoning regulations (Pennanen 2019).

The product differences that are caused by market factors are the direct result of the end-user's preferences that vary between target groups. These include, but are not limited to, the equipment level of the apartments, common spaces, car parking and architectural quality of the apartment building. Statistically these tend to correlate with the apartment price levels and new construction with higher priced apartments more often has for example more expensive equipment. The product differences can also be caused by external factors such as zoning regulations. In the inner-city areas of Helsinki, the zoning regulations are very different from those present in more peripheral neighborhoods and they can have a high impact on the minimum requirements of the product that is constructed. One such example is a parking requirement that sets a minimum of parking spaces that must be constructed in proportion to apartments. In inner city areas the parking needs to be mostly built underground and it has very high construction costs. When these parking spaces are not sold separately with a sum that represents the entirety of their cost, these additional costs must be distributed to the apartments as an overhead cost much in the same manner than the technical spaces and common spaces of the building. (Pennanen 2019)

As the end-product price data is planned to be gathered at the apartment level, the construction costs as well must be distributed to the individual apartments so that they can be analyzed together within the residual function formula. This requires the distribution of all construction costs to the apartments. This includes the cost of the apartment itself, but also the costs related to technical spaces, common spaces and all other functions and spaces that must be built along with the apartment building but cannot be sold separately from the apartments. Another option would be to construct the analysis on the property level and not on the apartment level. This would require that the observations of end-product prices are from similar apartment buildings with the same amount of common spaces, technical facilities and other functions. This approach could be useful if such data was available with a good statistical representation of the observation areas. As this kind of data is relatively hard to obtain and there is existing data available of the construction costs with apartment level distribution of total costs we have decided to use the existing data of apartment-level construction costs for the purposes of this research.

The third required dataset is the vacant lot sales data. There are two goals for the data. The first is to gain a representation of the price level of vacant lots in the different observation areas of the research. This information is used in the first part of our research as an independent variable in the residual valuation formula that is used to calculate the expected development returns for the different districts. For this purpose, it is most important to gain a statistical representation of the price level of vacant lots in a district that represents the most likely price that the developer must pay in order to obtain a suitable lot for the development. This price level should reflect market conditions to the best possible extent so that the expected development returns can be reliably estimated using the residual valuation method.

The second goal for the vacant lot sales data is to gain observations of individual lot sales that can be used as test data in the regression analysis that aims to test the capability of the residual valuation method to predict vacant lot sales prices. The same data can be used for the first and second purpose, but it will be used differently, and this leads to slightly different requirements. As the requirement for the first purpose was that the vacant lot sales data represents the price level of the observation areas on average, this is not required for the second purpose. When the residual valuation method is tested with the vacant lot sales data, each of the lot sales is considered individually using the linear least squares regression. For this purpose, it is equally important to have a broad statistical representation of the observation areas as the test results cannot be considered as reliable without sufficient test data.

In both purposes set for the vacant lot sales data it is important that the data represents actual vacant lot prices that fit the development purpose. For this reason, the data can only include lots that are designated for multi-story apartment buildings and have existing building rights in a detail plan. If the lot is zoned to another use or does not have building rights in the detail plan it does not represent a suitable lot for this kind of development. Also, it is important to exclude all transactions that are not actual sales, have special terms and conditions or are traded between a buyer and seller that have special relations (such as family relations). Also, any property scripts that are not developable properties according to zoning policies or lots that are too small for multi-story apartment buildings must be excluded so that the lot represents a viable situation that a developer may face when buying a lot for residential development.

The final required data component of the model is the price of the capital investment that reflects the value of capital and the risk related to the investment that the developer must consider in the valuation (Pagourzi et al. 2003, p. 7). This can be expressed as a required development return which can be calculated as a cash sum proportional to either development costs or scheme value (Coleman et al. 2012). The profit proportional to the developed property value is preferred especially when the construction costs are hard to estimate accurately (Crosby, Devaney & Wyatt 2018, p. 7). As there is little public data available for the profit expectations or requirements for development we would have to use other supporting data to fulfill this requirement in the research. This is a rather common problem as expressed by Geltner & Miller (2000), which we already discussed briefly in chapter 4. As the focus of this study is to address the use of the residual valuation method in housing project valuation problems by developers in general it should apply information that is accepted on the market by average rather than by any particular developer in any particular project. For this purpose, it was suggested by Geltner & Miller (2000) that historical financial data from listed property development companies could be a viable method to calculate a development profit that is common in the market. In Finland we have several listed property development companies that we can gather public financial data from and we can compare this data to the estimated development profits calculated with the residual valuation method.

The financial data from listed property development companies can be used, but it is not without its problems. This data is only available at the company level and it does not consider any differences in development profits regionally. If there are differences in the development profits, they cannot be seen from this macro-level data. For this reason, the financial data can only be used to evaluate if the profit estimates are in the same size-range with the actual results of the development companies, but it cannot be used to evaluate any regional differences. The financial data of the listed property development companies also has the problem that it is not project-specific but company specific on the macro-level. For this reason, there is very little to gain in the analysis by using this data as it is hard to say if it represents the organization more than the development projects. Another option here is to use the average estimates of the expected development returns that can be calculated with the residual valuation method by using the vacant lot sales data. This option will be considered in the actual analysis if there is no other viable data that can be used to represent the required development returns.

5.3 Research data

The gathered research data consists of three independent datasets that describe housing prices, vacant lot prices and construction costs. The housing price data was gathered from Oikotie.fi, which is one of the most popular services used for apartment sales advertisements in Finland. The data was limited to new construction only, as the older apartments are subject to depreciation and this can distort the results when the focus of the study is in new construction. Also, the data was limited to apartments where the building is constructed on an owned lot and any apartment buildings that were constructed on a leased lot were excluded from the data. The reason for this limitation is that the lot can be a large part of property value and the legal status of lot ownership can have a significant effect on the apartment prices. The residual valuation method also includes the assumption that the lot is part of the property that is being valued with the method. A more detailed description of the limitations considering the housing price data is provided in chapter 5.1.1 that discusses the housing price data more thoroughly.

The vacant lot sales data was gathered from the database of the National Land Survey of Finland (NLS). NLS has a right by law to gather information of all property transactions in Finland and as a public office their data is presumably the most reliable available source of this information. The vacant lot sales data was limited to registered building sites with residential use only. Any property scripts and non-residential properties were excluded so that the data better represents the assumptions that need to be made to use the residual valuation method. These limitations and other limitations considering the vacant lot sales data are further described in chapter 5.1.2 where the gathered data is discussed in more detail.

The construction cost data was gathered from the database of Haahtela-kehitys Oy that is the leading private company in Finland that focuses on construction cost analysis research and software applications. The reason that the data is gathered from this source is that the empirical analysis requires explicit data of the total construction costs of different types of apartments in the different areas in the Helsinki region. The data must take into account not only the costs related to the apartment itself, but also all the costs that relate to the common spaces, technical facilities and structures of the entire apartment building. Also, the differences in the product as well as the price level of construction in the different districts in the greater Helsinki area must be considered. There is no publicly available data that meets these requirements, but Haahtela-kehitys Oy has agreed to provide this data to be used in the research. The data was limited to residential multi-story apartment buildings and any other building types were excluded. These limitations are vital so that the construction cost data is compatible with the other datasets in the context of residual valuation. The construction cost data is further described in chapter 5.1.3.

The research data was gathered from different districts in the municipalities that make up the greater Helsinki metropolitan area (HMA). There are 14 municipalities that make up the larger Helsinki metropolitan area: Helsinki, Espoo, Vantaa, Kauniainen, Hyvinkää, Järvenpää, Kerava, Kirkkonummi, Nurmijärvi, Sipoo, Tuusula, Vihti, Mäntsälä and Pornainen. Further in the texts these are referred simply as Helsinki metropolitan area (HMA). These municipalities are further divided into sub-areas or districts. One way to divide them into districts is the national postal code area system. Postal code areas described in the sales advertisements in the data provided by Oikotie.fi and they can be easily linked to the vacant lot sales data provided by the NLS using real estate register codes. For these reasons we will use the postal code areas as a base for the division of areas.

As we are using three different datasets, the problem with data availability is amplified. The areas that were used in this study had to have data considering housing prices, vacant lot sales and construction costs. As there was not enough data for all the areas in the HMA, the analysis had to be limited to the municipalities of Helsinki, Espoo, Vantaa, Järvenpää and Kerava.

5.3.1 Housing price data

The gathered housing price data consists of the sales advertisements of newly constructed apartments that were posted to the service between January 2018 and November 2018 in the Helsinki metropolitan area. After the initial gathering of the data the data was filtered and only the sales advertisements that fulfilled the filtering criteria were accepted. There were two main criteria for the filtering of the housing price data. These criteria were aimed to include only the sales advertisements that can be reliably combined with the other datasets of this study, the vacant lot sales data and the construction cost data. First the apartment had to be new construction and the advertisements of old apartments were not accepted. The reason for these criteria is that old apartments are subjects to depreciation that can affect the price of the apartment. The construction cost data is also modeled to represent new construction and for this reason newly constructed apartments are more likely compatible with the construction cost data. Also, as the focus of this study is on development profitability, it is more correct to use new construction as it is more likely to represent the possible development opportunities in the area.

The second filtering criteria is that the apartment must be constructed on a lot that is owned by the apartment complex and thus is a part of the apartment price. The residual method assumes the lot to be a part of the owned property and a leased lot would violate the assumption made by the residual model and lead to problems with the empirical analysis. For these reasons any sales advertisements where the apartment building was constructed on a leased lot were excluded from the research data. Finally, there was no data considering larger apartments (more than 3 rooms) from many of the areas and a choice had to be made between analyzing all the different apartment types and the different municipalities and districts. As the focus of the study is in analyzing regional variation of housing prices and development profitability the larger apartments were excluded from the study to keep a broader representation of the different districts in the Helsinki region. Due to this limitation the data represents only 1-, 2- and 3-room apartments. The filtered housing price data consisted of 1,393 sales advertisements in total. The distribution of the advertisements by apartment type were the following: 625 advertisements were from single room apartments, 485 from 2-room apartments, 228 from 3-room apartments and 55 from 4-room apartments or larger. The number of sales advertisements for all the districts and apartment types are represented in table 2.

Table 2. Number of sales advertisements for different apartment types and districts

City	District	1 -room	2 -room	3 -room	4 -room	All
Helsinki	Jätkäsaari	30	45	29	4	108
Helsinki	Kalasatama	3	13	2	0	18
Vantaa	Viertola	106	74	38	9	227
Vantaa	Martinlaakso	74	44	35	3	156
Vantaa	Kivistö	158	68	37	5	268
Espoo	Niittykumpu	89	62	39	21	211
Espoo	Eestinlaakso	53	102	31	7	193
Espoo	Saunalahti	103	48	3	0	154
Kerava	Kerava keskus	7	24	10	6	47
Järvenpää	Järvenpää keskus	2	5	4	0	11
Total		625	485	228	55	1,393

The average asking price of the advertisements for all apartment types was 241,773 € with the standard deviation of 99,613 €. Single-room apartments had the average of 183,489 € with the standard deviation of 38,060 €, 2-room apartments had the average of 242,990 € with the standard deviation of 63,110 € and 3-room apartments had the average of 344,680 € with the standard deviation of 112,122 €. These averages give a good understanding of the average price levels in the Helsinki region, but it is important to understand that the prices varied significantly between different areas and apartment types. Apartment types also had variation in the floor area of the apartments. The variation of floor areas in different apartment of the same apartment type may be one cause for variation especially as there is a very high correlation of 0.81 between the apartment size and the asking price. One way to approach this issue is to use the price per floor area (€/m2) of the apartments in the analysis. The correlation between the price per floor area and the apartment size is -0.36 which implies that the price per floor area is the highest in smaller apartments and lower in larger apartments.

To get a more in-depth understanding of the price variations we continue to analyze the variation of prices per floor area (€/m2) between regions separately for each apartment type. Single-room apartments had the highest average price per floor area of 6,431 €/m2 and the lowest standard deviation of 1,193 €/m2, 2- room apartments had the average of 5,466 €/m2 with the standard deviation of 1,223 €/m2 and the 3-room apartments had the lowest average of 5,136 €/m2 with the highest standard deviation of 1,305 €/m2. The standard deviation is the lowest with the higher average price per floor area.

Regionally the highest prices per floor area of single and 3-room apartments were in Jätkäsaari, Helsinki and highest prices per floor area for 2-room apartments were in Kalasatama, Helsinki. The lowest prices per floor area of all apartment types were in Kivistö, Vantaa. The highest, lowest and average prices of all the areas are found in table 3.

Table 3, Apartment asking prices (€/m2)

City	District	Apartment type (rooms)	Average price (€ / net area)	Lowest price (€ / net area)	Highest price (€ / net area)	Standard deviation (€ / net area)
Helsinki	Jätkäsaari	1	9,838	7,845	11,632	1,042
		2	7,508	6,637	8,976	658
		3	7,548	6626	8,991	719
Helsinki	Kalasatama	1	10,414	10,262	10,594	168
		2	8,584	7,533	10,053	798
		3	7,716	7,454	7,977	370
Vantaa	Viertola	1	6,591	4,860	7,850	641
		2	5,423	4,263	6,624	510
		3	5,226	4,287	6,231	440
Vantaa	Martinlaakso	1	5,346	4,763	6,396	475
		2	5,362	4,377	6,089	474
		3	4,562	3,829	5,770	462
Vantaa	Kivistö	1	5,874	3,954	7,560	787
		2	4,334	3,318	5,731	488
		3	3,938	3,264	4,968	371
Espoo	Niittykumpu	1	7,250	6,469	8,483	436
		2	6,674	5,724	7,998	458
		3	6,099	5,313	7,712	535
Espoo	Eestinlaakso	1	6,027	5,000	7,446	781
		2	4,462	4,053	5,755	401
		3	4,005	3,801	4,782	222
Espoo	Saunalahti	1	6,405	5,461	7,293	566
		2	5,510	4,750	6,132	442
		3	4,599	4,359	4,851	246
Kerava	Kerava keskus	1	5,107	4,698	5,515	293
		2	4,650	4,295	5,144	208
		3	4,107	3,830	4,527	215
Järvenpää	Järvenpää keskus	1	5,377	5,154	5,600	315
		2	4,866	4,563	5,188	256
		3	3,968	3,832	4,056	95

5.3.2 Vacant lot sales data

The vacant lot sales data consists of individual property transactions that were gathered from the Helsinki metropolitan area (HMA) from the sales that took place between 2016 and 2018. After the initial gathering of the data the data was filtered and only the transactions that fulfilled the criteria described in Table 4 were accepted. These criteria are aimed to include only the transactions that relate to the focus of this study and where the statistical differences between the lots and the reliability of the transactions can be reasonably controlled.

Table 4: Filtering criteria for vacant lot sales data

Criteria considering the sold lot:
The lot has an existing detail plan and is designated in the plan as a multi-story apartment building
The lot is entirely in residential use
The lot has a building permits for residential use
The lot is a registered building site and not a property script
The lot has at least 500 square meters of lot area
The lot does not include shoreline and is not connected to bodies of water
Criteria considering the transaction:
The transaction does not include buildings or other commodities except land
The transaction is a true property sale and not any other form of transaction
The transaction parties are private owners or municipalities. The owner and the seller must not have a family relation
Other criteria
Only the districts that had 3 or more transactions were accepted

The criteria considering the lot limits the traded lots to those that have an existing detail plan. The reason for this limitation is that the detail plan grants building rights to a lot and all the lots that are analyzed must have building rights so that we can use them as a reference in the residual land valuation method where the construction of new buildings is presumed. The lots also had to be restricted to residential use only, as other uses can affect the land value and these other uses are excluded from this study. Only the lots that are registered building sites were included and property scripts were excluded from the transactions. This limitation results from the fact that property scripts are not valid building sites before they are registered. The registration includes additional costs and the building rights of the property script cannot be reliably calculated from this data. Some of the lots were very small compared to the average sold lots. This may have severely affected the price and to better control the statistical aspects of the data the lots that have an area of less than 500 square meters were excluded. Some of the lots also had direct access to shoreline or were connected to bodies of water. There are studies for example by Peltola & Väänänen (2006) that show that connection to bodies of water can affect the price of the lot, and for this reason these few transactions were excluded so that the data better represents the average vacant lots in the areas.

The final limitation to the data was the exclusion of areas that did not have at least 3 vacant lot sales in the same district. Without statistically sufficient data the areas could not be included to this study as the results would have been unreliable. Due to data availability, the observation areas were previously limited to the municipalities of Espoo, Helsinki, Vantaa, Järvenpää and Kerava. From these municipalities, only some of the districts had the enough sales data of at least 3 vacant lot sales. These districts and the vacant lot sales amounts are described in table 5. In total there were 57 vacant lot sales in the greater Helsinki area that fit the criteria described above.

Table 5: Distribution of accepted vacant lot transactions

City	District	Transactions
Helsinki	Jätkäsaari	8
Helsinki	Kalasatama	8
Vantaa	Viertola	3
Vantaa	Martinlaakso	7
Vantaa	Kivistö	13
Espoo	Niittykumpu	3
Espoo	Eestinlaakso	3
Espoo	Saunalahti	6
Kerava	Kerava keskus	3
Järvenpää	Järvenpää keskus	3
Total		57

The sales prices of the properties had substantial variation between the areas with the highest price overall being 12,342,759 € and the lowest 510,625 €. The highest average sales price of 5,401,000 € was in Niittykumpu, Espoo and the lowest average sales price of 463,750 € was in Kerava keskus, Kerava. The standard deviations between areas varied considerably from the highest (3,125,042 €) being in Jätkäsaari, Helsinki and the lowest (27,063 €) being in Kerava keskus, Kerava. The standard deviations in the sales prices however are not very reliable in predicting the actual deviation of price levels in the areas. This is because the lot areas and building rights also varied considerably and the comparison of very different sized lots and lots with very different amounts of building rights can be misleading. For this reason, the further analysis of the sales prices requires the analysis of prices per lot area and prices per building permit amount. These figures can better represent the deviation of the prices. The highest and lowest sales prices as well as the averages of all areas are listed in table 6.

Table 6: Lot sales prices (€)

City	District	Average (€)	Lowest (€)	Highest (€)	Standard deviation (€)
Helsinki	Jätkäsaari	4,033,700	7,075,425	12,342,759	3,125,042
Helsinki	Kalasatama	2,663,277	4,772,731	7,518,400	1,424,826
Vantaa	Viertola	1,227,655	1,590,795	1,986,050	380,216
Vantaa	Martinlaakso	713,765	1,245,129	1,537,640	338,300
Vantaa	Kivistö	895,488	1,880,624	4,922,500	1,094,821
Espoo	Niittykumpu	5,401,000	8,898,500	11,200,000	3,078,947
Espoo	Eestinlaakso	1,751,762	2,537,493	3,126,891	708,276
Espoo	Saunalahti	760,000	2,031,517	3,325,100	1,082,164
Kerava	Kerava keskus	463,750	495,000	510,625	27,063
Järvenpää	Järvenpää keskus	642,000	1,022,522	1,564,000	481,613

The price per lot area is a measurement of price that takes into account the lot size. It is fair to assume that the lot size does matter when considering the sales price of the property and this is confirmed by the data. There are significant differences in the prices per lot area with the highest average price per area of 5,299 €/m2 being in Jätkäsaari, Helsinki while the lowest average of 133 €/m2 was in Kerava keskus, Kerava. The standard deviation of the prices per lot area in the entire dataset was 953 €/m2 with the highest standard deviation of 4,242 €/m2 being in Jätkäsaari, Helsinki and the lowest of 16 €/m2 being in Kerava keskus, Kerava. The deviation of prices between areas and the standard deviation of prices within areas are both significant. The standard deviation of price per lot area correlates strongly with the average price per lot area. The standard deviation is highest in areas with the high prices per lot area. The differences between areas are explained in part by the differing amount of building rights per lot area with Jätkäsaari having the average of 3.93 efficiency rate compared to 0.71 in Kerava keskus. The prices per lot area are listed for each area in table 7. To further understand this relationship, we must also study the prices per gross building area which better considers the differences in building intensity between the areas.

Table 7: Lot sales price per lot area (€/m2)

City	District	Average (€ / lot area)	Lowest (€ / lot area)	Highest (€ / lot area)	Standard deviation (€ / lot area)
Helsinki	Jätkäsaari	2,114	5,299	13,117	4,242
Helsinki	Kalasatama	1,346	3,288	5,515	1,374
Vantaa	Viertola	491	695	829	180
Vantaa	Martinlaakso	651	1,287	1,620	406
Vantaa	Kivistö	396	915	2,349	586
Espoo	Niittykumpu	2,228	2,502	2,749	262
Espoo	Eestinlaakso	410	534	614	108
Espoo	Saunalahti	304	491	863	194
Kerava	Kerava keskus	133	151	161	16
Järvenpää	Järvenpää keskus	169	310	418	128

As the building rights and lot sizes vary between areas and transactions, the differences between areas can be measured with the price per gross building area (€/k-m2). The total highest price per gross area of 2,040 €/k-m2 was in a transaction from Kalasatama, Helsinki with the second highest of 1,991 €/k-m2 being in Jätkäsaari, Helsinki. The highest average price per gross area of 1,570 € however, was in Niittykumpu, Espoo. The second and third highest averages were in Jätkäsaari and Kalasatama. The standard deviation of the price per gross area was a 51 €/k-m2, which is a lot lower than the standard deviation in prices per lot area (953 €/m2). This supports the assumption that building rights have a significant effect on lot sales prices. The standard deviation in prices per gross area varied significantly between areas with the highest standard deviation of 512 €/k-m2 being in Kalasatama, Helsinki and the lowest of 5 €/k-m2 being in Järvenpää keskus, Järvenpää. The deviation in the price per gross area correlates heavily with the average price level of the area. The areas with the highest prices per gross area also had the highest standard deviation in the prices. The highest, lowest and average prices per gross building area are represented in table 8.

Table 8: Lot price per gross building area (€/m2)

City	District	Average (€ / gross area)	Lowest (€ / gross area)	Highest (€ / gross area)	Standard deviation (€ / gross area)
Helsinki	Jätkäsaari	961	1,223	1,991	338
Helsinki	Kalasatama	482	1,141	2,040	458
Vantaa	Viertola	400	503	558	90
Vantaa	Martinlaakso	395	565	695	138
Vantaa	Kivistö	318	412	550	63
Espoo	Niittykumpu	1,358	1,570	1,800	222
Espoo	Eestinlaakso	552	570	604	30
Espoo	Saunalahti	369	592	914	222
Kerava	Kerava keskus	206	212	215	5
Järvenpää	Järvenpää keskus	135	309	460	164

5.3.3 Construction cost data

The construction cost data was provided by Haahtela-kehitys Oy that is the leading private company in Finland that focuses on construction cost analysis research and software applications. The data was gathered at Haahtela-kehitys Oy by their leading cost expert, Erkki Teittinen. The data represents the knowledge acquired by the collective effort of Haahtela-kehitys Oy and relies on the expert opinion of Teittinen (2019). The expert opinion is based on TAKU® cost analysis model, reference projects and construction market research conducted in the year 2018.

TAKU® (Haahtela-kehitys Oy) is an information model meant for the budgeting and financial control of construction projects during their conception, planning and execution. It is based on a heuristic model that uses the input information of spaces and space attributes and other information that is available for the user in the pre-planning phase of development. These include, but are not limited to: space types, space sizes, floor height, lighting requirements, technical requirements and construction regulations. In the planning phase of development, TAKU® -model utilizes the actual design and technical plans of the building to model building elements that are priced with unit prices based on construction market research. The model also includes the costs related to the construction site, planning and project management. TAKU® -model is a market-tested tool that represents the cost differences of different kind of construction projects and different kinds of spaces and Haahtela-kehitys Oy is committed to continuous testing of the model as well as the market research required to uphold the relevant cost information related to the pricing. This testing and research are conducted with actual reference projects that are priced using the model. The results of the model are compared to the actual historical costs of the projects at a detailed level. The test results are used to calibrate the model including the calibration in the annual price level and regional price levels. The testing is conducted in six regional zones in Finland. These results are also published publicly as an index /Haahtela - index). TAKU® -model is the market leader in Finland as an information model for budgeting and property development management both on the public and private sector.

The reason that the data is gathered from this source is that the empirical analysis requires explicit data of the total construction costs of different types of apartments in the different areas in the Helsinki region. The data must consider not only the costs related to the apartment itself, but also all the costs that relate to the common spaces, technical facilities and structures of the entire apartment building. Also, the differences in the product as well as the price level of construction in the different districts in the greater Helsinki area must be considered. There is no publicly available data that meets these requirements, but Haahtela-kehitys Oy is committed to provide this data to be used in this research. The construction cost data is further described in chapter 5.1.3.

The construction cost data provided by Haahtela-kehitys Oy is a list of regional averages of the new construction costs related to different apartment types. These apartment types include 1-, 2- and 3-room apartments that were chosen previously as the focus of the empirical analysis. Unlike other available data these averages include not only the price of the apartment but also all the overhead costs that are distributed to the apartments from other spaces in the building such as common spaces, technical spaces, stairways, hallways and entrances. These spaces as well as the apartments are modeled to represent the product differences that are caused by both the market factors and external factors of the areas and they represent the best available understanding of regional product differences and price level differences that are available with current research (Teittinen 2019). The data was limited to residential multi-story apartment buildings and any other building types were excluded. These limitations are vital so that the construction cost data is compatible with the other datasets in the context of residual valuation. The estimated construction costs are represented in Table 9 by index area and by observation district.

Table 9: Estimated new construction costs based on Haahtela-index areas

City	District	Index area (Haahtela-index)	Apartment type (rooms)	New construction costs (€ / net area)
Helsinki	Jätkäsaari	1	1	3,464
			2	3,193
			3	3,168
Helsinki	Kalasatama	1	1	3,464
			2	3,193
			3	3,168
Vantaa	Viertola	1	1	3,464
			2	3,193
			3	3,168
Vantaa	Martinlaakso	1	1	3,464
			2	3,193
			3	3,168
Vantaa	Kivistö	1	1	3,464
			2	3,193
			3	3,168
Espoo	Niittykumpu	1	1	3,464
			2	3,193
			3	3,168
Espoo	Eestinlaakso	1	1	3,464
			2	3,193
			3	3,168
Espoo	Saunalahti	1	1	3,464
			2	3,193
			3	3,168
Kerava	Kerava keskus	2	1	2,801
			2	2,643
			3	2,581
Järvenpää	Järvenpää keskus	2	1	2,801
			2	2,643
			3	2,581
Average				3,155

5.3.4 Problems with the research data

In chapter 5.2 we introduced the requirements for the research data that would be needed to conduct this research accurately. These requirements set out the criteria for the data and if all the criteria are met the data fits the research well and we can expect the results to be more reliable than if the data violates some of the criteria. Optimally the data should fit both the more structural and theoretical criteria that are defined by the residual valuation method, but it should also have a good statistical representation. Unfortunately, not all the data meets all these criteria, and this leads to problems with the reliability of the results. As there are several different datasets that are used in this research we will analyze the problems related to each dataset separately. After analyzing the problems related to each separate dataset we will shortly discuss the problems related to their compilation in the residual valuation formula. These problems are come from the different formats of the datasets and some adjustments had to be made to use all the datasets together in the research.

The housing price data was gathered from the sales advertisements at Oikotie.fi. The data meets the requirements for statistical representation quite well in most areas as there were a total of 1393 sales advertisements and most of the observation areas had a relatively high number of advertisements (over 100). However, there were some problem areas such as Järvenpää keskus, which only had 11 sales advertisements and the district of Kalasatama had only 18 sales advertisements. These numbers would present less of a problem with statistical representation if they were from the same apartment type (such as 1-room apartments), but when the apartment types are considered there are even more problems with these areas. Järvenpää keskus only has 2 sales advertisements from 1-room apartments and Kalasatama has only 2 advertisements of 3-room apartments. This lack of statistical representation can lead to inaccuracies in the analysis. Although this is a problem, we decided to take these areas into the analysis as they are particularly interesting development areas, but it is important to remember these problems when the results are analyzed.

4-room and larger apartments lacked statistical representation in almost all the areas and for this reason the larger apartments had to be excluded from the analysis which makes the analysis narrower and less representative of actual apartment buildings that often have larger apartments as well. This presents a problem when the results are reviewed in relation to actual development possibilities that the developers face in these areas. As there is a lack of data of larger apartments it is harder to make accurate estimates related to the residual land value of any of these areas. Although the housing price data meets the requirements for statistical representation in most of the areas and apartment types it presents another problem with reliability as it consists of sales advertisements and not actual apartment sales. The use of sales advertisements is not an optimal solution. Sales advertisements do not represent the actual historical prices of apartments but instead represent the seller's expectations. This can lead to inaccuracies if the apartments are not actually sold at the listed price and it is hard to estimate the possible discounts that the seller might offer for the buyer at the actual time of purchase. This presents a problem with data reliability when it is used to measure the expected price that the developer can obtain from a development project.

The land prices are represented in the analysis by the vacant lot sales data provided by the National Land Survey of Finland (NLS). The data meets the requirements set for land price data quite well as it is based on actual vacant lot sales that have taken place during the years 2016-2018. Most of the problems that relate to the vacant lot sales data were already controlled by filtering the data that was gathered. This filtering was aimed to exclude any lot sales that do not reliably represent vacant lots that are zoned for residential use. Although the filtering of the data solved some of the problems as it led to the exclusion of properties that were not designated for multi-story apartment buildings, were already developed or had special terms or conditions, it caused problems with the statistical representation of the observation areas. Altogether only 104 vacant lot sales passed the filtering criteria, and this led to a very narrow sample in some of the observation areas. At best the individual districts had no more than 10 lot sales and the worst seven districts had only 3 lot sales per district.

This presents a significant problem with the statistical representation of the observation areas. Even though these lot sales are filtered to represent the residential vacant lots well, when there are only a handful of observations the results can be somewhat random. Individual sales may or may not represent the actual price level of lots in the area and with such a narrow sample it is hard to say if this data represents the price level well enough for the empirical analysis. These lot sales may represent only some situations and individual locations in the district and this can distort the results. It could be possible to obtain a larger sample by widening the time-range of the sales from 2016-2018 to include lot sales from earlier years as well. This would however cause another problem related to the price changes that may have occurred during that period. On a time-period of 2 years it is fair to assume that the changes in lot prices are relatively small but the longer the period the higher the chance is that the price changes are significant. As there are no public indexes that track price levels of lots on a district level that could be used to account for these changes it could lead to even higher inaccuracies when the differences of lot prices are compared between the districts. All things considered it must be noted that the data has a major problem with statistical representation that cannot be easily fixed for this research. These problems that relate to the vacant lot price data must be considered when the results are analyzed.

When it comes to the construction cost data the gathered research data fits the set requirements well when it comes to product-level differences and the level of detail, but it lacks a broad statistical representation as it is based on reference projects and expert analysis instead of a large statistical sample. A large statistical sample could have been obtained from the unit prices of apartments, but this data does not include the differences in product between the districts and only considers the regional unit price levels of materials and labor. This kind of data is much more statistically accurate but lacks the information needed to differentiate the construction cost differences between the different districts which is essential for this study. As there was no data available that fits all the criteria and the construction cost differences between districts was considered essential, the use of more detailed data instead of a broad statistical sample was considered a better option. However, it must be noted that the data and its capability to represent the regional differences of construction costs relies heavily on the expert opinion of Teittinen (2019) and the status of Haahtela-kehitys Oy as the leading expert of cost analysis in the industry. In terms of academic independence, it would have been preferable to use a broader statistical sample of reference projects if such data were available.

5.4 Research methods

The purpose of this research is to evaluate if the residual valuation method is a useful tool for the estimation of the differences in expected development returns and land values of different districts in the HMA for residential development. The research is done in two parts that have different roles in this analysis and use different research methods.

The first part of the research uses correlation analysis to study the inherent limitations and possible problems that relate to the residual valuation method. The aim here is to test all the used datasets for correlation with each other as well as with the estimates. The reason for this analysis is that in the second part of the empirical research we use several different datasets in a pre-established formula and this requires some quite heavy assumptions that must be made to conduct the analysis. As there are a lot of assumptions and some problems with the statistical representation of the data it is important to understand the problems that possibly misguided assumptions can cause when the model is used. Correlation analysis enables us to better understand which datasets have a relation with each other and how high is the correlation of each independent dataset and the residual valuation estimate. The different factors that are tested in the correlation analysis are: apartment prices, construction costs and vacant lot sales prices along with the estimated development returns and land residual values. This analysis enables us to critically look at the estimates in relation to the data they use and provide insights to the second part of the research.

The second part of the research uses regression analysis to test how well the residual land value estimates can predict individual vacant lot sales prices. The focus here is to test the pre-established residual valuation model that is based on the literature and find out if the model is useful for the estimation of vacant lot sales prices. As the literature findings give a well-established functional form for the residual valuation method that can be tested we focus on testing this form rather than speculate with other possibilities and combinations. The residual valuation form is not based on independent factors that influence the value "ceteris paribus" but on the subtraction of costs from the estimated end-product sales price and thus it cannot be classified as a hedonic model. As there are no independently acting factors in the formula the regression model is quite simple, consisting only of the residual land value as a predictor. The use of ordinary least squares regression however requires some conditions from the data, such as an assumption of a linear relationship between the predictor and the dependent value. These conditions must be first tested so that we know which method of regression can be used in the analysis. There can be other explanatory models that use the same data such as a model that uses apartment prices alone as a predictor. Since the residual land valuation method uses apartment price data but adds other complexity (subtraction of construction costs) it should either lead to better results or otherwise its use is poorly justified in the estimation of vacant lot sales prices. For this reason, we will also test the apartment prices alone as a predictor so that we can compare the results between a model with apartment price data and the residual land valuation model. A hedonic model could be established using the datasets and other possible price data as independents, but since this study is focused on testing the existing residual valuation framework we find the two described models as adequate for this purpose. Other opportunities for price models using the established data are left for recommendations for further study that are discussed in more detail in chapter 8.

5.4.1 Correlation analysis

In research, correlation analysis is used to measure the connection between two variables. This connection can be measured by using the Pearson correlation, which is a number between -1 and 1 that indicates if the two variables are linearly related and to what extent. A Pearson correlation of 1 is called a perfect positive linear correlation and indicates that the two variables move completely in the same direction. A Pearson correlation of -1 on the other hand is called a perfect negative correlation and indicates that the two variables move completely to opposite directions. A correlation of 0 indicates that there is no linear connection between the datasets and their movements are not connected to each other, at least with a linear connection. Even with a correlation of 0, there is a possibility that the variables share a non-linear connection that is not visible with Pearson correlation. With real statistical data the correlation is often somewhere between -1 and 0 or between 0 and 1. In these cases it is important to consider the context of the data and the statistical representation of the data when the results are analyzed. (Metsämuuronen 2010, s. 370).

The statistical representation of the data is important to the analysis at least for two reasons. First the sample size of the data can have a major effect on the statistical significance of the results. When sample sizes are low, a high value for the Pearson correlation does not necessarily mean that there is an actual real-world connection of this magnitude (Metsämuuronen 2010, s. 370). A second factor to be considered is that a measured correlation does not implicate a causal relationship between the two variables. Especially with phenomenon as complex as housing economics the data is heterogeneous by default and there might be other factors that influence the data which can lead to a situation where correlation appears, but the two variables do not share a systemic or causal relationship. This is a problem that cannot be accounted for in the correlation analysis, but it must be kept in mind when the results are analyzed. In this research, correlation analysis is used to answer the second research question: "Is there a correlation between the residual valuation parameters (apartment prices, construction costs, vacant lot sales prices) and the valuation results (expected development returns & residual land values)?"

This is done by calculating the Pearson correlation between all the datasets that were gathered for this research as well as with the results of the residual valuation. As the data was gathered from different sources and initially has different accuracy levels (for example individual apartments, or vacant lots) this analysis is done with the compiled data that represents the district-level values. The district-level compilation puts the datasets into the same level of accuracy and enables the comparison of same level data for each of the variables.

5.4.2 Regression analysis

Regression analysis is a statistical tool that is used in research to estimate relationships among variables. In regression the focus is on the relationship of a dependent variable and one or more independent variables (predictors). Regression analysis aims to understand how the typical dependent variable changes when the one of the independent variables is changed, while the other independent variables stay fixed. If the analysis contains more than one independent variable it is called multiple regression analysis. In these situations, where many independent variables are used it requires that the variables are truly independent and changing one of them does not affect the others. This condition can be violated by multicollinearity or autocorrelation that both can be tested statistically. (Pejic et al. 2013. pp. 43-44)

Regression analysis is a collection of different methods that can be used to estimate different kinds of relationships between variables, but the most basic method is the ordinary least squares linear regression. This method is used to establish linear estimators between an independent variable and a dependent variable that share a linear relationship. (Wilson 2012. pp. 23-24). There are some requirements for this method and it cannot be used in all cases. First the relationship between the independent and dependent value must be linear and it cannot be used with nonlinear relationships. A second important condition is that the data must be normally distributed. A third preferable condition is that the standard deviation of the probability distributions is the same for all values of the independent variable. If the condition of equal variances is fulfilled, the data is called homoscedastic and otherwise heteroscedastic. (Wilson 2012. pp. 30-31). In the analysis we will prefer the ordinary least squares regression analysis if the data meets these requirements. If the data is in contradiction to some of these requirements we will have to use transformations or other regression methods in the analysis. These aspects are discussed further in chapter 6 where we describe the actual regression models that were used and the justification for their use.

Regression analysis is used in this research to calculate the difference between the residual valuation estimates and real-world observations. At this point we have used the residual land valuation method to estimate the expected development returns and the residual land values for each of the observation districts. However, there are no real-world observations for the expected development returns that we could use in a regression analysis to test our estimates. This is unfortunate for the research, but we simply cannot use regression analysis without enough data in this case. The estimated residual land values however can be tested using the vacant lot sales price data and this is our focus in the regression analysis. We will test the residual land value as a predictor to the vacant lot sales prices and in comparison, we will also test the apartment price data as a predictor to vacant lot sales prices. Our hypothesis is that the residual land value is a better predictor, since it uses systemic information of both apartment prices and construction costs to estimate the land value. If the apartment price data alone is a better predictor to the vacant lot sales prices, then we must conclude that the residual valuation method is most probably not an optimal way of estimating vacant lot sales prices.

5.4.3 Limitations of the research methods

The research setting, and the research methods have their limitations. In this research we use two different research methods: correlation analysis and regression analysis. Besides these analysis methods we use the residual land valuation method to calculate the expected development returns and residual land values. The residual valuation method requires assumptions that affect the results in many ways. We will next go through the limitations of each of the steps that are taken in the research in total.

In the first part of the research we use the residual land valuation method to calculate the expected development returns and residual land values. There are two important limitations to this method of which the first considers the formula itself and the other considers the used data. The most important limitation of the residual valuation formula is that it assumes that all actors in the land market are property developers that share a unified logic for the appreciation of vacant lots. This logic is represented by the formula, which assumes that the developer will calculate the expected end-product value using prices of similar apartments in the same market. After this the developer subtracts all development related costs from the end-product price. Finally, the developer sets the required development return and subtracts the return resulting in a residue or "residual land value". After the calculation the developer compares the residual value to the asking price of the lot and only purchases the lot if the residual value is higher than the asking price. This logic is the basis of residual land valuation, but there can be other business logics that the method does not consider. For example, the developer might by land in advance and speculate with future construction costs and apartment prices that do not represent the current market situation. Also, not all of the actors in the market are developers. In some of the areas, such as Helsinki, municipalities are large land owners that can have other strategic goals than the maximization of revenue from lot sales. This can affect the vacant lot sales prices in ways that contradict the logic of the residual land valuation method.

Besides the limitations of the residual valuation formula itself, the residual valuation also has limitations considering the data. The residual valuation should apply the best market information about end-product prices (apartment prices), total development costs and required development returns. It is problematic to estimate the total development costs and required development returns. As there is no data available for total development costs we will use the estimated construction costs instead. These do not consider all possible development costs such as acquisition costs, zoning costs, marketing of the apartments etc. Also, there was no data available for required development return and it is used as a constant in the formula. All these problems related to the data must also be understood as limitations of the method itself and we can expect to have results that apply only within the constraints of the formula and the used data. As the residual valuation outcomes are what is being tested with the other methods these limitations have a major role in the rest of the analysis as well.

In the second part of the research we focus on the correlation analysis. The most important limitation here is that we can measure a correlation between two variables, but this does not ensure that there is a causal relationship between the variables (Metsämuuronen 2010, s. 370). This must be kept in mind when the results are analyzed. The correlation analysis is limited to pointing out possible problems in the residual valuation method and it helps to understand the results in the regression analysis, but it does not give a systemic understanding of the relationship between the different variables.

The third part of the research, regression analysis, also has its limitations. Regression analysis can be used to create estimators for the dependent variable (Vacant lot sales price) using other data as predictors. In linear ordinary least squares regression these estimators are constructed by minimizing the square sum of error between the dependent variable and the predictor. This method is very sensitive to the statistical representation of the data and only works if the data represents the real-world population of the measured variable. (Pejic et al. 2013. pp. 43-44)

The vacant lot sales data has some major problems in this aspect as the land market has very different types of vacant lots that are traded infrequently. It is extremely hard to know how well the data represents the land market in general. There is also a problem with data coherence and accuracy. The data that is used as a predictor (residual land value) is constructed on a district -level and it represents the district averages. The vacant lot sales data however is specific to the individual sites that can have many micro-locational attributes as well as individual characteristics. With a large sample size this would be a less of a problem as it would be more probable that individual characteristics do not have a major impact on the results. The sample size in this case however is small (57) with some districts only having a sample size of 3. With these sample sizes there is a high risk that some lot characteristics have a very high impact on the sales prices and therefore they do not represent the “average” vacant lot prices in the area. This is something that cannot be avoided in the analysis due to lack of data and it makes the analysis less accurate and reliable. All these limitations to the research methods are significant and while they cannot be avoided in this research setting, they must be considered when the results are analyzed, and we must take them into account when conclusion are drawn.

6 Results

In this chapter we present the results of the empirical research. As the empirical research was done in a multitude of steps, this chapter is divided into three parts that represent these steps that had to made to achieve the final results. In the first part we present the results of the residual valuation that is the basis of the research and these results are used in the second and third part of the analysis. In the second part we present the results of the correlation analysis, where the residual valuation function parameters as well as the results are tested for correlation. In the third part we present the results of the regression analysis, where the residual valuation estimates are tested against the vacant lot sales observations.

6.1 Residual valuation

The residual valuation was done with two different arrangements of the same functional form. In the first arrangement we used the apartment prices, construction costs and vacant lot sales data to estimate the developer's expected returns for each of the observation districts. In the second arrangement we used the apartment prices, construction costs and the average developers expected return (that was calculated in the first arrangement) to estimate the residual land values for each of the districts. Even though the second arrangement uses the average results of the first arrangement, it is important to notice that the results itself exclude each other as they use different assumptions. The first arrangement assumes that the lot price is constant and results in the expected development return. The second arrangement assumes that the development return requirement is constant and results in residual land value. The average expected development return is used here simply as there was no data available for required development returns. Even though the expected return (estimate) is different from required return (investor requirement) this test arrangement enabled us to look at the residual land values with the assumption that the required returns do not change between districts. This might not be true in the real world, but it is a better estimate than no estimate at all and the relevance of this assumption is tested further in the correlation and regression analyses.

6.1.1 Residual valuation variables

In both residual valuation arrangements, we first compiled the data to district level so that all the datasets are in the same form and same level of accuracy. As the datasets were initially very different, the compilation required different transformations for different datasets. The apartment price data was initially gathered from individual apartment level and for the compilation we used district-level average prices per apartment area (€ / m2). The vacant lot sales data was also gathered from individual vacant lots and the compilation was done similarly to the apartment price data, by calculating the district-level average prices. However, the vacant lot sales data initially uses price per gross area rather than price per apartment net area. This was a problem for the analysis as the data is not in the same format as the other data. To solve this problem, we used a transformation from gross area to net area with a multiplier of 1.4 (net area x 1.4 = gross area), which represents a rough average transformation rate for multi-story apartment buildings in Finland. This calculated gross area includes the common spaces and technical spaces that are not represented in the net apartment area. The transformation rate is a rough average approximation and while it represents the average building practices in Finland, there can be variation in individual projects especially due to the distribution of common spaces. (Haahtela & Kiiras 2014)

The construction cost data was gathered from price index areas in the Helsinki region. These index areas are larger than the observational districts, so each observational district was given the estimated construction costs from the index area where it is categorized. This makes the construction cost data the least accurate data in the analysis, since it was the only dataset where the accuracy of the data is lower than the accuracy of the analysis. The residual valuation parameters and their values are represented in table 10.

Table 10: Residual valuation data

City	District	Apartment type (rooms)	Apartment price (€ / net area)	New construction costs (€ / net area)	Lot price (€ / net area)
Helsinki	Jätkäsaari	1	9,838	3,464	1,712
		2	7,508	3,193	1,712
		3	7,548	3,168	1,712
Helsinki	Kalasatama	1	10,414	3,464	1,597
		2	8,584	3,193	1,597
		3	7,716	3,168	1,597
Vantaa	Viertola	1	6,591	3,464	704
		2	5,423	3,193	704
		3	5,226	3,168	704
Vantaa	Martinlaakso	1	5,346	3,464	791
		2	5,362	3,193	791
		3	4,562	3,168	791
Vantaa	Kivistö	1	5,874	3,464	577
		2	4,334	3,193	577
		3	3,938	3,168	577
Espoo	Niittykumpu	1	7,250	3,464	2,198
		2	6,674	3,193	2,198
		3	6,099	3,168	2,198
Espoo	Eestinlaakso	1	6,027	3,464	798
		2	4,462	3,193	798
		3	4,005	3,168	798
Espoo	Saunalahti	1	6,405	3,464	829
		2	5,510	3,193	829
		3	4,599	3,168	829
Kerava	Kerava keskus	1	5,107	2,801	297
		2	4,650	2,643	297
		3	4,107	2,581	297
Järvenpää	Järvenpää keskus	1	5,377	2,801	432
		2	4,866	2,643	432
		3	3,968	2,581	432
Average			5,912	3,155	994

The expected development returns, and the residual land values were calculated for 1, 2 and 3 room apartments for each of the 10 observation districts. Additionally, the average expected development returns, and residual land values were calculated for each of the districts using a 1:1:1 ratio between the apartment types. This ratio may not represent the most likely or typical development situation in terms of apartment type mix in a project, but it gives an impression of the expected development returns of the district without adding any further assumptions. Assumptions considering a typical unit mix of apartments may vary between development situations and this is the most neutral way of presenting the average results.

6.1.2 Arrangement 1 (Expected development returns)

The expected development returns were calculated using the first arrangement of the residual valuation formula. Here apartment prices, construction costs and vacant lot prices were used to calculate the expected development returns, using the following function:

Residual valuation function (arrangement 1):
Expected development return = Apartment price - Construction costs - Vacant lot price

On average the highest expected development returns were in Kalasatama, Helsinki, with the expected return of 45.3% for 1-3 room apartments on average. The lowest expected returns were in Eestinlaakso, Espoo, with the expected return of 15.7% for 1-3 room apartments on average. The average for all districts for 1-3 room apartments was 28.4%. All the average expected development returns for the observation districts can be found in table 11.

Table 11: Expected development returns, district average

City	District	Expected return (€ / net area)	Expected return / Apartment price (%)
Helsinki	Jätkäsaari	3,311	39.9
Helsinki	Kalasatama	4,032	45.3
Vantaa	Viertola	1,767	30.7
Vantaa	Martinlaakso	1,024	20.1
Vantaa	Kivistö	863	18.3
Espoo	Niittykumpu	1,200	18.0
Espoo	Eestinlaakso	758	15.7
Espoo	Saunalahti	1,401	25.4
Kerava	Kerava keskus	1,649	35.7
Järvenpää	Järvenpää keskus	1,630	34.4
Average		1,763	28.4

The expected development returns varied between the apartment types and we can get a more detailed view of the results by looking at the results by apartment types (1-3 rooms) instead of district-level averages. The highest expected development return was in 1-room apartments in Kalasatama, Helsinki, with the expected return of 51.4%. The lowest result was in 3-room apartments in Eestinlaakso, Espoo, with the expected return of 1.0%. The highest overall ranking district was Kalasatama, where the expected returns ranged from 38.2% (3-room apartments) to 51.4% (1-room apartments). The lowest overall ranking district was Eestinlaakso where the expected returns ranged from 1.0% (3-room apartments) to 29.3% (1-room apartments).

The highest variation between the apartment types was in Eestinlaakso, Espoo with the gap between 3-room apartments and 1-room apartments being 28%-units. The lowest variation between the apartment types was in Martinlaakso, Vantaa, with the gap between 3-room apartments and 1-room apartments being 7%-units. The variation between apartment types overall was also significant. 1-room apartments had the highest average expected development return of 35% and 2-room apartments had the average of 27.6% while 3-room apartments had the lowest average of 19.8%. This variation between apartment types is significant with a gap of 15.2%-units between 1-room apartments and 3-room apartments on average. The results for all the areas and apartment types can be seen from table 12.

Table 12: Expected development returns by apartment type

City	District	Apartment type (rooms)	Expected return (€ / net area)	Expected return / Apartment price (%)
Helsinki	Jätkäsaari	1	4,662	47.4
		2	2,603	34.7
		3	2,668	35.3
Helsinki	Kalasatama	1	5,353	51.4
		2	3,794	44.2
		3	2,951	38.2
Vantaa	Viertola	1	2,422	36.8
		2	1,525	28.1
		3	1,353	25.9
Vantaa	Martinlaakso	1	1,091	20.4
		2	1,378	25.7
		3	603	13.2
Vantaa	Kivistö	1	1,832	31.2
		2	563	13.0
		3	193	4.9
Espoo	Niittykumpu	1	1,587	21.9
		2	1,282	19.2
		3	732	12.0
Espoo	Eestinlaakso	1	1,765	29.3
		2	471	10.6
		3	39	1.0
Espoo	Saunalahti	1	2,112	33.0
		2	1,488	27.0
		3	602	13.1
Kerava	Kerava keskus	1	2,009	39.3
		2	1,710	36.8
		3	1,229	29.9
Järvenpää	Järvenpää keskus	1	2,143	39.9
		2	1,791	36.8
		3	955	24.1
Average			1,763	27.5

Based on the results the variation between the estimated expected development returns overall is relatively high, ranging from 15.7% to 45.3% of the total development value. The gap between the lowest and highest is nearly 30% (29.6%) of total development value and it can be considered as significant since small (1-3 room) apartments are relatively homogenous products compared to other forms of real estate and all the observation districts were from relatively limited geographical area (Helsinki region). We can also conclude that the estimates made with the residual valuation method give different apartment types very different expected returns. Overall the variation between the apartment types is significant (15.2%-units), but not as high as the variation between the districts (28.4%-units).

After this first step we have successfully answered the first research question regarding development profitability: “Are there significant differences in predicted housing development profitability or land residual values in different districts or apartment types in the Helsinki region, when the land residual method is used in the estimation?” The answer is yes to both the districts and apartment types. There are significant differences in expected development returns between the districts (28.4%-units) and between the apartment types (15.2%-units).

6.1.3 Arrangement 2 (Residual land values)

The residual land values were calculated using the second arrangement of the residual valuation formula. Here apartment prices, construction costs and required development returns were used to calculate the residual land values, using the following function:

Residual valuation function (arrangement 2):
Residual land value = Apartment price - Construction costs - Required development return

The second arrangement of the function required the data of required development returns. Unfortunately, such data was not available for this research. There was no public data available regionally or considering apartment types or any types of development projects. The developers were not particularly willing to share this information for the purposes of this research due to its nature as a business secret. To fill this gap of knowledge we used the results of the first arrangement of the residual valuation function that were introduced in the previous chapter. On average the first arrangement provided us with the expected development return of 27.5% that was used as a baseline scenario for the second arrangement. This figure represents the average expected development return for the districts and apartment types on average and it is used in the function as a constant. There is a relatively high chance that the average expected development return is overly influenced by some of the very high values present in some of the districts such as Kalasatama. Also, the construction cost data does not entirely represent the total development costs that relate to a development project (such as zoning and marketing). For these reasons we found it appropriate to add two additional scenarios that have a lower required development return than the baseline of 27.5%. These scenarios utilize required returns of 20% and 15%. These levels are justified by the fact that we know that the construction costs do not represent the total development costs entirely, so it is fair to assume that the returns would most likely be lower than the estimated results. The results of the residual land values are provided with each of these scenarios and we will discuss the relevance of these scenarios further in chapter 7.

The decision to use the previously estimated development returns here comes with the assumption that there are no differences in required returns between districts and apartment types. This assumption is most likely inaccurate and might not represent the real-world investment requirements. However, it is better to take a neutral approach here than to add any other assumptions that cannot be based on evidence. The results must be seen in a context where required returns stay the same even with different districts and apartment types. This assumption could be further tested with additional data and these possible additional studies are discussed further in chapter 8.

The results for the baseline scenario, where the required development return of 27.5% was used in the estimation is discussed in detail. Results for other scenarios of 20% and 15% can be seen from table 4 and are more briefly discussed in the summary of this subchapter. In the baseline scenario the highest residual land values were in Kalasatama, Helsinki, with the residual land value of 3,180 €/ net area for 1-3 room apartments on average. The lowest residual land values were in Kivistö, Vantaa, with the residual land value of 143 €/ net area for 1-3 room apartments on average. The highest average residual land value was 21 times (2100%) higher than the lowest average residual land value. The average for all districts for 1-3 room apartments was 1,131 € / net area. All the average residual land values for the observation districts with the different required returns can be found in table 13.

Table 13: Residual land values, district average (€ / net area)

City	District	Lot price (€ / net area)	Development return 27.5%	Development return 20%	Development return 15%
Helsinki	Jätkäsaari	1,712	2,741	3,363	3,778
Helsinki	Kalasatama	1,597	3,180	3,848	4,293
Vantaa	Viertola	704	891	1,322	1,609
Vantaa	Martinlaakso	791	415	797	1,051
Vantaa	Kivistö	577	143	497	733
Espoo	Niittykumpu	2,198	1,563	2,064	2,398
Espoo	Eestinlaakso	798	228	590	831
Espoo	Saunalahti	829	716	1,129	1,404
Kerava	Kerava keskus	297	675	1,022	1,253
Järvenpää	Järvenpää keskus	432	759	1,114	1,351
Average		994	1,131	1,575	1,870

The residual land values varied between the apartment types and we can get a more detailed view of the results by looking at the results by apartment types (1-3 rooms) instead of district-level averages. The highest result was in 1-room apartments in Kalasatama, Helsinki, with the residual land value of 4,086 € / net area. The lowest result was in 3-room apartments in Kivistö, Vantaa, with the residual land value of -313 € / net area. The highest overall ranking district was Kalasatama, where the residual land values ranged from 2,425 – 4,086 € / net area with the highest value in 1-room apartments and lowest value in 3-room apartments. The lowest overall ranking district was Kivistö, Vantaa where the residual land values ranged from -313 to 794 € / net area with the highest value in 1-room apartments and lowest value in 3-room apartments.

The highest absolute variation between the apartment types was in Kalasatama, Helsinki with the gap between 3-room apartments and 1-room apartments being 1660 € / net area. The highest relative variation was in Saunalahti, Espoo with the difference between 1-room apartments (highest) and 3-room apartments (lowest) being 609% (1,013 € / net area). The lowest absolute variation between the apartment types was in Kerava keskus, Kerava, with the gap between 3-room apartments and 1-room apartments being 504 € / net area. The lowest relative variation was in Kalasatama, Helsinki with the difference between 1-room apartments (highest) and 3-room apartments (lowest) being 59% (1,660 € / net area).

The variation between apartment types overall was also significant. 1-room apartments had the highest average residual land value of 1,615 € / net area and 2-room apartments had the average of 1,076 € / net area while 3-room apartments had the lowest average of 702 € / net area. This variation between apartment types is significant with a gap of 913 € / net area between 1-room apartments and 3-room apartments on average. On average 1-room apartments (highest) had a 130% higher residual land value than 3-room apartments (lowest).

There was one exception, Martinlaakso, Vantaa, where 2-room apartments had the highest residual land value (694 € / net area) and 1-room apartments were the second highest (412 € / net area). In all other areas the residual land value was higher in smaller apartments than larger ones. The results for all the areas and apartment types can be seen from table 14.

Table 14: Residual land values by apartment type (€ / net area)

City	District	Apartment type (rooms)	Lot price (€ / net area)	Development return 27.5%	Development return 20%	Development return 15%
Helsinki	Jätkäsaari	1	1,712	3,668	4,406	4,898
		2	1,712	2,250	2,813	3,189
		3	1,712	2,304	2,870	3,247
Helsinki	Kalasatama	1	1,597	4,086	4,867	5,387
		2	1,597	3,030	3,674	4,103
		3	1,597	2,425	3,004	3,390
Vantaa	Viertola	1	704	1,314	1,809	2,138
		2	704	738	1,145	1,416
		3	704	621	1,013	1,274
Vantaa	Martinlaakso	1	791	412	813	1,080
		2	791	694	1,096	1,364
		3	791	139	481	709
Vantaa	Kivistö	1	577	794	1,235	1,528
		2	577	-51	274	490
		3	577	-313	-18	179
Espoo	Niittykumpu	1	2,198	1,792	2,335	2,698
		2	2,198	1,645	2,146	2,480
		3	2,198	1,253	1,711	2,015
Espoo	Eestinlaakso	1	798	905	1,357	1,659
		2	798	42	377	600
		3	798	-265	36	236
Espoo	Saunalahti	1	829	1,179	1,660	1,980
		2	829	802	1,215	1,490
		3	829	166	511	741
Kerava	Kerava keskus	1	297	901	1,284	1,539
		2	297	728	1,077	1,309
		3	297	397	705	910
Järvenpää	Järvenpää keskus	1	432	1,097	1,500	1,769
		2	432	885	1,250	1,493
		3	432	296	593	792
Average			994	1,131	1,575	1,870

Based on the results the variation between the residual land values is relatively high, ranging from -313 to 4,084 € / net area. When apartment type is considered the differences are also significant with the residual land values ranging from 702 to 1,615 € / net area. The variation between the lowest and highest apartment types on average (913 €/ net area) is significant but not nearly as high as the variation between the lowest and highest district on average (3,037 € / net area).

After this part we have successfully answered the first research question regarding residual land values: "Are there significant differences in predicted housing development profitability or land residual values in different districts or apartment types in the Helsinki region, when the land residual method is used in the estimation?” The answer is yes to both the districts and apartment types as it was also with the expected development returns. There are significant differences in residual land values between the districts (3,037 € / net area) and between the apartment types (913 €/ net area).

6.2 Correlation analysis

The correlation analysis was done to all the residual valuation parameters and both of the residual valuation results (expected development return & residual land value). The aim of the correlation analysis was to find an answer to the second research question: “Is there a correlation between the residual valuation parameters (apartment prices, construction costs, vacant lot sales prices) and the valuation results (expected development returns & residual land values)?”

The answer to this research question is important, so that we can better understand the underlying structure of the residual valuation method and its possible problems in the estimation of expected development returns as well as land values. All the parameters and the results were tested for correlation and the results are presented in the next subchapter. Even though all the correlation results are reported, it is important to notice that the expected development returns, and residual land values are the results of different assumptions and they cannot be compared to each other. The expected development returns were calculated using the vacant lot sales data as a parameter in the residual valuation arrangement 1. The residual land values were calculated without using the vacant lot sales data and using a constant value to represent the required development return. For this reason, the correlation between the expected development returns and land residual values does not give us any information about their relationship as the assumptions and data that were used in their calculation are different and they are both estimated values and not observations.

Next, we will go through the correlations of the valuation parameters and valuation estimates. Pearson correlation factor was calculated for all the parameters as well as the results. The parameters and results are separated in the tables to clarify what is observation data and what is an estimate made with the residual valuation method. The first table (table 6) contains the correlations calculated for the entire datasets without separating the apartment types to 1, 2 and 3 room apartments. This is a broader analysis that gives a good over the top look at the results in general. In this part the room amount of the apartment was taken into the correlation analysis as a separate parameter for which correlation was also tested. The parameters tested in the general analysis were the following: apartment price, construction costs, lot price and apartment room amount. The residual valuation results that were tested were the expected development returns and land residual values that were calculated in the analysis of the previous chapter. To clarify the difference between the parameters and the residual valuation results, we will first go through the correlations between the parameters and then continue to the correlation between individual parameters and the residual valuation results.

Between the parameters, the highest correlation was between apartment prices and lot prices, with a Pearson correlation of 0.71. This is a high positive correlation as the range in Pearson correlation is from -1 to 1 with 1 being an absolute positive correlation. The second highest correlation was between new construction costs and apartment prices with the Pearson correlation of 0.52. This can also be considered as a significant positive correlation between the two datasets. New construction costs and lot prices also had a positive correlation with the Pearson correlation of 0.47. When we look at the apartment types by the amount of rooms in the apartment it is shown that the amount of rooms has a similar sized negative correlation with both the apartment price (-0.41) and construction costs (-0.42). The amount of rooms however has no correlation with lot prices, which is an interesting result since all of the above-mentioned parameters had a correlation with each other.

Based on the results the apartment prices, new construction costs and lot prices all have positive (although different) correlations with each other. Apartment room number correlates negatively with apartment prices and construction costs and based on this we can conclude that apartments with a higher number of rooms have both lower prices per net area and lower construction costs per net area. This result is not surprising especially considering construction costs as smaller apartments have a higher percentage amount of cost-intensive spaces, such as bathroom and kitchen space.

After analyzing the parameters in relation to each other we continue to the correlation between the individual parameters and the residual valuation results. The highest correlation was between land residual values and apartment prices with almost an absolute positive Pearson correlation of 0.98. The second highest correlation of 0.91 was with the apartment prices and expected development returns. The other factors had significantly lower correlations with the residual valuation results than the apartment prices. New construction costs had a 0.26 correlation with expected development returns and a 0.33 correlation with land residual values. Lot prices had a 0.38 correlation with expected development returns and a 0.67 correlation with residual land values.

When correlation with lot prices is considered it is important to remember that lot prices were used to calculate the expected development returns but they were not used to calculate residual land values. For this reason, the correlation between the lot prices and residual land values has to be seen separately from the other correlations. While the other correlations can be used to analyze how high of an impact a parameter may have had in the estimate outcome, the correlation between the lot prices and residual land values can be used evaluate how well the residual land value estimate relates to the real-world observations of lot prices. The correlation between these two is quite high (0.67) and based on this result there is a reason to believe that the residual land values move much in the same direction with the lot price observations. This result supports the structure of this study as the next planned step is the deeper analysis of the relationship between residual land values and lot price observations that is studied in the next chapter with a regression analysis. All the results of the correlation analysis made for the entire dataset can be found in table 15.

Table 15: Correlation between parameters and results, all apartment types

	Parameters				Results	
	Apartment price	New construction costs	Lot price	Rooms	Expected return	Land residual value
Apartment price	1.00	0.52	0.71	-0.41	0.91	0.98
New construction costs	0.52	1.00	0.47	-0.42	0.26	0.33
Lot price	0.71	0.47	1.00	0.00	0.38	0.67
Rooms	-0.41	-0.42	0.00	1.00	-0.46	-0.34
Expected return	0.91	0.26	0.38	-0.46	1.00	0.94
Land residual value	0.98	0.33	0.67	-0.34	0.94	1.00

After the first step, the correlation analysis was repeated by using the different apartment types as separate datasets that were tested for correlation. This analysis is more detailed as the correlations can be seen separately for the apartment types. For the parameters the correlation between apartment prices and lot prices ranged from 0.75 (1-room apartments) to 0.81 (2-room apartments) with 3-room apartments having a correlation of 0.82. The correlation between lot prices and new construction costs was the second highest and it was at the same level with all the apartment types with a correlation of 0.53. Although the level of correlation is not very different compared to the analysis with the entire dataset (0.47) it is still higher when the apartment types are analyzed separately. When analyzed separately, it is also higher than the correlation between apartment prices and construction costs, unlike in the first analysis. The correlation between apartment prices and construction costs however is notably lower when the apartment types are analyzed separately ranging from 0.37 (2-room apartments) to 0.45 (1-room apartments).

Based on these results the analysis of the apartment types separately gives different correlations for the parameters. In both cases the highest correlation is between apartment prices and lot prices but when the apartment types are analyzed separately the second highest is between construction costs and lot prices and only the third highest is between apartment prices and construction costs. All the above mentioned still have a positive and similar sized correlation between each other in both versions.

When we look at the correlation between the individual parameters and the residual valuation results there are more differences in the correlations between the general analysis and the analysis separated by apartment types. The highest correlation in the detailed analysis is still between the apartment prices and residual land values ranging from 0.94 (2-room apartments) to 0.97 (3-room apartments) while 1-room apartments have a correlation of 0.95. This is also a near absolute correlation between apartment prices and residual land values like in the first analysis. The second highest correlation is between the apartment prices and the expected development returns ranging from 0.85 (2-room apartments) to 0.92 (1-room apartments) which is also near the level of the first analysis (0.91). Lot prices also had a similar level of correlation with the expected development returns in the detailed analysis ranging from 0.41 (2-room apartments) to 0.45 (1 & 3 -room apartments). Lot prices had somewhat higher correlation with the land residual values than in the first analysis ranging from 0.69 (1-room apartments) to 0.75 (3-room apartments) when the first analysis had the correlation of 0.67 between the two. This is not a very significant difference, but it implies that the correlation is stronger when apartment type is taken into consideration and this is especially important when the regression analysis is considered in the next chapter.

The most notable difference in the general analysis (without apartment type) compared to the apartment type -level analysis is with the construction costs. In the detailed analysis the correlation between construction costs and residual land values ranged from 0.15 (2- room apartments) to 0.26 (1-room apartments) while in the previous analysis it was 0.33 between the two. The correlation was also lower between the construction costs and expected development returns ranging from -0.05 (2-room apartments) to 0.16 (1-room apartments) while in the first analysis it was 0.26. The correlation between construction costs and the residual valuation results is between small to none when the apartment types are analyzed separately and this raises a question of whether the correlation in the previous analysis was mainly related to the differences in the apartment types after all.

Based on the results of the more detailed analysis there is a significant positive correlation between all the residual valuation parameters: apartment prices, new construction costs and lot prices. Apartment prices have a very high correlation with both the expected development returns and the residual land values and lot prices have a smaller but significant positive correlation with them. Construction costs on the other hand have a relatively low correlation with the residual land values and a very low correlation with the expected development returns. All the correlations between the parameters and the residual valuation results can be seen by apartment type in table 16.

Table 16. Correlation between parameters and results by apartment type

	Parameters			Results	
	Apartment price	New construction costs	Lot price	Expected return	Land residual value
1-Room Apartments					
Apartment price	1.00	0.45	0.75	0.92	0.95
New construction costs	0.45	1.00	0.53	0.16	0.26
Lot price	0.75	0.53	1.00	0.45	0.69
Expected return	0.92	0.16	0.45	1.00	0.91
Land residual value	0.98	0.26	0.69	0.96	1.00
2-Room Apartments					
Apartment price	1.00	0.37	0.81	0.85	0.94
New construction costs	0.37	1.00	0.53	-0.05	0.15
Lot price	0.81	0.53	1.00	0.41	0.73
Expected return	0.85	-0.05	0.41	1.00	0.90
Land residual value	0.97	0.15	0.73	0.92	1.00
3-Room Apartments					
Apartment price	1.00	0.41	0.82	0.87	0.97
New construction costs	0.41	1.00	0.53	0.02	0.19
Lot price	0.82	0.53	1.00	0.45	0.75
Expected return	0.87	0.02	0.45	1.00	0.93
Land residual value	0.97	0.19	0.75	0.93	1.00

6.3 Regression analysis

The aim of the regression analysis was to answers the third research question: "Can the residual land valuation method be used to predict vacant lot sales prices in the Helsinki region? " To approach this question, we need to compare the estimates made with the residual land valuation method to the actual vacant lot sales price observations from the observation areas. Regression analysis is a useful tool for this kind of analysis as it can show us how well the residual land values can predict the vacant lot sales prices.

In the regression analysis we can already use some of our previous findings to support the research setting of the analysis. For example, we know based on the correlation analysis of the previous chapter that residual land values and vacant lot sales prices have a 0.67 Pearson correlation which gives us a good reason to analyze the relationship with regression analysis. If there was no correlation between the two, then it would be most likely that the residual land values could not be used to predict vacant lot sales prices and regression analysis would give us little additional information. Based on the correlation analysis we also know that apartment prices have an even higher correlation (0.71) with the vacant lot sales prices than the residual land value estimates have. This means that the apartment prices could predict vacant lot sales prices better than the residual land value estimates. For this reason, it is appropriate to test this possibility as well by constructing two regression models: one with the residual land value as predictor and another with the apartment price as predictor. The main difference between the apartment prices and the residual land value estimates in this research is that the residual valuation takes into account construction costs. The required development profit is dealt with as a constant and as such has no effect on the regression analysis. If the residual land value cannot predict vacant lot sales prices better than apartment prices can predict them alone, then the residual land value estimate is not a very good predictor to vacant lot sales prices.

In the previous parts of this research we have used data that separates apartment types to individual datasets that have been studied apart from each other. In this part of the analysis this cannot be done as the vacant lot sales data does not contain any information on the possible apartment types that could be constructed on the lot. For this reason, we use the average residual land values and apartment prices that have been calculated with a 1:1:1 -ratio between 1, 2 and 3 -room apartments. This mix was used before in the previous chapters, and as discussed it does not represent any typical apartment type mix in a development project. An assumption of a typical mix would add additional assumptions to the analysis that cannot be based on evidence as development projects can be very different considering the unit mix (Pennanen 2019). It is a more neutral approach to use an even mix, although it might not represent a typical situation. This research can always be repeated with another apartment type mix to achieve results that better suit a particular type of development project and these suggestions are further discussed in chapter 8.

6.3.1 Regression data

Before going to the regression analysis, let us first look at the data that we are using. The residual land values were calculated for each of the districts by using the average residual land value of 1, 2 and 3 room apartments with a 1:1:1 ratio for apartment mix. All the apartment types are equally represented in the average residual land value. In the residual valuation we used the average expected development return that was calculated using the first arrangement of the residual valuation, where expected development returns were calculated with the function using vacant lot sales prices. This average expected development return was used as a constant when the residual land values were calculated. For this reason, it is not surprising that for roughly half of the areas the residual land value is higher than the vacant lot sales price on average and similarly roughly half of the areas have a lower residual land value than the vacant lot sales prices. There are differences on the district level and some of the estimates are significantly different from the district average lot price. Overall the differences range from the absolute values of 113 to 1,584 € / net area with the highest absolute difference in Kalasatama, Helsinki. The relative differences range from 14% to 127% with highest difference in Kerava keskus, Kerava (127% higher residual land value than average sales price) and the lowest difference in Saunalahti, Espoo (14% lower residual land value than average sales price). The district averages for vacant lot sales prices, residual land value estimates and the absolute and relative differences can be seen from table 17.

Table 17. Regression data, district averages

City	District	Residual land value (€)	Lot price (€ / net area)	Difference (€)	Difference (%)
Helsinki	Jätkäsaari	2,741	1,712	1,029	60
Helsinki	Kalasatama	3,180	1,597	1,584	99
Vantaa	Viertola	891	704	187	27
Vantaa	Martinlaakso	415	791	-376	-48
Vantaa	Kivistö	143	577	-434	-75
Espoo	Niittykumpu	1,563	2,198	-635	-29
Espoo	Eestinlaakso	228	798	-570	-71
Espoo	Saunalahti	716	829	-113	-14
Kerava	Kerava keskus	675	297	378	127
Järvenpää	Järvenpää keskus	759	432	327	76
Average		1,131	994	138	15

As the residual land valuation was done on a district level the result is that the residual land value estimates are always district-level data. Vacant lot sales prices however are the prices of individual lots that are categorized to a district geographically. When the vacant lot sales prices are predicted for individual lots we must use the district-level estimates as it is impossible to calculate the residual land value for the individual lot using the data we have. This results in major inaccuracies considering individual lots as they have individual properties that may affect their price, but the residual land valuation deals with them as though they were completely identical. The regression analysis is done by using the lot price data of the individual lots and comparing this to the district-level estimate that the lot is given based on its regional location. This results in the fact that for example the residual land values of lots in Kalasatama all are the same (2,272 € / net area) while the individual lot prices range from 482 to 2,040 € / net area. This is something that cannot be avoided in this research setting, but it must be kept in mind when the results are analyzed. The lot price data and the residual land value estimates for all the individual lots can be found in table 18.

Next, we will take a brief look at the statistical factors of the regression data to better see if the data fits the requirements of linear least squares regression. The sample size of the lot price data is 57 individual vacant lot sales observations. This is a rather small sample size, but as a total it should be enough for regression analysis. However, the sample sizes from individual districts have some variation and some of the districts have very few observations. The most observations are in Kivistö, Vantaa with a sample of 13 vacant lot sales observations. There were only 3 observations in 5 of the districts (Viertola, Niittykumpu, Eestinlaakso, Kerava keskus, Järvenpää keskus) for each district. This is a very low sample size for districts and the lot price information representation is statistically less reliable for these districts. All the districts from the region that had even lower amount of lot price observation were discarded in the collection of the data as this is the lowest sample size that we can accept in any circumstance. It is unfortunate that there is no more data available and thus the sample sizes are low for this analysis. The trading in the vacant lot market in the general area is not frequent and for this reason the low sample sizes are unavoidable in a research that aims to compare multiple different districts. The sample sizes for each of the districts can be seen from table 19.

Table 18. Regression data by individual lot transaction

City	District	Lot area (m2)	Building rights (m2)	Lot price €	Lot price (€ /gross area)	Residual land value (€ /gross area)
Helsinki	Jätkäsaari	1,888	8,737	10,062,500	1,152	1,958
		883	6,800	9,429,842	1,387	1,958
		941	6,200	12,34, 759	1,991	1,958
		2,040	5,400	5,745,600	1,064	1,958
		1,933	5,296	6,169,091	1,165	1,958
		2,191	4,820	4,632,526	961	1,958
		1,909	4,300	4,187,383	974	1,958
		1,367	3,705	4,033,700	1,089	1,958
Helsinki	Kalasatama	3,228	8,800	7,518,400	854	2,272
		2,615	5,970	4,884,422	818	2,272
		1,979	5,530	2,663,277	482	2,272
		900	4,250	4,963,552	1,168	2,272
		1,227	3,550	4,502,000	1,268	2,272
		1,049	3,550	4,385,000	1,235	2,272
		899	2,904	3,655,200	1,259	2,272
		1,685	2,750	5,610,000	2,040	2,272
Vantaa	Viertola	1,880	3,900	1,558,680	400	636
		4,046	3,600	1,986,050	552	636
		1,602	2,200	1,227,655	558	636
Vantaa	Martinlaakso	1,044	3,160	1,497,840	474	296
		949	2,365	1,537,640	650	296
		942	2,190	1,522,950	695	296
		1,124	2,160	853,200	395	296
		979	2,010	1,392,950	693	296
		796	1,840	1,197,560	651	296
		1,097	1,807	713,765	395	296
Vantaa	Kivistö	3,149	5,890	1,873,020	318	102
		2,794	4,400	1,652,838	376	102
		1,381	3,700	1,176,600	318	102
		3,316	8,950	4,922,500	550	102
		1,727	6,080	2,796,800	460	102
		1,183	6,040	2,778,400	460	102
		1,587	4,458	1,890,192	424	102
		1,952	4,100	1,439,100	351	102
		2,715	3,600	1,565,196	435	102
		2,137	3,380	1,385,800	410	102
		2,457	2,700	1,061,100	393	102
		2,367	2,350	1,011,075	430	102
		2,263	2,050	895,488	437	102
Espoo	Niittykumpu	5,028	8,250	11,200,000	1,358	1,117
		3,991	6,500	10,094,500	1,553	1,117
		1,965	3,000	5,401,000	1,800	1,117
Espoo	Eestinlaakso	5,426	5,650	3,126,891	553	163
		4,452	4,950	2,733,826	552	163
		4,268	2,900	1,751,762	604	163
Espoo	Saunalahti	3,851	4,100	3,325,100	811	511
		5,756	4,050	2,227,500	550	511
		7,557	3,550	3,243,000	914	511
		2,411	3,200	1,180,000	369	511
		3,079	2,850	1,453,500	510	511
		2,501	1,900	760,000	400	511
Kerava	Kerava keskus	3,851	2,375	510,625	215	482
		3,213	2,375	510,625	215	482
		2,881	2,250	463,750	206	482
Järvenpää	Järvenpää keskus	3,796	4,760	642,000	135	542
		4,553	3,400	1,564,000	460	542
		2,063	2,600	861,566	331	542

Table 19: Sample size by observation district

City	District	Observations
Helsinki	Jätkäsaari	8
Helsinki	Kalasatama	8
Vantaa	Viertola	3
Vantaa	Martinlaakso	7
Vantaa	Kivistö	13
Espoo	Niittykumpu	3
Espoo	Eestinlaakso	3
Espoo	Saunalahti	6
Kerava	Kerava keskus	3
Järvenpää	Järvenpää keskus	3
Total		57

To use a linear least squares regression analysis, there are some important assumptions that the data must meet in addition to a sufficient sample size. First there must be a linear dependence between the two datasets. This is not an easy assumption as even if there is a relationship it can be something else than linear. The first way to test if there is any relationship at all is the correlation analysis that was done in the previous chapter. From those results we know that there is a 0.67 Pearson correlation between the lot prices and residual land values. Based on this result we can expect that there is a relationship between these datasets and that this relationship is positive in terms of correlation. This is a good starting point for the regression analysis, as if there was no correlation, it would be unlikely that there is a linear relationship between the lot prices and residual land values. A more detailed way to analyze the relationship is to look at the individual lot price observations and their residual land values. When analyzed in a scatterplot it seems that the relationship is linear at least to some extent (figure 1). The relationship looks to be quite linear with low values, but the data gets more scattered when the lot prices and residual land values are high. This fan-shape form shows that the data is heteroscedastic and does not fulfill the assumption of equal variances. Heteroscedastic data is not preferable for linear regression analysis as the inequality of variances makes it less reliable to achieve the best linear unbiased estimators of the population parameters (Goldberger 1964. pp. 238-243). The use of homoscedastic data (data with equal variances) would make the regression analysis more reliable but it is not an absolute condition for simple linear regression analysis. (Fox 1997. p. 306)

Based on the analysis of the data we know that the data has some problems regarding its use in linear least squares regression analysis. The sample size for the data in total is adequate but the sample sizes for the individual districts are low and this causes inaccuracies and some reliability issues as individual observations can have an overly high impact on the results. The data is also heteroscedastic and does not fulfill the equal variances condition. The heteroscedasticity of the data makes the resulting linear estimators of the analysis less reliable and this must be considered when analyzing the results. The relationship between the two datasets seems linear based on the scatterplot analysis. This is the most important condition, since if the relationship was something else than linear the linear least squares regression analysis could not be used without transformations. In this case, the regression analysis can be done using the ordinary least squares method as there is no reason at this point to doubt the linearity of the relationship which is the zero hypothesis in this case. The linearity assumption can also be later tested using the residual data of the regression. If the residual data gives us a reason to doubt that the relationship is linear, then we must reevaluate this assumption using the results to provide a better iteration of the regression model. The scatterplot of the lot prices and residual land values is shown in figure 1.

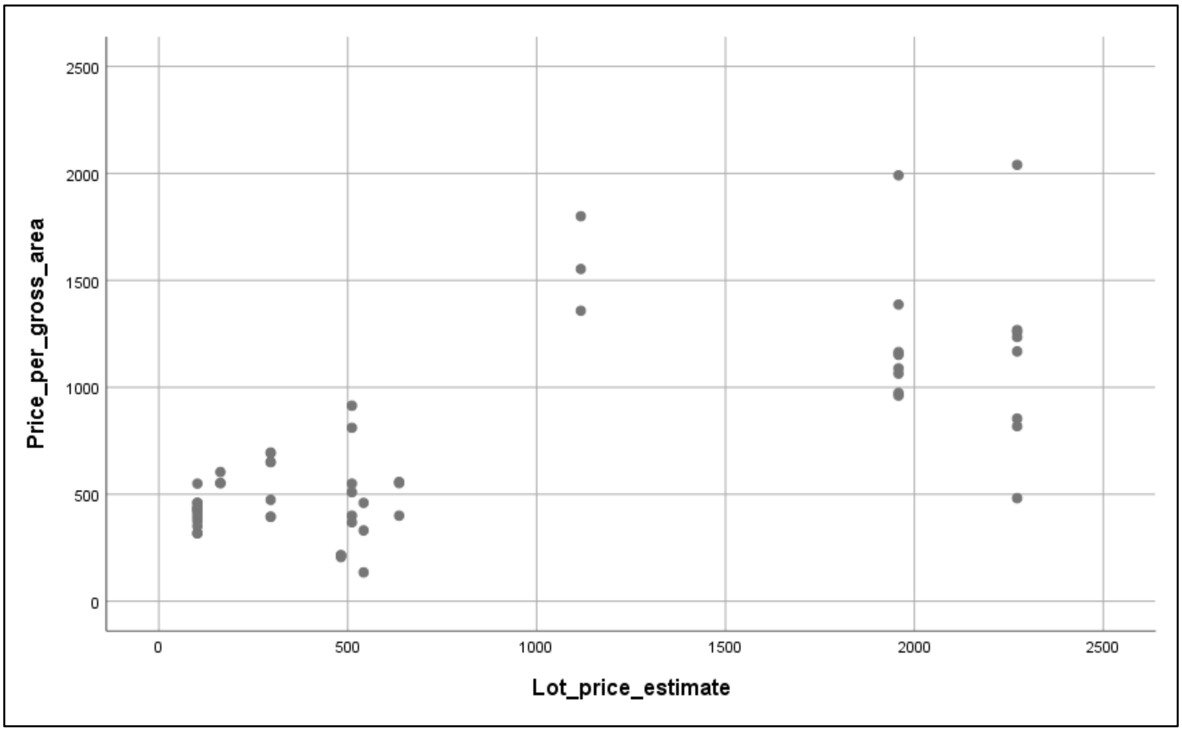

Figure 1: Scatterplot, Lot price per gross area (€/m2) & Residual land value (€/m2)

6.3.2 Model 1 (Residual land value)

The main regression model was constructed by using the residual land value as an estimator. The aim of this analysis is to test if the residual land values can predict the vacant lot sales prices as was stated in the third research question. As discussed in the previous chapter the data meets the conditions of a linear-least-squares regression analysis well enough, although the data is heteroscedastic, and the sample size is quite low. There is no evidence at the moment that contradicts the linearity assumption, but the linearity must be confirmed in the analysis of the residuals.

The regression analysis was done by using the IBM SPSS Statistics software. As there is only one predictor in the regression model there was no need to make a distinction between the different model construction methods (enter, remove, forward, backward, stepwise) and the most straightforward method "enter" was used. This method simply adds the predictor to the model and calculates the regression analysis.

Let us first look at the conditions and assumptions that are required so that linear least squares method results can be regarded as reliable. One of the preferable conditions in linear regression is that the data is normally distributed. We can analyze at the normal distribution of the data by looking at the regression standardized residual frequencies. In a normally distributed data the frequencies are higher with low absolute standardized residual values and they follow a normal distribution. The data is fairly normally distributed as can be seen from figure 2. There is a peak in values on the center of the histogram that goes beyond that of normal distribution but the data is not inclined to any direction so there is no skewness in the data. With such a small sample size it is hard to say whether the peak at the center shows a violation of normal distribution or if it is only a byproduct of the small sample. The most important thing is that there is no skewness and based on this the data is fairly normally distributed.

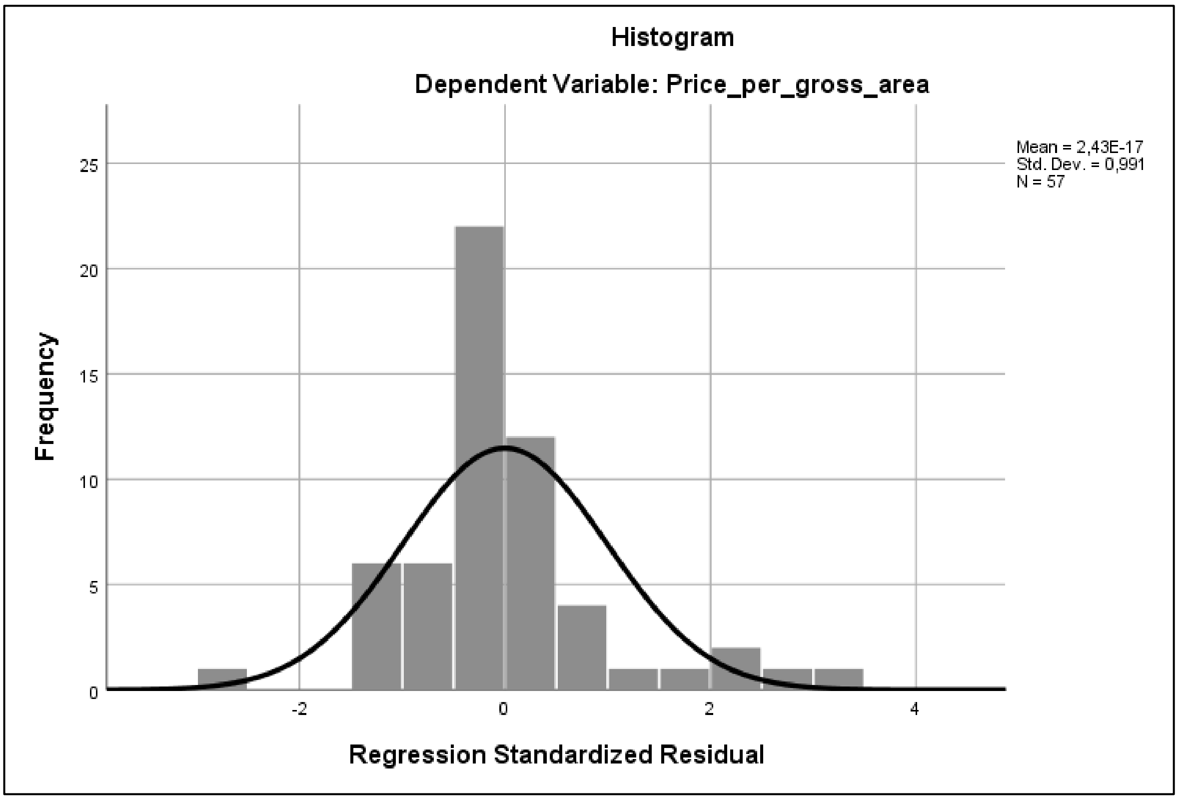

Figure 2: Histogram, Regression Standardized Residual Distribution (Model 1)

The second condition that is analyzed through the residual values is the linearity assumption and the equality of variances in the residual values. This can be analyzed in a scatterplot (figure 3) by looking at the regression standardized predicted values and the regression standardized residuals. As the dots in the scatterplot are equally and randomly distributed and there is no pattern that shows a relationship between the predicted values and residuals, we have no reason to doubt the linearity assumption. The data is homoscedastic in the residual scatterplot, which means that the linear least squares regression analysis can be used to provide unbiased linear estimators that can describe the statistical relationship of the two datasets.

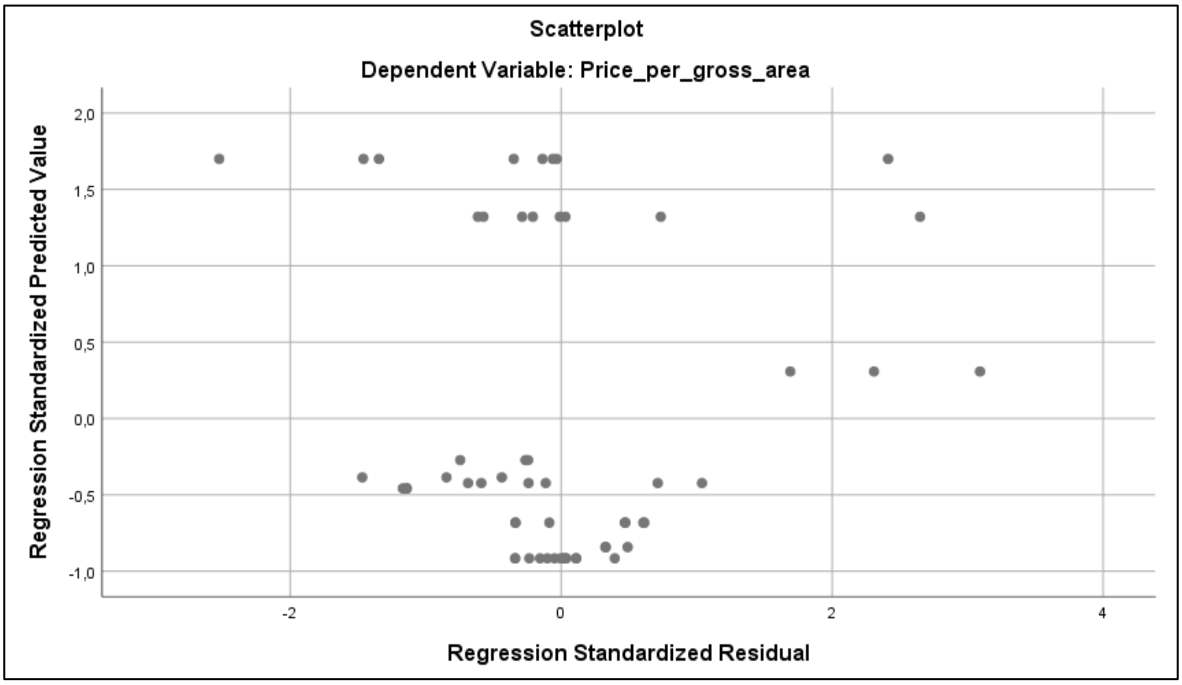

Figure 3: Residual Scatterplot (Model 1)

Next, we will go through the constructed regression model. The model has one dependent variable (Lot price, € / gross area) and one predictor (Residual land value, € / gross area). The model does not have any other predictors and can be summarized as the following regression function:

Lot price = B1 x Residual land value + C (constant)

The coefficients and model summary are shown in table 20. The model has an adjusted R square of 0.512 which means that the residual land value as a predictor can predict 51.2% of the adjusted variation in vacant lot sales prices. An explanation rate of over 50% implies that the residual land value is a good predictor for the vacant lot sales prices, but still almost half of the actual variation in the lot prices is left unexplained using this model.

Table 20: Regression model summary (Model 1)

Model Summary			
R	0.722		
R square	0.521		
Adjusted R square	0.512		
Std. Error of the estimate	315.572		
Durbin-Watson	1.200		
Coefficients			
		(Constant)	Residual land value
Unstandardized coefficients	B	385.128	0.393
	Std. Error	60.542	0.051
Standardized coefficients	Beta		0.722
	t	6.361	7.737
	Sig.	0.000	0.000

The conclusions that we can draw from these results and the applicability of this model in valuation problems is further discussed in chapter 7, but before that we will construct another model using only the apartment prices as a predictor. It is important to remember that the residual land values were constructed using apartment price data and construction cost data. The apartment price dependent model serves as a comparison point for this model when we follow to the conclusions. If the apartment prices are a better predictor for the lot prices alone, then we must question the usefulness of the residual valuation method when it comes to the estimation of vacant lot sales prices.

6.3.3 Model 2 (Apartment price)

As described above, in addition to the main regression model, a comparison model was constructed by using the district average apartment price as an estimator. The aim of this analysis is to find out if the apartment price alone is a better predictor to vacant lot sales prices than the residual land values.

First, let us look at the two datasets in this analysis. When the data is analyzed in a scatterplot (figure 4) it is clearly visible that the two variables have a very similar relation than the lot prices and residual land values. The relationship seems linear or at least there is no indication of a nonlinear relationship. The data is heteroscedastic and does not fulfill the condition of equal variances. In these terms the apartment price data is on a similar level of reliability than the residual land value data for which the same plot was introduced in the previous chapter.

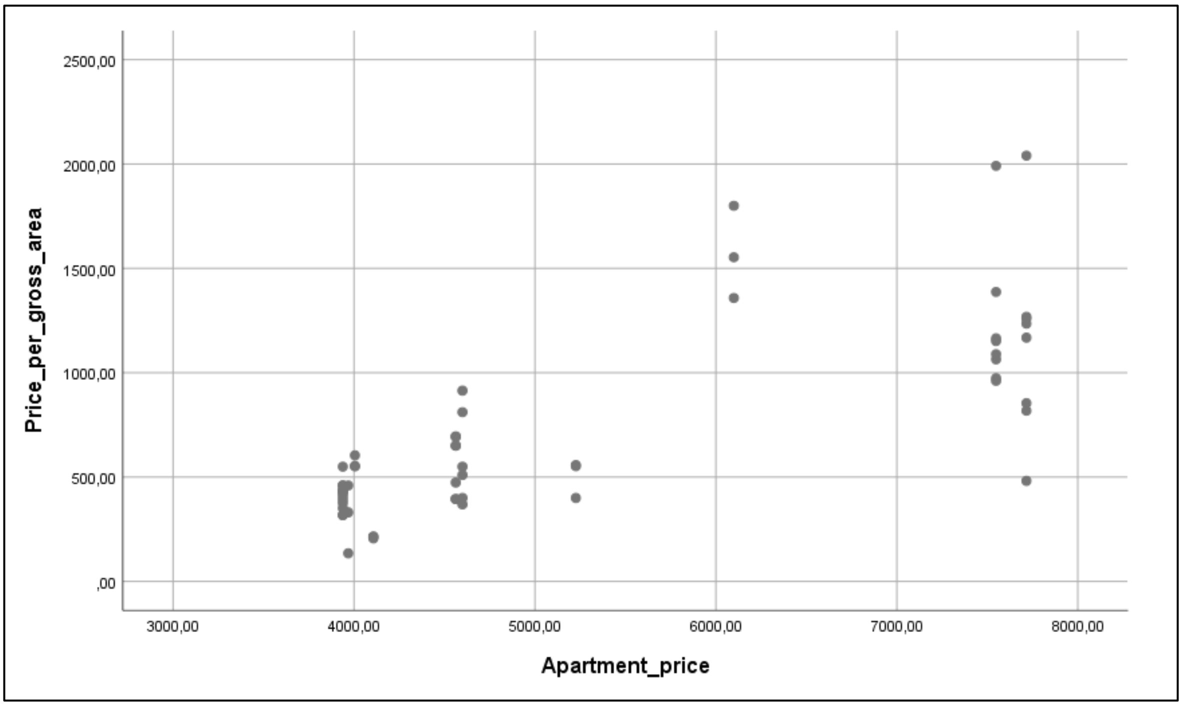

Figure 4: Scatterplot, Lot price per gross area (€/m2) & Apartment price (€/m2)

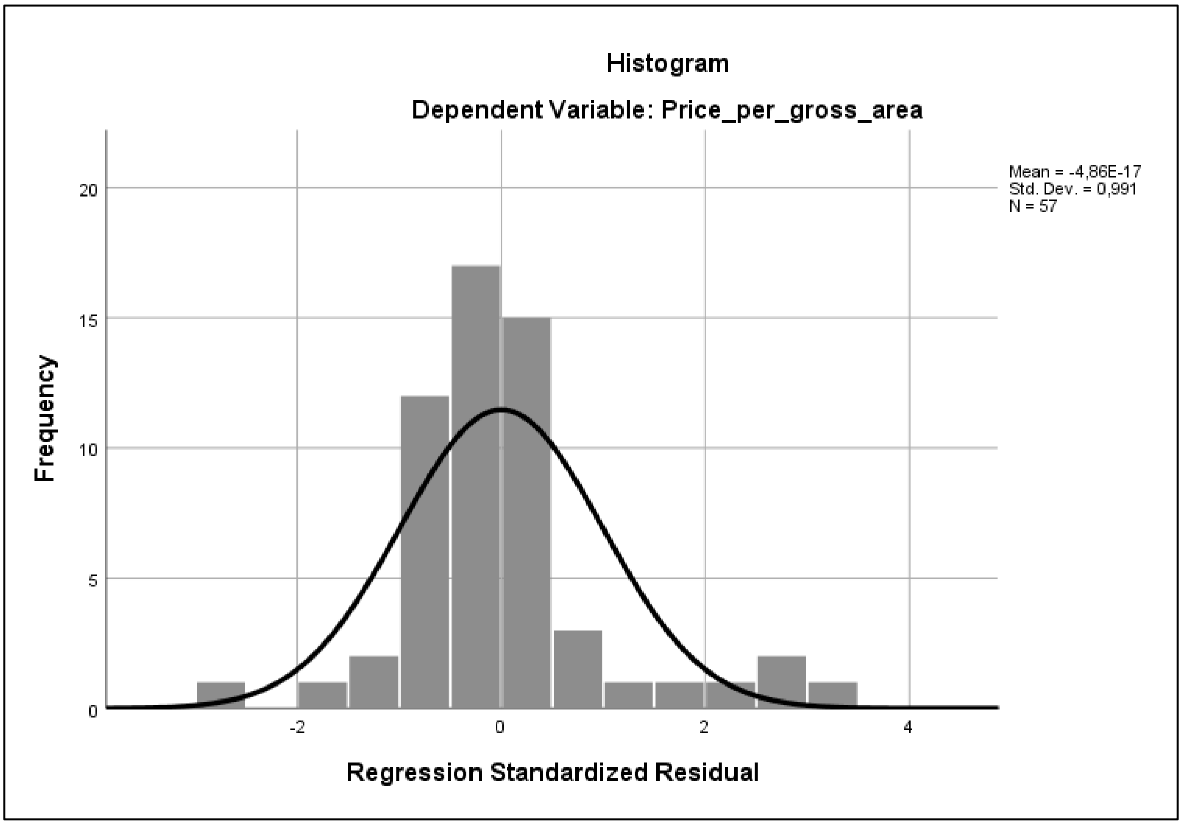

Figure 5: Histogram, Regression Standardized Residual Distribution (Model 2)

When we look at the normal distribution of the data (figure 5) we can see that the data is mostly normally distributed with some left inclination or skewness on the histogram and a higher than normal peak in the center. The skewness is minor and as the sample size is quite low it is hard to make any conclusions out of this skewness. Based on the histogram we can assume that the data is fairly normally distributed in a similar manner than with the residual land values. When we analyze the residual scatterplot (figure 6) we can see that the situation is similar than with the previous data. The dots in the scatterplot are equally and randomly distributed and there is no pattern that shows a relationship between the predicted values and residuals. We have no reason to doubt the linearity assumption. The data is homoscedastic in the residual scatterplot, which means that the linear least squares regression analysis can be used to provide linear estimators that can describe the statistical relationship of the two datasets.

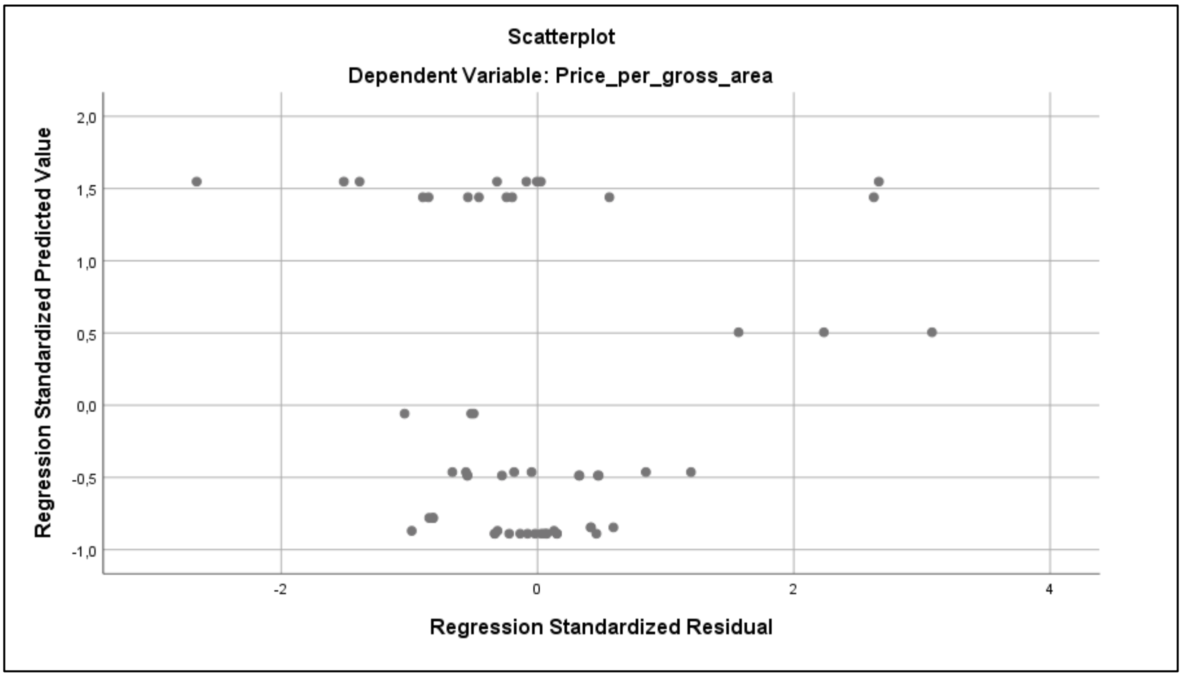

Figure 6: Residual Scatterplot (Model 2)

Next, we will go through the constructed regression model. The model has one dependent variable (Lot price, € / gross area) and one predictor (Apartment price, € / net area). The model does not have any other predictors and can be summarized as the following regression function:

Lot price = B1 x Apartment price + C (constant)

The coefficients and model summary are shown in table 21. The model has an adjusted R square of 0.581 which means that the district average apartment price as a predictor can predict 58.1% of the variation in vacant lot sales prices. This is a slightly higher, but similar value than in model 1 that used residual land values (50.1%).

Table 21: Regression model summary (Model 2)

Model Summary			
R	0.767		
R square	0.588		
Adjusted R square	0.581		
Std. Error of the estimate	292.696		
Durbin-Watson	1.248		
Coefficients			
		(Constant)	Apartment price
Unstandardized coefficients	B	-464.834	0.224
	Std. Error	139.651	0.025
Standardized coefficients	Beta		0.767
	t	-3.329	8.861
	Sig.	0.002	0.000

We can also compare the models 1 and 2 by looking at the normal probability plots of the models (figure 7 and figure 8). The aim of the probability plot is to show if the expected and observed cumulative probabilities of the model are the same or how much they differ. If the expected and observed probabilities differ significantly the model is considered as less reliable. The probability chart (figure 8) for apartment prices has a slightly better distribution (dots distributed closer to linear line) than the chart with residual land values (figure 7). This result in addition to the explanatory rate of the models shows that the model based on the apartment prices alone is a better and more reliable predictor to the vacant lot sales prices than the model with residual land values.

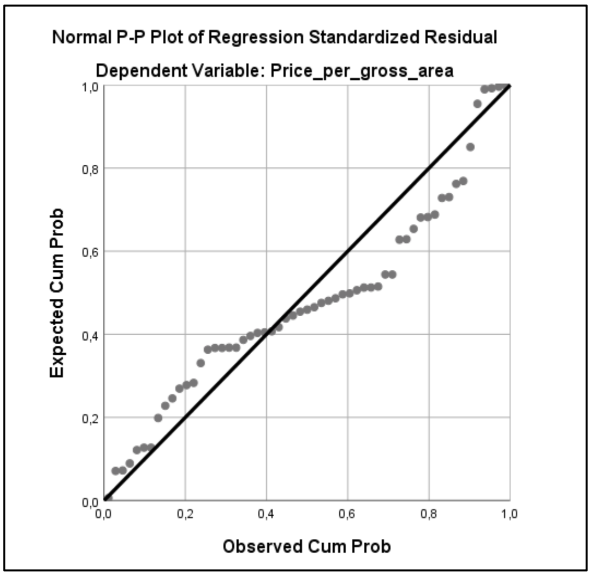

Figure 7: Normal probability distribution (Model 1)

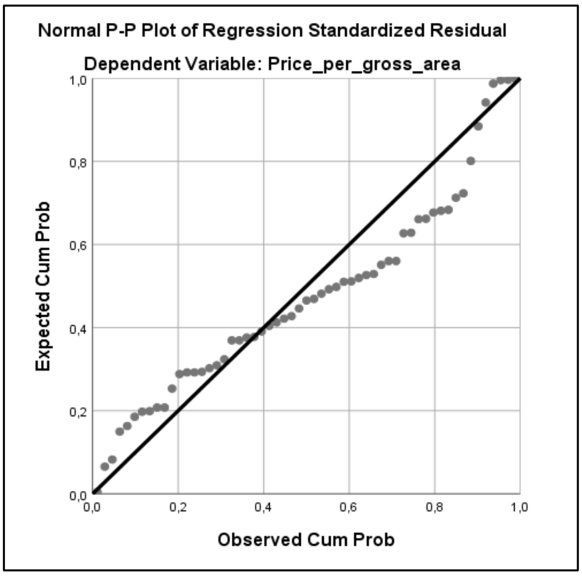

Figure 8: Normal probability distribution (Model 2)

Based on the results it seems that the apartment prices alone have a higher explanation rate and a better explanation reliability for the vacant lot sales prices than the estimated residual land value. As the residual land value uses the same apartment price data with the addition of the construction costs it is highly likely that the explanation ability of the model is mostly due to the explanation ability of the apartment prices. Hence the usefulness and the applicability of the residual land valuation method in the prediction of vacant lot sales prices is questionable. The implications of these results are further analyzed in the next chapter where we will draw the conclusions for the research.

7 Conclusions

The purpose of this research was to study the residual land valuation method that is commonly used by property developers to estimate the investment value of vacant land when considering different types of development projects. The residual method is predominantly used for performing profitability analyses of development and redevelopment projects or either to define the value of the land component or the value of the buildings as a part of developed property (Skarzyński 2006, p.1). Our main objective in this study was to find out if there are large differences in residual land value in the Helsinki region and is this residual value represented in the vacant lot prices in the area. Residential land values have been studied widely and also in the Helsinki region, but there were no examples in the literature where the residual valuation method had been used for this purpose.

The research was structured to three main topics that were approached with different research methods and all of them had individual results that were analyzed separately. In the first part we aimed to find out if there are significant differences in predicted housing development profitability or in residual land values in different districts and apartment types in the Helsinki region, when the residual land valuation method is used in the estimation. In the second part of the research we used correlation analysis to better understand the relationships between the different variables (apartment prices, construction costs, vacant lot sales prices) and the valuation results (expected development profits & residual land values). Finally the third part of the research used regression analysis to find out how well the residual land value estimates can predict the vacant lot sales prices. As the research consisted of three separate parts we will first go through the conclusions from each individual part independently and then summarize the conclusions of these results so that we can get a broader perspective of all of the results and how they relate to the literature findings.

The results of the first part showed us that there are significant differences in expected development returns between the districts (28.4%-units) and between the apartment types (15.2%-units). The average expected development returns in the highest ranking district, Kalasatama were 45.3%, while they were only 15.7% in the lowest ranking area, Eestinlaakso. The residual land values were also calculated using the same residual valuation formula, but with the assumption that development profit requirements are the same for all areas (27.5%). Here we made the assumption that the difference between the apartment prices and construction costs are capitalized entirely to land value without a difference in development profitability between areas as suggested in the Ricardian land rent theory (Ricardo 1809). With these assumptions our results show that there are significant differences in residual land values between the districts (3,037 € / net area) and between the apartment types (913 €/ net area). The value in Kalasatama, which had the highest average residual land value, was 21 times higher (3,180 € / net area) than in Kivistö, which had the lowest average residual land value (143 €/ net area). The range in residual land values is significantly larger than in vacant lot sales prices that ranged from 297 to 2,198 €/ net area.

Based on the results of the first part of the research we can come to the conclusion that the price gap between apartment prices and construction costs is not entirely capitalized to land prices in the land market, but leads to large differences in expected development returns. The current vacant lot price differences between areas are significantly smaller than the differences in residual land value and this leads to higher expected development returns with the residual valuation method.

The reasons behind the results of the first part are intriguing and the aim of the second part was to find out where these differences come from by calculating the correlation between all of the parameters of the residual valuation formula. The results of the correlation analysis show that there is a significant positive correlation between all of the residual valuation parameters: apartment prices, new construction costs and lot prices. However the correlation between apartment prices and lot prices is significantly higher on average (0.71) than between apartment prices and construction costs (0.52). Based on these results it seems that the construction costs correlate with apartment prices but not to a full extent (absolute correlation) and this leads to a higher development residual in the areas where apartment prices are higher. Based on the theories of land value this residual should be the result of the value of location and should be capitalized to land values. However even though the vacant lot prices and apartment prices have a very high correlation the vacant lot price differences do not represent the entire residual value differences between the areas. This leads to a situation where the residual land value is not capitalized entirely to lot prices but some of it is capitalized in higher expected development returns. When the correlation analysis is extended to include the correlation between the individual parameters and the residual valuation results this effect is further confirmed. Apartment prices have a very high correlation with both the expected development returns and the residual land values while construction costs have a relatively low correlation with the residual land values and a very low correlation with the expected development returns. Based on this we can conclude that the expected development returns are mostly influenced by apartment prices and higher expected development returns can be capitalized in the areas where the apartment prices are highest. Even though the construction costs also rise with the apartment prices the effect of the higher construction costs on development profitability is minimal compared to the higher overall residual that can be gained from the higher apartment prices with the current lot price levels.

The first two parts of the research have shown us that there are significant differences between the residual land values of the observation areas and that these differences are not entirely capitalized to lot prices but lead to higher expected development returns. In the third part of the research the aim was to measure how different the estimated residual land values are from the vacant lot prices. This was done by using regression analysis where the individual vacant lot prices are being predicted with the residual land value as a predictor. The regression model that used residual land values to predict lot prices could predict 51.2% of the variation in lot prices while a comparison model that used apartment prices as a predictor had a slightly better explanation rate of 58.1%. Based on these results the residual land value is not a very good predictor to vacant lot sales prices as the model that uses apartment prices alone can achieve better results. The main difference between these models is that residual land value takes into account the construction costs while the other model only uses apartment prices. These results show that vacant lot prices seem to be determined in the market more based on the apartment prices than the residual land values. These results confirm the results of the first two parts of the research that the vacant lot sales prices do not seem to be determined in the land market by the residual land values and this can be one of the reasons that leads to the differences in expected development returns as the higher residual is not entirely capitalized to higher lot prices.

There were no previous research results to be found in the literature that could be compared to the results of this study. The main reason for this is that the residual valuation method is not commonly used in research that is focused on regional land values. However there are a number of papers published that focus on residential land values and their underlying causes as well as the dynamics of housing prices and their relation to residential land values. Some of the older governing theories of the field are somewhat contradicted to the study results. For example the Ricardian rent theory assumes that the differences in property values are capitalized in differences in land value and rising property prices also result in rising land prices (Ricardo 1809). This assumption was tested by Ooi & Lee (2006) who found out that there is a causal relationship that verifies this systemic relationship from property prices to land prices. The residential property prices and the urban land prices are integrated in the long term and this should lead to the increase of land values in the areas where apartment prices increase. Our study however has shown that in the Helsinki region the residual that can be gained from different locations is not represented entirely in vacant lot prices.

There is evidence from studies that the real world land market differs from the theoretical framework in some key aspects. The land market is not a very efficient market especially as land is not traded frequently and there is a constant lack of relevant and observable market information (Evans 2004, p.60). The lack of information may cause problems in the estimation of sales prices and it emphasizes the effect of particular buyers and sellers in the market. The buyer and seller characteristics were studied by Isakson (1997) who found out that different types of buyers and sellers had a significant effect on the land prices. For example the highest prices were paid by governmental entities, while the lowest prices were paid by individuals. The combinations of the buyer and seller were also significant and the prices paid between two individuals were significantly lower than the prices paid between two corporations. (Isakson 1997, p.113)

The inefficiency of the land market combined with the significance of buyer and seller characteristics can explain why the vacant lot sales prices differ from the residual land values. If a transaction between two corporations results in higher valuation of the land it can result from the fact that the professional developers can better estimate the residual value of the land than other landowners. When the balance of power and information is equal between the buyer and the seller this can result in higher land prices as the latent value of the land is better understood between the buyer and the seller. This also leads to the conclusion that if the residual value is not well understood by one of the participants, it can lead to undervaluation of the land. The differences in expected development returns that were observed in this study are hard to explain if at least some vacant lots were not sold under their development value that is represented by the residual valuation method.

There are also recent studies from the Helsinki region that support the results of this study. The adjustment of housing prices and residential land prices was studied by Oikarinen (2014) who found out that land prices react much more slowly to shifts in market situations than apartment prices. Oikarinen argues that the reason behind this lag is the same than with Evans (2004) that there is not enough publicly available sales data to make reliable predictions of land prices and this leads to problems in determining the land prices. (Oikarinen 2014, p.17). The explanation presented by Oikarinen (2014) and Isakson (1997) can explain most of the results of this study. If the land owners do not have the adequate tools and information to make good predictions of the highest possible land value that can be obtained, then the increases in apartment prices might not immediately affect the land prices and land is sold under its highest possible value. When this is combined with a developer that has better valuation tools and a good understanding of the residual value of the land the developer can seek out the land that has the best residual value compared to its sales price and focus development on these areas that give the highest expected development returns. Based on the results of this study, these areas seem to be the areas with the highest apartment prices, since the construction costs do not increase in the same proportion with the apartment prices and higher apartment prices result in higher residual land values. With these results it is to be expected that the developers are most willing to initiate new construction in these areas that have the highest apartment prices, since the land prices indeed seem to lag behind and this would result in the highest expected development returns.

8 Discussion

The emphasis on this research was to study the differences in residual land values and vacant lot sales prices in the different districts in Helsinki and the surrounding Helsinki region. The results of the study were interesting and lead to the conclusion that there are significant differences in the residual land values between the observation areas and that these differences are not entirely represented in the vacant lot sales prices. We can find explanations for the results based on previous research and most of them tend to explain this phenomenon with the inefficiency of the land market, availability of relevant market information and the complexity in the determination of land values. However there are other aspects that may have affected the results and they are in large part caused by the way this study was conducted. We will next go through the major problems of this study and the ways that these problems may have affected the study results. From here we will suggest some ways for further study that could be used to provide more reliable results of the same topic as well as to find out some of the causes for the results.

One of the major issues in this study was the availability of relevant data with sufficient amount of detail. The residual valuation method is somewhat complex to use as it requires many different types of data that have to represent the same individual property or area. To calculate the residual land value we need data of the property value (which was here represented with apartment prices) and the construction costs required to develop the property. If we want to compare this to actual vacant lot sales prices we would preferably need to have the sales price of the specific lot that is being valued. This is less of a problem when an individual development project is being valued using the method as this information can be acquired and the valuation is often done to determine whether the lot should be bought with the asking price or not. However it is much harder to obtain this information regionally as the lots being sold are not yet developed and there is little information available on their development possibilities or the apartment prices that could be obtained from the specific lots. For this reason we had to use regional averages of both lot prices and apartment prices and this has most probably caused major inaccuracies in the results. In further study this problem could be approached differently by studying individual development projects from different areas instead of the average prices of the areas. This kind of study would require detailed data of the development projects such as the actual construction costs, apartment sales prices and the price that was paid for the vacant lot. This more detailed study would give more reliable information on the actual profitability of the development projects in different areas. The residual valuation method could be used here to determine the residual land value of the development property which could then be compared to the price that was paid for lot. This kind of study could provide much more reliable results but it would require extensive amount of research and many different development projects to gain regional information of the differences between areas in the Helsinki region. This was one of the main reasons that this approach was not taken in this research as the regional differences were of particular interest and the amount of time and data required with this approach to represent the regional differences is significantly higher.

Another major problem in this study was that we do not have sufficient understanding of the developer's perceived risk in different development projects and how the regional area affects the risk of the project. In this study we had the assumption that if the residual land values are represented entirely in the lot sales prices, then the land value represents the Ricardian rent theory (Ricardo 1809). This assumption is only justified if the developer's required return is represented correctly in the residual valuation method. The residual value of the lot in the mind of the developer is not only the difference between the end product sales price and the construction costs. The developer must take into account the risk level of the project and a higher risk project requires a higher expected return to compensate for the risk. Here we faced a major challenge as there is very little public information about the required returns of the developers and while there are studies on the subject such as those of Crosby, Devaney & Wyatt (2018), Curry (2013) and Dong & Sing (2014) they do not provide any practical tools that could be used to compensate for this lack of information. This information could be obtained from the developers if they were willing to provide it, but the nature of the information is often considered a business secret which is one of the reasons that this information could not be obtained for this research. Due to lack of information this research was done assuming that the different areas have the same required development return and this may have distorted the results.

As the information of the required returns is hard to obtain, the development risk levels in different regional areas could be studied individually in further study. This could be done by analyzing different development projects with the residual valuation method and calculating the expected development returns in a more detailed level. Also the price changes in apartment prices in different areas could be used to represent the risk relating to the sales prices of the apartments that most likely has a significant impact on the development risk. The results of this study show that the most important factor that determines the expected development return is the sales price of the apartments. A study of apartment price volatility in different areas could give more light on the possible risk levels on of developing apartments in different areas and this could be used as a framework to categorize different areas with different expected risk levels. This information could then be used in the residual valuation format to better analyze if the differences in the expected development returns are actually based on differences in development risk in the areas or on land being sold under value.

The developer's perceived development profitability of the different areas could be studied indirectly by assuming that the developers are most willing to initiate new development in areas that are most profitable. This could be approached for example by comparing the construction volumes of the districts to the expected development returns calculated with the residual valuation method. The developers should rationally prioritize projects that have the best return compared to the risk of the project. The areas where land is being sold most under its highest development value should be the areas that the developers are most interested in and this should show in construction volumes as a relatively higher amount of new projects than in other areas. Of course this suggested study is indirect in nature and it cannot provide very detailed information on the required returns of the developers but it could be used to further analyze the results of this study. If the construction volumes are indeed significantly higher in the area where the highest expected development returns were calculated, then we could verify that these areas most likely have higher residual land values than the lot prices and land is being sold under its development value which is being capitalized as higher returns for the developers.

These suggestions for further study still leave one final question open: if land is being sold under its development value in some of the areas, then what is the reason behind this? Can it be explained entirely with the inefficiency of the market and the availability of information combined with problems in valuation as suggested by Oikarinen (2014) and Evans (2004)?

Could there be other understandable causes for the land sales prices? One of the interesting aspects that has not yet been studied in this context is the influence of municipalities and their strategic decisions in land management policies. For example the city of Helsinki is a major landowner in Helsinki and a significant amount of new development land is either owned by the city of Helsinki or acquired by it prior to the new area development. For this reason the decisions made by the city can have a major impact on the lot sales prices. If the city sells land under its development value in an area it will instantly affect the land market and other landowners in the area as well. The city of Helsinki could prefer to sell the lots in a strategic development area at a lower price to speed up development and incentivize developers to prioritize projects in that area. This could be more profitable for the city in the long run than to maximize the revenue from land sales in the short run. The city has to invest in new development area infrastructure in advance and the strategic gains from the area only come after the area is sufficiently developed.

However, the EU legislation prohibits the sale of land under market price to buyers that are market subjects, such as property developers (Vedenoja 2017. p. 22). With the current legislation the intentional sale of development land under its market value is prohibited and thus it is reasonable to assume that the possible under value selling is caused by inaccurate valuation rather than strategic decisions by the city. Nevertheless the decisions of the city can have a large impact on the land market and many cities and municipalities in Finland use somewhat simple and inaccurate methods in land valuation such as price-zone-based models (Vedenoja 2017. p. 83). The accuracy of these zone-based prices can be questioned if the residual values of the lots have significant differences inside a price-zone. The EU commission has recommended the use of open auctions or separate professional valuations as the best ways to ensure that land is sold at market value (Vedenoja 2017. p. 22). If the developers are willing to pay more for the land in an open auction, then the zone-based prices can lead to undervaluation of the land. This is something that could be studied by taking an approach to the land management and lot pricing strategies of one or more of the major cities in the Helsinki area. This further study could reveal more information on the actors of the land market and explain some of the underlying reasons behind the results of this study. It is possible that the lot sales prices have less differentiation between areas than the residual land values at least partly because the lot sales price is determined with methods that do not consider all the differences between the lots. This can lead to a situation where a detailed analysis made by the developer can reveal the best deals that are available in the land market and enable the exploitation of these differences so that the developer can achieve higher development returns. The open auction policy that was recommended by the EU commission could be a better way to determine lot prices than price-zone based models. It would be worth a further study to compare if the residual values and vacant lot sales prices are closer together in cities where open auction policy is predominantly used in the sale of city-owned development land.

References

Adair, A., S. McGreal (1988). The Application of Multiple Regression Analysis in Property Valuation. *Journal of Valuation*. Vol. 6 Issue. 1 pp. 57 – 67. Available: http://dx.doi.org/10.1108/eb008022.

Ahlfeldt, Gabriel M. (2008). *If Alonso was Right: Residual Land Price, Accessibility and Urban Attraction.* University of Hamburg, Department of Economics. MPRA Paper No. 11707. Available: http://mpra.ub.uni-muenchen.de/11707/.

Ahvenniemi, H., Pennanen, K., Knuuti, A., Arvola, A., & Viitanen, K. (2018). Impact of Infill Development on Prices of Existing Apartments in Finnish Urban Neighborhoods. *International Journal of Strategic Property Management*. Vol. 22(3). pp. 157–167. Available: https://doi.org/10.3846/ijspm.2018.1540

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent.* Cambridge, Massachusetts: Harvard University Press. ISBN-13: 9780674537002.

Archer, W. & D. Ling (1997). The Three Dimensions of Real Estate Markets: Linking Space, Capital and Property Markets. *Real Estate Finance*. Vol. 14:3. pp. 7–14. ISSN 0748-318X.

Boyd, T., S. Boyd (2012). Valuing the Land Component of Improved Investment Property. *Journal of Property Investment & Finance.* Vol. 30 Issue: 4. pp. 338–353. Available: https://doi.org/10.1108/14635781211241752.

Capozza, D., Sick, G. (1994). The Risk Structure of Land Markets. *Journal of Urban Economics.* Vol. 35. pp. 297–319.

Capozza, D.R., P.H. Hendershott, C. Mack & C.J. Mayer (2002). *Determinants of Real House Price Dynamics.* NBER Working Paper Series. Available: http://www.nber.org/papers/w9262.

Coleman, C., Crosby, N., McAllister, P., & Wyatt, P. (2012). Development Appraisal in Practice: Some Evidence from the UK Planning System. *Journal of Property Research*. Vol. 30. No. 2. pp. 144–165.

Crosby, N., S. Devaney, P. Wyatt (2018). The Implied Internal Rate of Return in Conventional Residual Valuations of Development Sites. *Journal of Property Research*. ISSN: 0959-9916. Available: https://doi.org/10.1080/09599916.2018.1457070.

Cunningham, C.R (2006). House Price Uncertainty, Timing of Development and Vacant Land Prices: Evidence for Real Options in Seattle. *Journal of Urban Economics*. Vol.59. pp. 1–31. Available: www.sciencedirect.com.

Curry, B.J (2013). The Trouble with Rates in the Subdivision Development Method to Land Valuation. *The Appraisal Journal*. Vol. 81. pp. 151–171.

Darlow, C. (1982). *Valuation and Development Appraisal*. Estates Gazette. London. ISBN-13: 978-0728200753.

DiPasquale, D., & W.C. Wheaton (1992). The Markets for Real Estate Assets and Space: A Conceptual Framework. *Journal of the American Real Estate and Urban Economics Association*. Vol. 20:1. S. 181–197. ISSN 0092-914X.

Dong, Z., T.F Sing (2013). Developer Heterogeneity and Competitive Land Bidding. *Journal of Real Estate Finance Economics*. Vol. 48. pp. 441–466. DOI 10.1007/s11146-013-9408-1.

Evans, A. W. 2004. *Economics, Real Estate, and the Supply of Land*. Oxford, UK. Malden, Mass. Blackwell. ISBN: 1-4051-1862-8.

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods.* California: Sage Publications. ISBN: 0-8039-4540-X

Fujita, M. 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge: University of Cambridge. ISBN-13: 978-0521396455.

Gallimore, P., M. Fletcher, M. Carter (1996). Modelling the Influence of Location on Value. *Journal of Property Valuation and Investment*. Vol. 14 Issue 1. pp. 6 – 19. Available: http://dx.doi.org/10.1108/14635789610107444.

Geltner, D. M., N.G Miller (2000). *Commercial Real Estate Analysis and Investment*. South-Western Publishing. ISBN 0-324-13676-5.

Goldberger, Arthur S. (1964). *Econometric Theory*. New York: John Wiley & Sons. pp. 238–243. ISBN-13: 978-0471311010

Greenhalgh, P.M., R.S Bendel (2015) An Investigation of Development Appraisal Methods Employed by Valuers and Appraisers in Small and Medium Sized Practices in Brazil. *Journal of Property Investment & Finance.* Vol. 33 Issue: 6. pp. 530–547. Available: https://doi.org/10.1108/JPIF-03-2015-0019.

Grissom, T., Berry, J. & Lim, L. (2010). Economics of Development Strategies Utilizing Option and Portfolio Analytics. *Journal of European Real Estate Research*. Vol. 3. No. 2. pp. 117–137.

Haahtela (2015). *Rakennuksen hinnan arviointiohje 2015*. Haahtela-kehitys Oy. Saatavissa: https://www.haahtela.fi/fi/julkaisut/painojulkaisut.

Haahtela, Y. & J. Kiiras (2014). *Talonrakennuksen kustannustieto 2014*. Helsinki: Haahtela-kehitys Oy. ISBN: 978–952-5403-22-0.

Havard, T. (2014). *Financial Feasibility Studies for Property Development: Theory and Practice*. Routledge. New York. ISBN-13: 978-0415659178.

International Valuation Standards Council (2017). *International Valuation Standards 2017*. Norwich: Page Bros. ISBN: 978-0-9931513-0-9.

Isakson, H.R (1997). An Empirical Analysis of Determinants of the Value of Vacant Land. *Journal of Real Estate Research*. Vol. 13. Issue 2. pp. 103–114.

Jud, G.D. & D.T. Winkler (2002). The Dynamics of Metropolitan Housing Prices. *Journal of Real Estate Research*. Vol. 23:1. pp. 29–45. ISSN 0896-5803.

Kester, W. (1984). Today's Options for Tomorrow's Growth. *Harvard Business Review*. Vol. 62. Issue 2. pp. 153–160.

Laakso, S. (1997). *Urban housing prices and the demand for housing characteristics: a study on housing prices and willingness to pay for housing characteristics and local public goods in Helsinki Metropolitan Area*. Helsinki: ETLA, the Research Institute of the Finnish Economy, 1997. Vol. 275 No. 27. ISBN 951-628-260-1.

Lujanen, M. (2004). *Housing and Housing Policy in the Nordic Countries*. Nordic Council of Ministers, Copenhagen 2004. Akaprint as, Århus. ISBN: 92-893-1027-8.

Lönnroth, T. (2017). *The Value of Urban Land and Residential Land Leasing in Helsinki.* Master's Thesis. Aalto University School of Business.

McDonald, R. and Siegel, D. (1986). The Value of Waiting to Invest. *Quarterly Journal of Economics.* Vol. 101. pp. 707–727.

Metsämuuronen, J. (2010). *Tutkimuksen tekemisen perusteet ihmistieteissä*. e-kirja Tutkija-laitos. 4. Edition. Helsinki: International Methelp Oy. ISBN-13 978-952-5372-29-8.

Mills, E.S (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *The American Economic Review, Papers and Proceedings*. Vol. 57. No. 2. pp 197–210.

Muth, R.F. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press, Chicago. ISBN-978-022-655-4136

Myers, S. (1984). Finance Theory and Financial Strategy. *Interfaces*. Vol. 14 No. 1.

Mäkinen, R. (2017) *The Effect of Residential Building's Present Price and Maintenance Backlog on the Price of Its Dwellings.* Master's Thesis. Aalto University School of Engineering.

Newell, M. (1989). Development appraisals. *Journal of Property Valuation and Investment.* Vol. 7. No. 2. pp. 126–133."

Ogunbayo, O.T, A.A Odebode, J.B Oyedele & O.T Ayodele (2018). The Significance of Real Estate Development Process Analysis to Residential Property Investment Appraisal in Abuja, Nigeria. *International Journal of Construction Management*. DOI: 10.1080/15623599.2017.1423164"

Oikarinen, E. (2007). *Studies on Housing Price Dynamics.* Doctoral Thesis, Series A-9: 2007. Turku School of Economics. Esa Print Tampere. ISBN: 978-951-564-507-4.

Oikarinen, E. (2014). Is Urban Land Price Adjustment More Sluggish than Housing Price Adjustment? Empirical Evidence. *Urban Studies.* Vol 51(8). pp. 1686–1706. Available: https://doi.org/10.1177/0042098013497409

Oikarinen, E., R. Peltola & E. Valtonen (2015). Regional Variation in the Elasticity of Supply of Housing, and its Determinants: The Case of a Small Sparsely Populated Country. *Journal of Regional Science and Urban Economics.* Vol. 50. pp. 18–30. Available: www.elsevier.com/locate/regec

Ooi, J.T.L., S.T Lee (2006). *Price Discovery between Residential Land & Housing Markets.* American Real Estate Society.

Owusu-Ansah, A. (2011). A Review of Hedonic Pricing Models in Housing Research. *Journal of International Real Estate and Construction Studies*. Vol. 1:1. pp. 19–38. ISSN 2153-6813.

Pagourtzi, E., V. Assimakopoulos, T. Hatzichristos, & N. French (2003). Real Estate Appraisal: A Review of Valuation Methods. *Journal of Property Investment and Finance*. Vol. 21:4. pp. 383–401. ISSN 1463-578X.

Pejic, M., B. Bozic, D. Milicevici, S. Marosani. (2013). The Use of Multiple Linear Regression in Property Valuation. *Geonauka.* Vol. 1(1). pp. 41–45. doi:10.14438/gn.2013.06

Peltola, R., J. Väänänen (2006). *Asuntotontin hinta*. National Land Survey of Finland. Vol. 105. ISBN 951-48-0198-9

Pendleton, E. C. (1965). Statistical Inference in Appraisal and Assessment Procedures. *The Appraisal Journal.* Vol. 33. pp. 73–82.

Pennanen, A. (2019). Interview on construction cost modeling using TAKU-model. Haahtela-kehitys Oy.

Pindyck, R. (1991). Irreversibility, Uncertainty and Investment. *Journal of Economic Literature.* Vol. 29. Issue 3. pp. 1110-1148.

Renshaw, E. F. (1958). Scientific Appraisal. *National Tax Journal.* Vol. 11. pp. 314–322.

Ricardo, D. (1809). *On the Principles of Political Economy and Taxation*. Dover Publications. 5.12.2004 edition. ISBN: 978-0486434612.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy*. Vol. 82. No 1. pp. 34–55. Available: http://links.jstor.org/sici?sici=0022-3808%28197401%2F02%2982%3A1%3C34%3AHPAIMP%3E2.0.CO%3B2-U.

Skarzyński, A. (2006). Residual method of property valuation. *Technological and Economic Development of Economy*. Vol. 12:3. pp. 253–256. Available: https://doi.org/10.1080/13928619.2006.9637749.

Teittinen, E. (2019). Interview on construction cost modeling using TAKU-model. Haahtela-kehitys Oy.

Vedenoja, S. (2017). *Effect of state aid rules on establishing market compliance when selling and leasing municipality owned land.* Master's thesis. Aalto University School of Engineering.

Vimpari, J. (2014). *Is There Hidden Value in Real Estate Investments? Real Options Analysis Provides Rationale to Contingent Investment Decisions.* Doctoral Thesis. Aalto University School of Engineering. ISBN 978-952-60-6007-1.

Vimpari, J., J.K. Kajander & S. Junnila (2014). Valuing Flexibility in a Retrofit Investment. *Journal of Corporate Real Estate*. Vol 16. No 1. pp. 3–21. Available: DOI 10.1108/JCRE-07-2013-0017.

Vimpari, J., S. Junnila (2014). Value of Waiting: Option Pricing as a Tool for Residential Real Estate Fund Divestment Management. *Property Management*. Vol 32. No 5. pp. 400–414. Available: DOI 10.1108/PM-03-2014-0013.

Vimpari, J., S. Junnila (2015). *Real Options Analysis as a Decision-Making Tool: a Preliminary Investigation of the Real Estate Investment Executive Perspective.* Aalto University Publications. Department of Real Estate, Planning and Geoinformatics. ISBN 978-952-60-6206-8.

Virenius, O. (2014). *Advantages of Real Option Analysis in Urban Development Project.* Master's Thesis. Aalto University School of Engineering.

Von Thünen, J.H. (1826). *The Isolated State in Relation to Agriculture and Political Economy*. Palgrave Macmillan. 2004 edition. ISBN 13: 9780230222519.

Wilson, J. H., Keating, B. P. & Beal-Hodges, M. (2012). *Regression Analysis: Understanding and Building Business and Economic Models Using Excel*. Business Expert Press. New York, N.Y. ISBN-13: 978-1-60649-435-6

Wolverton, M.L (1993). Case Study: An Alternative Technique to the Land Residual Method. *The Appraisal Journal*. Vol. 4. pp. 239–244.

Yao, H., F. Pretorius (2014). Demand Uncertainty, Development Timing and Leasehold Land Valuation: Empirical Testing of Real Options in Residential Real Estate Development. *American Real Estate and Urban Economics Association*. Vol. 42. Issue 4. pp. 829–868. Available: DOI: 10.1111/1540-6229.12052.

Zahirovich-Herbert, V., K.M Gibler (2014). The Effect of New Residential Construction on Housing Prices. *Journal of Housing Economics*. Vol. 26. pp. 1–18. Available: http://dx.doi.org/10.1016/j.jhe.2014.06.003.

Özdilek, U. (2011) Land Value: Seven Major Questions in the Analysis of Urban Land Values. *American Journal of Economics and Sociology*. Vol. 70. Issue 1. pp. 30–49.