Lappeenranta-Lahti University of Technology LUT
School of Business and Management
Degree program in Strategic Finance and Business Analytics

Melisa Manninen

Predicting the development of housing prices in Westmetro's phase 2 areas

Examiners: 1st Supervisor: Post-doctoral researcher Mariia Kozlova
2nd Supervisor: Post-doctoral researcher Jan Stoklasa

ABSTRACT

Author:	Melisa Manninen
Title:	Predicting the development of housing prices in Westmetro's phase 2 areas
Faculty:	School of Business and Management
Master's Program:	Strategic Finance and Business Analytics
Year:	2021
Master's thesis:	Lappeenranta-Lahti University of Technology LUT 98 pages, 25 figures, 27 tables, 8 appendices
Examiners:	Post-doctoral researcher Mariia Kozlova Post-doctoral researcher Jan Stoklasa
Keywords:	Housing price, price prediction, Westmetro, rail traffic, regression analysis

Investments in transport infrastructure are affecting land value and through that housing and commercial premise prices. The housing price of otherwise similar apartments might vary a lot according to location. When accessibility improves, housing prices can be expected to increase in the areas, but a relatively larger increase can be expected in the areas with moderate travel distance to business and service concentrations. In 2017, Westmetro started operating from Helsinki to Southern Espoo. As a result, Westmetro's station areas housing prices increased significantly. In December 2014, construction work for Westmetro's phase 2 in Espoo started and it is expected to start operating in 2023.

The objective of this study is to predict the housing price development in 2020-2023 for Westmetro's phase 2 areas Soukka, Espoonlahti, and Kivenlahti. The quadratic function of Ordinary least squares regression is used as a research method in this study. Data is collected from Hintaseurantapalvelu which is upheld by the Central Federation of Finnish Real Estate Agencies. Collected data includes 3204 observations based on realized real estate sales in 2009-2019 in a 1-kilometer radius from the Westmetro's phase 2's station areas.

The main findings of the study are that the overall housing price development for all areas is predicted to be significantly increasing in 2020-2023. There are large increases predicted for apartment buildings and terrace houses in all three areas in 2023, within a 1-kilometer radius from the metro station, despite the size or year built of the apartment. However, the price development for houses is predicted to be decreasing, which could be partially explained by the small number of observations and variation in price behavior in training set data.

The data analysis and results of this study give detailed information about the housing market in the observed areas for operators in the real estate business, individuals, seeking to buy or sell real estate, and for future research of a similar field. The results can be used in the evaluation of the value of the real estate and to support decision-making in investments, as well as in municipal zoning decisions.

TIIVISTELMÄ

Tekijä: Melisa Manninen
Tutkielman nimi: Ennuste asuntojen hintakehityksestä Länsimetron 2-vaiheen alueilla
Akateeminen yksikkö: Kauppakorkeakoulu
Pääaine: Strategic Finance and Business Analytics
Valmistumisvuosi: 2021
Pro Gradu -tutkielma: Lappeenrannan-Lahden teknillinen yliopisto LUT
98 sivua, 25 kuviota, 27 taulukkoa, 8 liitettä
Tarkastajat: Tutkijatohtori Mariia Kozlova
Tutkijatohtori Jan Stoklasa
Avainsanat: Asunnon hinta, hintaennuste, Länsimetro, raideliikenne, regressioanalyysi

Investoinnit liikenneinfrastruktuuriin vaikuttavat maan arvoon ja sitä kautta asuinkiinteistöjen ja liiketilojen hintoihin. Kahden muuten keskenään samanlaisen asunnon hinta voi vaihdella paljon riippuen sijainnista. Kun saavutettavuus paranee, voidaan asuntojen hintojen odottaa nousevan alueilla, mutta suhteellisesti asuntojen hintojen nousun voidaan odottaa olevan suurempaa alueilla, jotka sijaitsevat kohtuullisen matkustusetäisyyden päässä työ- ja palvelukeskittymistä. Vuonna 2017 Länsimetro aloitti liikennöinnin Helsingistä Etelä-Espooseen. Länsimetrohankkeen myötä asema-alueiden asuntojen hinnat ovat nousseet merkittävästi. Joulukuussa 2014 Länsimetron 2-vaiheen rakennustyöt aloitettiin Espoossa ja liikennöinnin suunniteltu aloitusajankohta on 2023.

Tämän tutkimuksen tavoitteena on ennustaa asuntojen hintakehitystä vuosille 2020–2023 Länsimetron 2-vaiheen alueilla Soukassa, Espoonlahdessa ja Kivenlahdessa. Tutkimusmetodina käytetään kvadraattista funktiota pienimmän neliösumman regressiomallista. Tutkimusaineisto on kerätty Hintaseurantapalvelusta, joka on Kiinteistövälitysalan keskusliiton ylläpitämä järjestelmä. Tutkimusaineisto sisältää 3204 toteutuneen asuntokaupan hintatiedot vuosilta 2009–2019, 1 kilometrin säteellä tutkimukseen valikoiduilta Länsimetron 2-vaiheen asema-alueilta.

Tutkimuksen tärkeimpinä tuloksina voidaan nostaa esiin, että kokonaisuudessaan asuntojen hintakehityksen ennustetaan olevan nouseva kaikilla kolmella alueella vuosina 2020-2023. Kerros- ja rivitaloasunnoille 1 kilometrin säteellä metroasemasta on ennustettu merkittävää hintojen nousua vuodelle 2023, riippumatta asunnon koosta tai rakennusvuodesta. Omakotitalojen hintakehityksen ennuste on kuitenkin laskeva. Tämä voi ainakin osittain selittyä ennustemalliin käytettyjen omakotitalojen havaintojen pienellä otannalla sekä vaihtelevalla hintakehityksellä.

Aineiston analyysi sekä tutkimuksen tulokset antavat yksityiskohtaista tietoa asuntomarkkinoista tutkituilla alueilla kiinteistöalan toimijoille, asunnon ostosta tai myynnistä kiinnostuneille yksityishenkilöille sekä aiheeseen liittyville tuleville tutkimuksille. Tutkimuksen tuloksia voidaan hyödyntää kiinteistöjen arvon määrityksessä ja päätöksenteon tukena, esimerkiksi sijoitus- tai kaavoituspäätöksissä.

ACKNOWLEDGEMENTS

Studying at LUT has been a memorable experience and as I am writing the final words of this thesis, I would like to thank everyone who has supported me during this journey.

First of all, I would like to thank my supervisors Mariia Kozlova and Jan Stoklasa for their valuable insights as well as for guiding and supporting me through this process. Also, I would like to express my gratitude to Kiinteistömaailma Espoonlahti, Espoon Kodit Oy, and the Central Federation of Finnish Real Estate Agencies for collaboration and providing the data for this thesis.

Lastly, I would like to thank my family and friends for their ongoing encouragement and support during my studies and this thesis.

Helsinki, June 2021

Melisa Manninen

Table of Contents

List of figures

List of tables

1 INTRODUCTION

In this chapter, the topic of the thesis is introduced. Background of the study, research problem, objectives, data, and method are presented. Lastly, the structure of this thesis is introduced.

1.1 Background of the study

An apartment could be considered essential for a person as everyone needs to live somewhere and as a commodity, it is not easily changeable as the apartment itself cannot be moved (Laakso & Loikkanen 2004, 251). Public transportation infrastructure and the housing market are strongly correlated as the changes in transportation infrastructure can create externalities that affect housing prices. (Dubé et al. 2013) Investments in transport infrastructure are affecting land value and through that on housing and commercial premise prices. When accessibility improves, housing prices are expected to increase in the areas that benefit from the investment with better access to business and service concentrations. (Laakso et al. 2016, 431) Households and companies are willing to pay more for the location with the improved traffic system. As a result, the price for land increases in the areas, and attraction of the area for companies and households grow resulting increase in workplaces and population. (Laakso et tal. 2016, 440; Mulley et al. 2016)

There are different traffic system projects under construction in Finland and the most significant of them are Westmetro's phase 2 in Espoo, tram line in Tampere, and Jokeri light rail in Helsinki and Espoo. Metro started first operating in Helsinki in 1982, from the Central railway station to the Eastern suburbs of Helsinki and the metro line has been expanded during the 1990s. In May 2008 Espoo and Helsinki city councils decided about the construction of the Westmetro's phase 1. Construction work for phase 1 started in December 2009 and it started operating in November 2017. (Länsimetro 2020a) Westmetro is an extension to the metro line in Helsinki and it connects Southern Espoo to the metro line (Länsimetro 2020b). In June 2012 city council of Espoo accepted a project plan for Westmetro's phase 2 and the decision about the construction was made in September 2014 (Espoo 2012; Espoo 2014). In December 2014, construction work for phase 2 started. At the time of the start of the construction, it was planned that phase 2 will start operating

earliest in 2020. (Länsimetro 2014) Based on the updated project plan, accepted by the city council of Espoo, phase 2 will start operating in 2023 (Länsimetro 2020c).

Figure 1 Route of Westmetro's phases 1 and 2 (Länsimetro 2020d)

In figure 1 route of Westmetro's phases, 1 and 2 are presented. In the yellow lined area are the phase 1's stations from Lauttasaari, Helsinki towards the final station of phase 1 in Matinkylä, Espoo. In the phase 2 area, there are 5 new stations in Espoo that are presented in the red-lined area.

1.2 The objective of the study and research questions

The objective of this study is to predict the housing price development in Westmetro's phase 2 areas Soukka, Espoonlahti, and Kivenlahti in 2020-2023. Because the study is completed when part of the prediction period is already realized, predictions for Q1-Q3/2020 are created only to sold real estate in the observation areas, so that it is possible to compare the predicted and realized housing prices and to get the accuracy of the predictions for that time frame. Commissioner company of this thesis is Kiinteistömaailma Espoonlahti, Espoon Kodit Oy.

The main research question is:

What model should be used to predict housing prices for Westmetro's phase 2 areas?

There are previous studies about the effect of traffic systems on housing prices, including Westmetro's phase 1, but there are no previous academic studies that would have predicted the housing price development in the area of a new traffic system. There is variation in the results of previous studies, but in Finland, the studies have shown significant positive effects in the proximity of the station areas on housing prices. The previous studies have shown that housing prices had a significant increase in Westmetro's phase 1 areas during the construction time and after the operating started. The results of this study will give detailed information about the expected housing price development in selected phase 2 areas for 2020-2023 including the time during the construction and the estimated start of the operating. The results are beneficial for the employer of the thesis, but also for other operators in the real estate business as well as individuals, who are interested to buy or sell real estate. The results can be used in the evaluation of the value of the real estate and to support decision-making in investments, as well as in municipal zoning decisions.

Sub-questions that support the main research question in the process to create an accurate prediction model:
-*What variables should be chosen to predict the housing prices?*
-*Does the used interest rate in the dataset have an effect on prediction accuracy based on historical data?*
-*Does the COVID-19 pandemic affect housing price predictions?*

Variables for the prediction model are chosen based on literature and previous research. Bank of Finland's predictions for 2020-2023 is used for chosen macroeconomic variables in the housing price predictions. They include predictions for Euribor 3 months, however, Euribor 12 months is the more commonly used interest rate in mortgages and for this reason, it is observed whether it affects prediction accuracy, when Euribor 3 months is used in the prediction model instead of Euribor 12 months. Because the study is conducted during the COVID-19 pandemic, also the possible effects of the pandemic on housing price predictions are analyzed.

1.3 Research method and data

This study is conducted as quantitative research and consists of a literature review and empirical part of the study. For the empirical part of the study, the main source of research data is Hintaseurantapalvelu (HSP), which is upheld by the Central Federation of Finnish Real Estate Agencies. Permission to collect and use the data from HSP in this thesis is cleared by the Federation of Finnish Real Estate Agencies. The data in HSP is based on realized housing sale prices delivered by real estate agents and contractors. In addition to data collected from HSP, data for macroeconomic variables is collected from Statistics Finland and the Bank of Finland. Data is collected from years 2009-2019 from three phase 2 areas: Soukka, Espoonlahti, and Kivenlahti. By using the collected dataset, linear and quadratic ordinary least squares regression models are conducted by using a different kind of combinations of the dataset and evaluated to find the best fit for the model. As a result, a quadratic function of ordinary least squares regression is used as a method in this research for predicting the housing prices for the years 2020-2023. For the predictions, the Bank of Finland's prediction values for chosen macroeconomic variables are used.

However, as the aim of this study is to predict the housing price development also for the time Westmetro's phase 2 starts operating and the estimated time for Westmetro's phase 2 to start operating is in 2023, the effect of the start of the operating cannot find out based on data from 2009-2019 from Soukka, Espoonlahti, and Kivenlahti, which is used for conducting the regression models for phase 2 areas. As mentioned above, Westmetro's phase 1 started operating in 2017. For this reason, a separate regression model for phase 1 areas is conducted, to get the effect of the metro's start to operate. The data from Westmetro's phase 1 areas is collected from the years 2007-2019. 2007 was used as a starting year for phase 1 data collection to get the data from a time before the official decision about the Westmetro's phase 1 as well, as the official decision about phase 1 was made in May 2008. By conducting quadratic ordinary least squares regression model from phase 1 data, it is possible to get the effect of the metro's start of operating. The coefficient of the metro's start of operating is added into the equation of the chosen prediction model for phase 2 areas.

1.4 Structure of the thesis

The thesis consists of 5 main chapters in following order:

1. Introduction
2. Background
3. Data and Methodology
4. Results
5. Discussion and conclusions

In the introduction, the background of the study as well as the research problem, objective, data, methodology, and structure of the thesis is presented. In chapter 2, the theoretical background of housing price formation, the effect of macroeconomic factors along with transport infrastructure on housing prices, and housing markets in Finland are presented. The chapter continues with the state-of-the-art literature review including previous studies from Finland and abroad including studies analyzing the effects of the transport system on housing prices and studies that predict the development of housing prices. The chapter ends with a section about the methodological background. The empirical part of the study starts from the third chapter, where collected data is presented to get a better insight into how the prices developed in the observation areas in 2009-2019. In the same chapter, the actual research method and comparison of different regression models are presented. At the end of the chapter, COVID-19 and the changes it might bring into housing price predictions are discussed. In the fourth chapter the actual prediction results are presented and finally, in the fifth chapter are the conclusions and evaluation of the study as well as recommendations for further research.

2 BACKGROUND

The theoretical background of the research presents the formation of housing prices, the factors affecting the price, and the development of the housing market in Finland. International and Finnish academic publications about the effect of investments in transport infrastructure on housing prices are covered in chapter 2.5, along with the studies focusing on the prediction of the housing prices. At the end of the chapter methodological background of the Hedonic price model and the Ordinary least square method are presented.

2.1 Housing price formation

An apartment could be considered as essential for a person as everyone needs to live somewhere. The apartment itself cannot be moved and it is not an easily changeable commodity. As a commodity, an apartment is relatively expensive. In Finland, medium-sized apartment's market price is approximately four times the yearly net income of average households. (Laakso & Loikkanen 2004, 251) Approximately 64 % of households in Finland are homeowners (Putkuri 2018).

There are different kinds of apartments based on their size, type, quality, and structural features. Housing price does not format based on only these features; it also considers the location of the apartment. As there are different features in apartments, also the consumers, apartment buyers differ from each other as they have different ages, family types, income, and phase of life. When the buyer chooses the apartment, it also means choosing the environment, access to public transportation, services, and many other things that are depending on the location. These features are strongly affecting on the choice of the apartment and the housing price. (Laakso & Loikkanen 2004, 241). The housing price of apartments with otherwise similar features might vary a lot according to location. Housing price can be considered to format from two parts: the value of its' physical features and land value. (Lönnqvist 2015, 28). When considering the physical features of the apartment, some renovations might have significant effects on prices per square meter. For example, in Helsinki, the difference between real estate from the apartment building, where the plumping renovation has been done compared to the apartment where it has not been done is 850 euros per square meter. (Yle 2013) Plumping renovations should be done every 40-60 years (Re/max 2018). According to Talouselämä (2015), having sauna in the apartment might

increase the housing price in total for 15000 euros and apartment in the upper floor might increase the housing price in total 7000-20000 euros compared to similar in lower floor. However, it is also mentioned that having an elevator in the building cannot be determined to have increasing effect on housing prices. The value and location of a property are strongly connected, and transportation infrastructure affects both of these property qualities. Accessibility can be considered as a main aspect of the location. Determination of physical accessibility can be given by the time and cost of travel to other locations. (Henneberry 1998) For different groups of people, accessibility can be determined differently. It can refer to accessibility to the city center, transportation crossroads, or nature for example (Laakso & Loikkanen 2004, 145).

Debrezion et al. (2007) sum up a basic theory of real estate pricing in their research: when a location becomes more attractive, based on some qualities, it increases demand which increases the prices. Supply of many activities is often focused on city centers and that way closeness of city center is considered as a positive quality of real estate that increases the price. With the investments for transport infrastructure, demand for the closeness of city center decreases as that increases attractive qualities on real estate around the stations. (Fejarang, 1994) Also service concentrations, transportation crossroads, environmental features such as parks or seaside might affect that housing prices are not decreasing as the distance increases as they might create local price changes depending on how attractive those features are seen (Laakso & Loikkanen 2004, 145). For example, a seaside view from the apartment might increase the price per square meter by 2000 euros compared to a similar apartment without the view (Talouselämä 2015). Value for land that is used for housing constructions is mainly based on transport infrastructure. If there is no transport accessibility to the area, the land does not have significant value. (Laakso & Loikkanen 2004, 363)

With the investments in new transport infrastructure, relative accessibility of locations changes, creating localized and general changes in land values (Henneberry, 1998). As a result of the investment, real estate close to the station area benefit from transportation time and cost-saving. According to Agostini and Palmucci (2008), real estate closer to public transportation have a greater market value than the ones with otherwise similar qualities but poorer access to transportation. With the possible lower traveling cost to workplaces and

shopping areas, investments for transport infrastructure are capitalized into housing and land prices partially or totally.

2.2 Macroeconomic factors

The housing market interplays with several areas of the economy and those ways have a significant impact on the economic cycle because the expectations on the housing market affect the companies' investment decisions and households' consumption decisions. Ownership of real estate is the favored form of accommodation in Finland, and it increases the position of the housing market in the economy. Hence, the changes in the housing market are strongly connected to the changes in the level of households' wealth. Changes in housing prices have an effect on a household's wealth as well as on market rents and hereby affect purchasing power and level of consumption. In economical literature, the correlation of a household's wealth and level of consumption has been recognized for a long time. Macroprudential policy and the state of the housing market are strongly correlating and affecting the households' decisions about housing. That way the housing market affects the economic cycle and the stability of the financial system. As the changes in the housing market can be seen broadly in the economy, similarly the changes in the economy and financial markets can be seen in the housing market. (Lindblad et al. 2019)

There are macroeconomic factors that affect housing prices such as inflation, income, interest rates, stock markets, and unemployment rate (Abelson et al. 2005, 1). To buy an apartment, one often needs debt financing, which connects real estate markets to capital markets and macroeconomic developments (Lönnqvist 2015, 27). 97 % of the mortgages are tied to Euribor, which is following the interest policy of the European Central Bank (Brotherus 2019). In Finland, most of the mortgages are tied to Euribor 12 months. (Nordea 2020, OP 2020a). Development in the real estate market can also affect macroeconomic development in different ways, for example through the household wealth effect created by the increase in the real estate value. On the other hand, the result of rigidity of real estate supply when there is growing demand leads to an increase in housing prices and constrains other consumption as the money spent on real estate is reduced from other consumption. (Lönnqvist 2015, 27) Other factors affecting housing prices are demand, supply, and housebuilding, real interest rate as well as a financial crisis (Laakso & Loikkanen 2004, 275-277).

When households' demand for real estate purchases intersects with the supply of real estate available, the equilibrium is achieved in the housing market. Demand and supply affect the housing market in tandem, but their relative and absolute influence can be unequal. There are several factors affecting supply and demand. Many of these factors can be affected by policy actions. Some of them can be observed more easily than others and some factors can affect similarly both demand and supply, for example, interest rates or availability of finance. (Lindblad et al. 2019)

Municipal zoning decisions are strongly affecting real estate investment, for example, if a municipality decides to expand residential areas and there is land available for that, it is the way to support housebuilding. Also changes in building regulations or land taxation, availability of labor, changes in productivity, fiscal policy, and increasing competition in the construction sector might affect housing supply and new-build construction. (Lindblad et al. 2019)

Demographic change and age structure can be identified as factors affecting housing demand and as a factor that affects preferred housing type. In areas that have aging population, the demand for small houses and apartments might increase instead of larger ones further away from services. Instead, growth in the working-age population (20-64 years) in the area keeps up the demand. (Laakso & Loikkanen 2004, 275; Lindblad et al. 2019) Housing demand is affected by monetary policy, mortgage reference rates, and other terms attached to mortgages as they change the cost of borrowing. Furthermore, taxation can affect the housing market, for example, change in form of mortgage deductions or transfer taxes. (Lindblad et al. 2019)

2.3 Effect of transport infrastructure on housing prices

Traffic-related accessibility is mostly based on the transport infrastructure and the transport system that is built around it, which enables traveling, transporting, and producing different kinds of trafficking systems. Transport infrastructure is usually built for public use, investments for that are large and expensive while they have wide economic, environmental, and social effects. (Laakso et al. 2016, 427)

Investments in transport infrastructure are affecting land value and through that on housing and commercial premise prices. When accessibility improves, housing prices are expected to increase in the areas that benefit from the investment with better access to business and service concentrations. Improvement in accessibility capitalizes on land value in all areas but it is relatively larger in areas that have moderate travel distance. Change in land value affects the level of house building as well as the supply of apartments and commercial premises. (Laakso et al. 2016, 431) When considering transport infrastructure construction projects, the lifetime of a project is relatively long as planning, design, and construction normally take time. It is probable that the effects of the project can be seen before the project is completed. As there is competition in real estate markets, the buyers will consider the real estate prices based on the information available, including the expected improvements in accessibility in the future. (Banister & Thurstain-Goodwin 2011; Yiu & Wong 2005)

If the change in traffic system does not create weaker accessibility in other locations, the improved accessibility can increase the total value of the whole area's real estate. Due to changes in the traffic system, relations in housing prices change in a way that probable demand in new land use is more focused on the areas near to new traffic system and its benefit area. (Laakso et al. 2016, 442)

The higher the market price for land, the more effective it is aimed to use. Respectively, the better the accessibility the higher is the land market price, so effective land use is based on accessibility. The realization of effective land use depends on zoning. For example, in downtown of Helsinki, zoning limits the land use effectiveness to achieve the demanded level. On the other hand, in other areas zoning would allow more effective land use but there might not be enough demand for that. Investments in transport infrastructure can have a significant effect on housing and commercial premises demand. To realize the change in demand as more effective land use in better accessibility locations depends a lot on the community's zoning decisions. (Laakso et al. 2016, 440)

If there is an improvement in the traffic system in the area, households benefit from the faster transport connection to central and sub-central areas. Households are willing to pay more about the location with the improved traffic system. Respectively companies are willing to pay more about the location with the improved accessibility. As a result, the price for land increases in the area that benefits from the improved traffic system and attraction of the area

for companies and households grows resulting increase in workplaces and population. Improvement in accessibility through the change in transport infrastructure and market price for real estate are strongly correlated. (Laakso et tal. 2016, 440; Mulley et al. 2016) Change in accessibility is an economic advantage that households and companies are willing to pay. As the traffic system does not fully charge for better accessibility for example, through a high increase in ticket prices, the benefit capitalizes on the value of the real estate, which can be seen as changes in land, housing, and commercial premises market prices. (Laakso et al. 2016, 441)

According to Vallinkoski in Yle's article, improved accessibility usually increases interest in the real estate market in the area, but if the area already had working transport infrastructure the effects might be relatively small and can be seen as a shorter sale time for properties. Real estate should be located within 5 to 10 minutes walking distance from the transport's operation area to get the benefit from improved accessibility in a form of a higher increase in housing price. (Yle 2017) In the case of Westmetro's phase 2, if price behavior follows phase 1, it can be expected that the increase of housing prices starts to accelerate right before the metro starts operating and in the following years (Kodit.io 2018).

2.4 Housing markets in Finland

In Finland, there are approximately 110 000 real estate sold in a year and one-third of them are new-build real estate. New-build real estate have a significant impact on the housing market as often the person moving into new-build real estate sells the old one and creates a chain for the housing market. (Brotherus 2019) At some level, housing markets are always local and often limited by job market areas. Real estate markets are often considered as an individual but, there are several submarkets connected to each other. The main factor that divides the real estate market is the form of holding. To have an owner-occupied apartment, decent payment ability is required for loan payments and often increases savings. In the urban regions with high housing prices, housing demand might be directed towards rental apartments. (Lönnqvist 2015, 30) Housing prices have a significant influence on the growth of regional economies. The growth of regional price differences weakens the ability of labor movement from recessive areas to growing urban areas. Increase in income level and population increase the level of regional housing prices and rental fees. (Oikarinen 2011, 143)

2.4.1 Regional differences in housing prices

In Finland, the housing market has been historically developing very similarly in all regions. As urbanization has progressed during the 2010 decade, the housing price development between regions has started to diverge from previous and in the future, the housing price increase will focus geographically in a small area. (PTT 2020) Regional differences in housing prices have affected the size of mortgages held by households. Mortgages are large and have grown in euros as well as in relation to income level especially in urban areas, as the housing prices are more expensive, and the regional demand is higher. (Putkuri 2018)

It is expected that during the 2020 decade the increase in housing prices will be focused on the central cities of growing regions. The change in price development in regions with different populations has developed in the early 2010s. Earlier banks did not have risks when they gave mortgages as in the long run housing prices increased in all regions in Finland. Savings which were invested into real estate by the homeowners were safe and an increase in housing prices has increased households' wealth but due to regional differences, this might not be the case in the future. If the value of the real estate decreases in relation to mortgage the situation could be, that the value of the mortgage is bigger than the value of the real estate. This might affect negatively the labor movement, which has a negative effect on the labor market. A fast decrease in housing prices could be a consequence of a shock in the housing market such as a recession. (PTT 2020)

Urbanization is changing Finnish society in an unparalleled way: the regional structure of Finland is changing faster than it has in decades. In the metropolitan area and a few other university cities, populations are growing fast. In some cities' population growth is moderate but, in most regions, the population is declining, and this can be noticed in the development of the housing market. Despite the economic growth, the housing market is growing only in the growing regions, while in smaller cities prices have been decreasing. Urbanization is the most affecting factor of housing market development. In the growing areas demand in the housing market is growing and there are a lot of new-build constructions and housing prices increase. On the other hand, in the recessive areas, the issues in the housing market are getting worse. Urbanization in Finland is focusing only on few cities and from these mainly in metropolitan area and Tampere. (Keskinen et al. 2020)

The fastest price increase in 2019 was in the metropolitan area and Tampere. Population growth is focusing on these areas, which can be seen in the form of a larger increase in price development. In the housing prices, polarization can be seen within regions and cities. In the metropolitan area, increases in housing prices are maintained by the most expensive areas in Helsinki. Differences in price development between Espoo and Vantaa compared to Helsinki are getting more prominent. Especially in Vantaa, new-build construction has been keeping the price development at a moderate level. In Tampere, new-build constructions on the way of new tram line are popular and housing prices in the city center have been increasing 4 % a year since 2015. (Keskinen et al. 2020)

An example of regionally diverged price development is that the price of studio apartments in rest of the Finland increased less than 2 % in 2015-2019, while at the same time increase in Helsinki was 18 %. Also, the price of a one-bedroom apartment or a bigger one increased 10 % more in Helsinki than in other areas in 2015-2019. During the last ten years, the increase in housing prices has been fastest in Helsinki, where the prices have been increasing faster than the income level. (Keskinen et al. 2020) In Hypo's housing market review Q3/2019 the regional housing price development in Finland is also pointed out. Based on statistics, the housing prices in Helsinki have increased more than 17 % during the last six years while in the Kainuu region the decrease has been 19 %. It is also noted that the statistics give more optimistic results than the reality is. In reality, the differences between the regions are even larger as the housing prices are calculated from the sold real estate. In the recessive areas, only part of the real estate is sold, and some will be on sale for a long time or uninhabited, which are not included in the statistics. Instead in the bigger cities, even real estate in a need of big renovation are sold. But at the same time, as the rental land becomes more common in urban areas, it also skews the statistics. The value of a land share is approximately 20 % of the value of an apartment and for that reason, the price per square meter is lower for real estate in rental land. The real increase in housing prices in urban areas is even higher as in the statistics as they do not separate the owning type of the land. (Brotherus 2019)

2.4.2 Differences in housing prices related to housing type and size

Prices of studio apartments have separated from other apartment or housing types, especially in Helsinki but also in other areas in Finland. Price differences between different sized apartments were recognized early in 2012. Since 2015, with one-bedroom apartments and bigger ones, the price development has been more consistent. (Keskinen et al. 2020; OP 2020b) In 2019, the difference between studio and one-bedroom apartments grew significantly as the prices of studio apartments increased two times faster than the prices of one-bedroom apartments. Also, two-bedroom and larger apartments had a bigger increase in price development than a one-bedroom apartment in Helsinki. In other areas in Finland, the difference between the price development of studio apartments and other apartments got more consistent. When considering the housing price development from all the other areas from Finland than Helsinki, overall, only the prices of studio apartments were increased from 2015. The popularity of studio apartments has been increasing during the past years. In other bigger cities in Finland except in Helsinki, the increased supply of studio apartments has prevented the huge increase in studio apartment prices compared to larger apartments. Despite the grown supply of studio apartments through new-build construction, it has not been effective enough to prevent huge differences in price development between apartment types. (Keskinen et al. 2020)

There are differences in price development based on the number of bedrooms. In growing regions, the price development of studio apartments is different from others. Investing in apartments is still a popular investment. The growing development of investing in apartments is probably decreasing when the interest rates start to increase, or the growth of rental revenues starts to decrease. Investors' interest is mainly focused on small apartments located in central locations, which have been offering steady income during the time of low interest rates. A big part of small apartments sold in growth regions ends up being owned by an investor or investment fund. When the investors are buying the apartments, it does not create a chain for the housing market as they probably are not selling the apartment when they buy one. Especially the price increase of studio apartments combined with tightened up loan terms has increased the demand for rental studio apartments. (Keskinen et al. 2020)

The price development of detached houses has been varying during the years. At the country level, prices of detached houses have been decreasing in the 2010 decade, apart from the years 2015 and 2016. (Keskinen et al. 2020; OP 2020b) Price development of detached houses has been more consistent in small cities than in the metropolitan area, where price differences, as well as the number of houses sold, are varying a lot even in every quarter of a year. It can be explained through a small number of detached houses sold, so even a single sale might affect price variation in the metropolitan area or in some other bigger cities. Also, the statistics do not consider detached houses in rental land, which play a significant role in detached house sales in bigger cities. (Keskinen et al. 2020)

2.5 State-of-the-art literature review

There are several studies on how transportation infrastructure investments affect housing and land prices. There are differences in the focus of the research, methodology, and variation in results, whether the effect on housing and land prices is positive or negative, also whether the effects capitalize on prices before or after the transportation starts operating. There are studies from all over the world and Finland-based studies are mainly focused on the Metropolitan area.

In chapter 2.5.1 studies on different kinds of transport projects affecting real estate prices from all over the world are presented. In chapter 2.5.2 the focus is on studies from Finland and towards the end of the chapter, the focus is on other studies about Westmetro's effect on housing prices. As the aim of this study is to predict housing prices, chapter 2.5.3 concentrates on previous research about housing price prediction.

2.5.1 Previous research abroad

There are variations in the results whether changes in transport infrastructure have a positive effect on housing prices or not. According to Bae et al. (2003), Brandt & Maennig (2011), McDonald & Osuji (1995), and Pan et al. (2014) new and existing transit lines have significantly increasing effect on housing prices. The research of Zhang et al. (2014) revealed a positive effect on housing prices when considering light rail transit or metro rail transit but no effect for bus rapid transit. Some of the reviewed studies had partially differing

results. Mohammad et al. (2015) concluded that the metro had a negative effect on housing prices when the distance to the station is less than 500 meters. On the other hand, their research also showed a significant positive effect on housing prices but only when the distance is more than 500 meters and within 1 kilometer from the station. According to Brandt & Maennig (2011), if the distance to the station is less than 250 meters, housing prices in the vicinity of underground stations were 4,6 % higher than in the vicinity of aboveground stations. If the distance is more than 250 meters, there is no significant difference between that station types. In the research of McDonald & Osuji (1995), the data is collected from the time there was no official decision made about the transit line and from the year the line was under construction. It was found that the housing prices start to increase already during the time of construction, while the research of Pan et al. (2014) focused on the time period after the new transit line started operating. However, Bae et al. (2003) gave partially divergent results as according to their research, the new metro line had positive effects on housing prices only during the construction time and after the line started operating the price effects were evaporated three years later.

Other observations were that the distance to the city center is insignificant, but density and employment rate are significant factors. (Bae et al. 2003; Pan et al. 2014) According to McDonald and Osuji, long distances to shopping centers have significant negative effects and Pan et al. showed that the values of properties increased by 6,5 % if there is a shopping site nearby.

The hedonic price model is used in all the observed international studies, but the used methods are varying. Zhang et al. (2014) have used Ordinary least squares (OLS) regression as a method as well as Pan et. al (2014), alongside the Multilevel regression model (MLR). Generalized least squares (GLS) regression is used in the research of Bae et al. (2003) and Difference in Difference (DID) along with the hedonic model in the research of Mohammad et al (2015). The use of semi-logarithmic form is favored despite the chosen method. (McDonald & Osuji 1995; Bae et al. 2003; Brandt & Maennig 2011; Pan et al. 2014; Zhang et al. 2014).

2.5.2 Previous research in Finland

In Finland big part of studies on the effect of rail traffic system on housing prices are focused on metropolitan area and are conducted as theses work. However, Seppo Laakso has done studies about the effect of the metro line on housing prices in Helsinki, when the metro first started operating in the 1980s, but he has also done following research about the topic. In the theses work during the 2010 decade the effect of different rail traffic systems on housing prices is studied including Westmetro in Helsinki and Espoo (Peltomäki 2017), Ring Rale Line (Laine 2017), and tram line in Tampere (Valaja 2018). The most relevant research concerning this thesis is the research of Harjunen (2018) for the city of Helsinki, where he has analyzed the anticipation effect of Westmetro on housing prices.

In Laakso's 1986 study the results showed that the metro had a significant effect on housing prices, especially in Eastern suburbs. Housing prices near metro stations have increased more than they would have without the metro. The main reason for this is the improved accessibility to the city center. The real estate further away from metro stations or in feeder transport areas has had a negative effect on housing prices since the metro started operating. Also, areas near to new shopping center and improved services benefitted increase in housing price. Laakso used the housing price data from the years 1980 and 1985, where 1980 represents housing prices before the metro started operating and 1985 represents housing prices 2,5-3 years after the metro started operating. (Laakso 1986, 30-31) Hedonic price model was used in his research as he did variance and regression analysis (Laakso 1986, 12). The biggest increase in housing prices was in real estate that were within a 400-meter radius from metro station (Laakso 1986, 21).

In 1997 Laakso continued his research about the effect of the metro on Helsinki housing prices. Since the research in 1986, the metro was expanded, and it had a new operating route further to the East of Helsinki and the line from Itäkeskus to Vuosaari was still under construction during the time of the research. (Laakso 1997, 230) The data in this research was collected from years 1980 to 1993. In the radius of 0-1000 meters from the metro station, the increase in housing prices was in total 3,8 %. The amount of increase varies a lot depending on the distance to the metro station, for example, if the distance is 0-250 meter, the increase in housing prices was 6,3 % while in area 750-1000 meter from the

metro station, the increase in housing prices was 1,3 %. In the feeder transport areas, the housing prices decreased approximately 5 % from 1980 to 1993. (Laakso 1997, 232-233)

According to Laine (2017), Peltomäki (2017), Harjunen (2018), and Valaja (2018) rail traffic systems have an increasing effect on housing prices. Positive effects on housing prices are indicated already in the construction time of the rail traffic system. (Peltomäki 2017; Harjunen 2018; Valaja 2018) Valaja's research is from construction time of the tram line. Results indicated a positive effect on housing prices in a 800-meter radius from the tram stop. However, it is stated in the results that the research could not prove that the increase in prices is due to the tram line. Peltomäki's (2017) results also indicated that the rail traffic system started affecting increasingly on housing prices during the construction time of Westmetro's phase 1. The highest increase during construction time is in the 400 to 800-meter radius from the metro station, while the highest increase in less than 400-meter radius from the metro stations is close to the originally estimated start of the operation date. Harjunen's research showed that an increase in housing prices can be seen near the new metro stations even five to six years before the metro starts operating. Housing prices increased approximately 4 % within 800 meters from the new metro station. If the distance to the nearest metro station was more than 800 meters, there was no anticipation effect in housing prices.

*Table 1 Latest Finnish studies about the effect of transport system on housing prices. * refers to thesis work.*

Researcher	*Year*	*Transport system*	*Method*	*Observations*	*Type of housing price data*	*Timeframe for data collection*
Laine*	2017	Train	OLS	18119	Sale price	2004-May 2017
Peltomäki*	2017	Metro	DID	11431	Sale price	2006-2017
Harjunen	2018	Metro	DID	43025	Sale price	2003-2016
Valaja*	2018	Tram	OLS	8460	Sale price	2015-May 2018

Hedonic price models were used in the observed studies and both OLS and DID methods are used in the latest Finnish studies. As presented in table 1, OLS was adopted into two of the observed studies (Laine 2017; Valaja 2018) as well as DID method (Peltomäki 2017; Harjunen 2018). For all of these four studies the housing price data is collected from HSP and purchase prices are used. Use of semi-logarithmic form (Laakso 1997; Peltomäki 2017; Harjunen 2018) and log-logarithmic form (Laine 2017; Valaja 2018) is common in the studies explaining housing prices.

As presented in table 1, Valaja has used a notably smaller number of observations and a shorter timeframe compared to Laine, Peltomäki, and Harjunen. Despite the smaller number of observations, like the others, Valaja has also collected the data from the time before the official construction decision of the tram line.

2.5.3 Previous research about housing price prediction

Various methods are used for predicting housing prices. Dubin (1998) adapted OLS regression and Maximum Likelihood (ML) methods for his predictions, and OLS was also adapted by Ottensmann, Payton, and Man (2008). Limsombunchai, Gan, and Lee (2004) have compared the performance of the Hedonic price model, Weighted Least Squares (WLS) regression, and Artificial neural network (ANN) model for housing price predictions.

The semi-log form is used in housing price predictions as well, to reduce heteroskedasticity. (Dubin 1998; Limsombunchai et al. 2004; Ottensmann et al. 2008) However, Dubin refers to Goldberg (1968) about a problem that even though logarithmic forms reduce heteroscedasticity, when predicting housing prices, the results will be biased when transformed back to functional form. On the other hand, he also states that to produce superior predictions, the theoretical superiority of the log form will overcome the bias. In his research, results are presented in linear functional form.

Table 2 Previous studies about predicting housing prices.

Researcher	*Year*	*Predicting method*	*Observations*	*Type of housing price data*	*Timeframe for data collection*
Dubin	1998	OLS and ML regression	1493	Sale price	1978
Limsombunchai, Gan and Lee	2004	WLS regression and ANN-model	200	Market price	May 2003
Ottensmann, Payton and Man	2008	OLS regression	8772	Sale price	1999

According to Dubin (1998), ML regression predicted housing prices better than OLS regression and Limsombunchai et al. (2004) showed that ANN-model has a better predictive power on housing prices over the WLS technique. However, Limsombunchai et al. recognized that the poor performance of the Hedonic price model could be caused by a lack of environmental variables and a small number of observations. Also, instead of purchase prices, they used market prices and assessed that economic factors, such as exchange rate or interest rate, that might affect housing prices are not included in the model. As presented in table 2, there is a lot of variation in the number of observations used in predictions. Also,

the used timeframe for data collection is relatively short, especially in the research of Limsombunchai et al. (2004), where it is only one month. In his research Dubin (1998) used a 66-33 ratio to split the data. Adjusted R-squared for the model is 0,731 and the sum of squared errors in functional form 81116,89. In their research, Ottensmann et al. (2008) have tested the performance of alternative measures by OLS, in their research the data was not split. The tests showed that location in relation to several employment centers is a significant predictor of housing prices and should be included in the model. Either distances, travel times, or measures of accessibility to the employment center should be included in variables of location in relation to employment. The tests also resulted that a model using several employment centers and accessibility performed better than the models using only distance to the center. The best results were from the model with a combination of accessibility to employment and change in accessibility.

2.5.4 Summary of results in previous research

The results are affected by the used data. All the observed publications presented in chapters 2.5.1-2.5.3 have a relatively big variation in the used number of observations, used timeframe for data collection and there is also some variation in the quality of the data especially between countries. Some of the studies abroad used market prices instead of purchase prices (Limsombunchai et al. 2004) or the average price of apartments with similar characteristics (Bae et al. 2003).

Based on the reviewed studies rail traffic systems have an increasing effect on housing prices in the vicinity of the stations and already during the construction time. (McDonald & Osuji 1995; Peltomäki 2017; Harjunen 2018; Valaja 2018) If the distance to the station is a maximum of 800 meters it seems to have a significant positive effect on housing prices. (Laakso 1987; McDonald & Osuji 1995; Peltomäki 2017; Harjunen 2018) There is also some divergent results considering the immediate vicinity of the station as Mohammad et al. (2015) discovered the increasing effects on housing prices starting from the 500-meter distance from the station and Brandt & Maennig (2011) concluded that if the distance to the station is less than 250 meters, the station type effects on housing prices as the prices are 4,6 % higher in the vicinity of underground stations than in the vicinity of aboveground stations. Unlike in the other publications, according to Bae et al. (2003), the new metro line had positive effects on housing prices only during the construction time and immediately

after the line started operating, but the price effects evaporated after that. The hedonic price model is used in all observed studies when evaluating the effect of the transport system on housing prices, but the used method varied between publications.

Various methods are used for predicting housing prices. As the number of observations, time frame, and variables also varied a lot between studies, there was no model that would have performed always better than the others. However, the use of hedonic price models seems to be common in predictions as well.

2.6 Methodological background

Based on previous studies presented in chapter 2.5, Hedonic price models are commonly used for analyzing the effects of transport infrastructure on land values and housing prices as well as for predicting housing prices. This chapter covers the background of the Hedonic price model, its problems with the analyzes of housing prices, and the advanced background of the ordinary least squares regression, which is the research method of this study.

2.6.1 Hedonic Price Model

In the literature, there are several methods that are used to evaluate the effect of transport infrastructure on land values and housing prices. One way is to simply compare the before and after prices from the commuting area of the new transport infrastructure. In the repeat sales method, the data collected for comparison is from the real estate that were sold more than once during the observed time. The method gives accurate results about the actual housing price increase, but it does not consider the change in characteristics of the property. (Garg 2016) However, there are several other factors affecting housing prices such as apartment and neighborhood characteristics besides the accessibility factor. An increase in housing price and land value cannot be valued without considering the other influencing factors. (Mulley & Tsai 2016) The most common way in housing market related studies is to use the hedonic price model, which considers real estate as a bundle of different attributes based on Lancaster (1966) and Rosen (1974) (Lönnqvist 2015, 62; Mulley & Tsai 2016). In the model of Rosen (1974), the relationship between the characteristics of the commodity and its market price can be non-linear. Also, all characteristics have an impact on housing

market price and the market price of a commodity is considered as a sum of prices of characteristics of the specific commodity.

The quality or characteristics of real estate varies and as a product, it consists of numeral qualitative and quantitative characteristics. Still, the product is sold as a unit in the real estate market with a single total price as the individual characteristics do not have distinct prices. (Laakso 1997, 25) The price of real estate is determined by different characteristics and qualities of its construction, environment, and location. At any point in time, demand and supply determine the paid price for the real estate in the specific market. A hedonic price model explains how the quality and quantity of a real estate's characteristics affect its price in the real estate market. (Banister 2007, 16)

Every product, in this case, real estate, has a market price, P, which is connected to a specific value of vector Z:

$$P = P(Z)$$

It connects the prices and the characteristics into each other:

$$P = P(Z_1, Z_2, \dots, Zj)$$

The equation for hedonic price function can be written as follows:

$$P = f(A, L, E)$$

Where the relationship between the housing price, P, and all of its attributes is estimated, such as various characteristics of the apartment (A), location (L), and environment (E).

The hedonic price model can be seen as a straightforward model as it only needs to have specific information for instance the housing price, the group of housing attributes, and appropriate specification of the functional relationships. No information about the housing buyers or sellers is needed. (Chin & Chau, 2003) With good quality data from the time frame before and after the transport investment, the hedonic price model can provide strong methods for the analysis and separate effects of different variables can be isolated. (Banister 2007, 17; Chin & Chau 2003)

The aim of forecasting is to create a prediction about the future values of the data and one way to achieve it is regression. Through regression analysis, it is possible to find out the correlation between the variables, but also analyze the quantitative data for estimating the parameters of a model to create the predictions for future values. (Prabhu, Chivukula, Mogadala, Ghosh & Livingston 2019, 200) Regression analysis can be conducted with available information on housing price data, where the real estate price is the dependent variable and other physical, economical, or quality characteristics are independent variables. The regression results provide information about how much a specific attribute would affect the housing price and if it is a positive or negative impact on the housing price. (Banister 2007, 17) The aim of a regression model is to explain the variation in exploratory (dependent) variable by the variations in explanatory (independent) variables (Mellin 2006, 267).

When predicting values, splitting the data into training and test sets is a crucial part of evaluating the models. When the data is separated into training and test sets, most of the observations are in the training set. 70-30 ratio is typically used in data splitting. When similar data is used on training and test sets, it is possible to minimize the discrepancy effects of the data and to understand better the qualities of the model. Once the model is improved through the training set, the model can be tested by making predictions against the test set. As the testing data already has values for the dependent variable, it can be determined if the model's predictions are correct. (Microsoft 2018) When determining the accuracy of the predictions, the predicted values are compared to actual values and the difference is aggregated in some way. The error of the prediction is the difference between the predicted and actual value and depending on whether the prediction is too high or low, the forecast error is positive or negative. If the values are summed, the positive and negative errors will cancel each other out and for that reason, it is better that the difference is squared or the absolute value taken, when all the values are positive, for example by calculating the Mean Squared Error (MSE) and Mean Absolute Error (MAE) values of the forecast errors. When evaluating the accuracy of the model, MSE or MAE values can be compared with those of other models for similar data and prediction timeframe and the model with the lowest error measuring values is the most accurate. (Brooks 2014, 292-293)

Regression analysis can be used for non-linear relationships as well. This requires transformations of the variables and transformation can be done for both dependent and independent variables. Logarithmic transformations can be done to reduce non-linearity. (KvantiMOTV 2003) The choice of the functional form is an empirical issue concerning the hedonic price model, as the hedonic theory does not tell the best functional form to use. There are many basic functional forms that can be used in the hedonic price model such as linear, semi-log, and log-log. It affects the results of the hedonic model which form of estimates are used. (Dunse & Jones 1998) In most studies about housing prices, the form of the hedonic price function is log-linear or semi-log, but Box-Cox transformation is common as well. (Laakso 1997, 49)

2.6.2 Problems with Hedonic Price Model

The use of the hedonic price model in housing price analysis is also criticized. In the model equilibrium is assumed in the housing market and that the price of variables is not interrelated. Because there are imperfections in the real-world housing market, the market equilibrium is not probable. The view that there would not be interrelationships between the implicit prices of attributes is incorrect as it would mean that the implicit price of attribute would not vary through areas and real estate types. Not all attributes give the same level of convenience or inconvenience. (Chin & Chau 2003; Dunse & Jones 1998)

The use of the hedonic price model in the real estate market has many key assumptions. Homogeneity of real estate is assumed and it is an arguable assumption. Real estate can be seen as heterogenous as they differ from each other for example by locational or structural characteristics or by the housing type. Another assumption is that the real estate market works under perfect competition and there are several buyers and sellers. In the real estate market, there are several buyers and sellers, but an individual buyer or seller cannot substantially affect housing prices. Buyers and sellers are considered to have the freedom to enter and exit the market and no constraints are set for the demand and supply of housing or limitations for housing producing resources. Nevertheless, there can be limitations with the budget for buyers and new-build constructors. A reasonable assumption is that buyers and sellers have perfect information about the real estate product and price, but it can be still claimed that in reality, perfect knowledge is impossible to achieve. Buyers aim to get as much information as possible about the housing-related attributes before making the

purchase decision. For sellers, the perfect information enables possibly to increase the housing price. The most relevant information such as price and attributes are available from brokers and real estate agents, but perfect information might be never fully realized in reality. (Chin & Chau 2003) Despite the problems, the hedonic price model is the most common method applied to housing market analysis (Dubé et al. 2013; Dunse & Jones 1998).

2.6.3 Ordinary Least Squares

Ordinary least squares (OLS) regression is a generalized technique for linear modeling that can be applied to model a dependent variable which is presented in a continuous data scale. OLS can be applied to a single or several independent variables. Independent variables can be also categorical variables that are appropriately coded into dummy variables. (Hutcheson & Sofroniou 1999, 55; Hutcheson 2011)

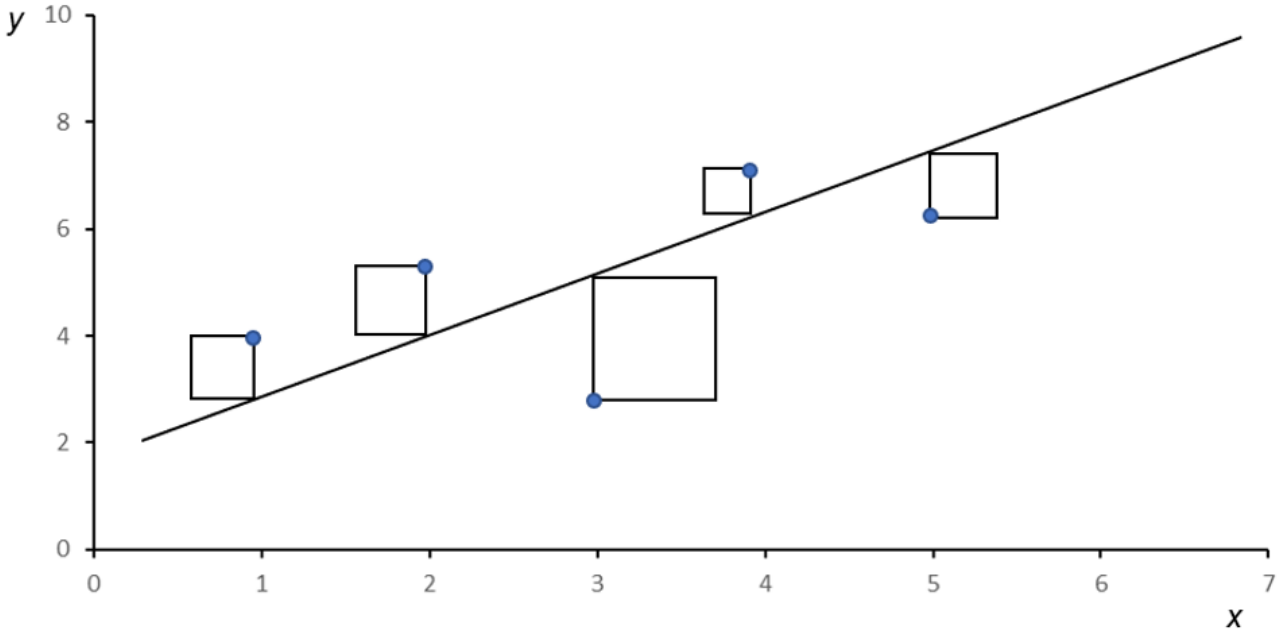

Figure 2 Method of OLS (Brooks 2014, 79)

OLS considers vertical distance from each point to the fitted line, squares it, and minimizes the sum of square areas, which is also known as the residual sum of squares (RSS). The method of OLS is illustrated in figure 2. The values that the model would have predicted are on the fitted line and the blue spots denote the actual given values of the observations. (Brooks 2014, 78-80) In OLS least squares procedure is used to minimize the sum of squared error terms. (Hutcheson & Sofroniou 1999, 57; Mellin 2006, 270) By using OLS

regression, it is also possible to model non-linear relationships as long as appropriate transformations are applied to one or more variables to transform the relationship linear. For modeling continuous data, especially when used in combination with dummy variables and transformed data, OLS regression is a powerful technique. (Hutcheson & Sofroniou 1999, 55-56)

Equation for linear regression can be written as follows:

$$y = \alpha + \beta x$$

Where the intercept of the line is presented by α and β presents the slope of the line. (Brooks 2014, 77; Hutcheson & Sofroniou 1999, 56)

OLS looks for the estimates $\hat{\alpha}$ and $\hat{\beta}$ of the parameters α and β such that $(\hat{\alpha}, \hat{\beta})$ = arg min $L(\alpha, \beta)$ where:

$$L(\alpha, \beta) = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} (y_t - \alpha - \beta x_t)^2$$

The summation is taken over all the observations from $t = 1$ to T, where T is the number of observations and with respect of $\hat{\alpha}$ and $\hat{\beta}$, L is minimized to find the values of α and β, so that the line is fitted to the data (Brooks 2014, 80-81).

Under the following assumptions, OLS will have best linear unbiased estimators (BLUE):

1) Specification: The relationship between dependent and independent variable has to be linear in parameters.
2) Unbiased: The expected value of the residual is 0.
3) Homoskedasticity: The constant variance of the errors.
4) No autocorrelation: Residuals are independent from each other.
5) No multicollinearity: No very high correlation between explanatory variables.
6) Additional assumption of normality: Dependent variable and residuals are normally distributed.

(Brooks 2014, 179–219; Puumalainen 2019)

If these assumptions are violated parameter estimates are biased and their standard errors might be wrong. If the assumptions fail, the model can be improved by transforming variables for example into logarithmic or dummy form. (Puumalainen 2019)

OLS is a commonly used empirical research method in studies about housing prices that apply the hedonic price method (Lönnqvist 2015, 72). As mentioned above OLS regression is powerful for modeling continuous data, especially when used in combination with dummy variables and transformed data (Hutcheson & Sofroniou 1999, 56).

2.6.4 Choosing the variables

It is common that when conducting a hedonic price model for housing markets, the variables are classified into groups such as environmental, locational, socio-economic, and structural variables. The aim and extent of the research reflect the choice of the variables. The prevalent factor in several studies seems to be data sources instead of theoretical reasons. Whereas it is problematic to choose the individual variables. As the variables are chosen from the available data, there are problems connected to limitations of data sources and difficulty to get reliable data as well as problems related to the definition, measuring, and qualification of variables, for instance, the environment or socio-economic structure of the neighborhood. (Laakso 1997, 46)

Another problem is multicollinearity, which was already mentioned in previous chapter 2.6.3 as one of the assumptions for OLS. Correlation between explanatory variables exists in regression analysis but if the correlation is too strong it leads to multicollinearity. The multicollinearity problem might appear if the correlation between explanatory variables is over 0,9. However, it is not always possible to observe multicollinearity problems from the correlation between variables. One way to measure multicollinearity is the Variance Inflation Factor (VIF). (KvantiMOTV 2003) VIF value of 10 indicates that the variance of the specific regression coefficient is 10 times higher than it would have been if the specific variable did not have a strong dependency with the other independent variable in the model. Often 10 has been used as a limiting value in the literature that multicollinearity problems occur. However, the VIF values should be evaluated in the context of other factors that affect the specific variable. The effects can decrease the variance of the regression coefficients even

when VIF is even 40 or more. When VIF has high values, one option is to respecify the model by removing one or more of the independent variables that suffer from multicollinearity. However, sometimes respecifying might do more harm than improvements to the model. (O'Brien 2007, 683-685)

If several variables explaining the variation of housing prices are multicollinear with each other, it weakens the econometric analysis. If multicollinearity problems are ignored in the empirical analysis, estimates of parameters are inconsistent and it results in unreliable test statistics. On the other hand, the risk of misspecification of the model exists, if the model is reduced and simplified too much. (Laakso 1997, 46-47) In the misspecification of variables, the model might be over-specified as an irrelevant independent variable is included or under-specified if the relevant independent variable is excluded from the model. (Chin & Chau 2003) According to Puumalainen (2019), if the model is only used for predicting and the coefficients are not interpreted, multicollinearity will not cause large damage.

To measure how well the regression model fits the data, there are measures known as the goodness of fits statistics. R2 is the most common goodness of fit statistic. R2 expresses a square of the correlation between the values of the exploratory variable and corresponding predicted values from the model. As the correlation coefficient is expressed on a scale of -1 to 1, due to that R2 is expressed between 0 and 1. If the model fits the data well, the R2 value is close to 1 and if it is close to 0, the model is not fitting to the data. The problem with R2 is that it does not decrease if more variables are added into regression and the R2 value will be at least as high for the updated regression as it was for the previous one. This makes it difficult to use R2 as a determinant measure of whether the added variable should be included in the model or not. A modification to pass these problems can be done to R2, which considers the loss of degrees of freedom related to adding more variables. Adjusted R2 can be used to make a decision whether the added variable should be included in the model or not. If the value of adjusted R2 increases, the variable should be included, and if it decreases it should not be included. (Brooks 2014, 151-155)

Outliers are observations that are away from the most of the observations on one variable or a combination of variables (Hutcheson & Sofroniou 1999, 19). These single divergent observations might have an effect on the results of regression analysis (KvantiMOTV 2003). Outliers are points in the data that do not fit in the pattern and are away from the fitted model

(Brooks 2014, 690). There are several reasons for why outliers exist as for example, data might include misspellings or problems with missing values. For checking whether there are outliers in the data, it is possible to use graphical methods such as histograms, box plots or normal probability plots. (Hutcheson & Sofroniou 1999, 19-20) One way to improve the model is to remove the outliers. When outliers are removed, standard errors and residual sum of squares are reduced and therefore R2 increases, which means better fit of the model for the data. Outlier observations are away from the rest of the observations and do not fit in the pattern of the rest of the data. If there are outliers in the data, it might have significant effect on the coefficient estimates in OLS. OLS tries to minimize the distances between points. When there are points far away from the fitted line and the residual, the distance from the outlier point to the fitted line, is squared, it leads to increase in the RSS. Even though, each observation represents a valuable part of information, there is a benefit to remove outlier observations that could have an excessive effect on the OLS estimates. (Brooks 2014, 211-213)

3 DATA AND METHODOLOGY

In chapter 3.1, the dataset and some descriptive statistics are presented. The chapter also includes an analysis of the housing price development on the timeframe of 2009-2019 to get a better insight, how the prices have developed in different areas during the different stages of Westmetro's construction. Chapter 3.2 includes sub-chapters where the data is tested, whether the OLS estimator is BLUE, regression models for different datasets are conducted and performance of the conducted models are compared and finally the model with the highest prediction accuracy is chosen for the actual predictions. As this study aims to predict the housing price development for the time Westmetro’s phase 2 starts operating and the estimated time for Westmetro’s phase 2 to start operating is in 2023, the effect of start of the operating cannot be find out based on data from 2009-2019 from phase 2 areas. A separate regression model for phase 1 areas is conducted, to get the effect of the metro's start to operate and equation for phase 2 prediction is formed by combining the chosen regression model and the effect of the metro's start to operate from phase 1 regression model. Towards the end of the chapter, sub-chapters present the observations that are used to create the predictions for the years 2020-2023 as well as two options for alternative predictions.

3.1 Data

Data is collected from Hintaseurantapalvelu (HSP) which is upheld by the Central Federation of Finnish Real Estate Agencies. Data on the website is based on realized housing sale prices delivered by real estate agents and contractors. The data has been collected in HSP since 1999. The data is available for parties working in the real estate business. (KVKL 2018) When analyzing the housing market in the Metropolitan area, using the data based on realized housing prices gives a truthful image of the price development. As mentioned in chapter 2.4.1, according to Brotherus (2019), in bigger cities even real estate in need of renovation are sold, which describes the observed areas in this study, but in recessive areas, only part of the real estate supply is sold, so for recessive areas, the realized housing price data would not give a truthful image of the areal housing market.

Data is collected from selected Westmetro's phase 2 areas: Soukka, Espoonlahti and Kivenlahti. In addition to phase 2 areas, data is also collected from all the phase 1 station

areas. By collecting the data from phase 1 areas as well, it is possible to analyze the housing price development during the different stages of the construction project and finally to analyze and to get the effect from the time the Westmetro's phase 1 started to operate. The desired data collection area is determined to be a 1 km radius from the metro station, so a 1 km radius area for each station was determined. These 1km radius areas are presented below in figure 3. Areas' zip codes were used in HSP to collect the data but also in some cases, the more exact border areas had to be determined based on street addresses, for example in a case where only a small part of the specific zip code area is fitting into 1 km radius area.

Figure 3 Determined areas for data collection (Mapdevelopers 2020)

In figure 3, phase 2 areas Soukka, Espoonlahti, and Kivenlahti are presented on the map and a 1km radius from the metro station is presented by the circles. 1km radius is determined on Google maps by using Mapdevelopers' draw a circle tool. It can be seen from figure 3, that in Soukka and Kivenlahti, the seaside area is within a 1km radius. Seaside is not far from the fringes of the 1km radius in Espoonlahti either. As it was mentioned in chapter 2.1, environmental features affect housing prices and seaside view might create a large increase in price per square meter compared to the similar apartment without the view (Talouselämä 2015).

Besides the housing sale price, HSP provides more information about the sold real estate, such as house type, zip code, municipality, area, street address, square meters, year built, number of rooms, floor, the total number of floors, sale price, debt share, sale price free of debt, price per square meter, condition of the real estate, date of sale, site ownership, maintenance fee, sauna, balcony, selling time and building material. The data in HSP is entered manually, so there is a chance for misspelling or missing values as not all the information is mandatory to enter. The collected data was reviewed carefully, and some small defects were corrected from the data. Rows with missing data were deleted, usually, the missing information was from the year built, number of rooms, or site ownership variable. Some variables had missing values more often, for example, variable maintenance fee, sauna, or balcony had several missing values so the whole variable was dropped from the dataset instead. This decision might affect the reliability of the predictions, as it was mentioned in chapter 2.1, according to Talouselämä (2015), having a sauna in the apartment might increase the total value of the real estate by 15000 euros. However, due to a large number of missing values in those specific variables, the decision was made that it was more reasonable to remove those variables. If those variables would have been included in the data and then observations with missing values would have been removed, it would have meant a large decrease in the total number of observations. The original number of observations in the dataset is 3431 and the number of observations after data treatment is 3204.

Table 3 Presentation of variables and descriptive statistics

Dependent variable		***Data source***	***Variable type***
	Price per square meter	HSP	Continuous
Independent variables			
Apartment	Year built	HSP	Continuous
	Number of rooms	HSP	Categorical
	Floor	HSP	Continuous
	Square meters	HSP	Continuous
	Housetype	HSP, modified	Dummy
		Dummy variables: Apartment buildings and Houses	
		Refrerence variable: Terrace houses	
	Site ownership	HSP, modified	Dummy
		0=Rental, 1=Own	
Location	Area	Created	Dummy
		Dummy variables: Espoonlahti and Kivenlahti	
		Refrerence variable: Soukka	
	Distance to Helsinki city center	Created	Continuous
		Distance measured in meters in the accuracy of hundred meters	
Variables related to Westmetro	Stage of metro construction	Created	Dummy
		*Dummy variables: Official decision about the construction, Under construction and *Metro is operating*	
		**refers used only for predictions, not in 2009-2019 dataset*	
		Reference variable: No official decision about the construction	
	Date of sale	HSP, modified	Categorical
		2009/Q1-2019/Q4 in graphs, 1-44 in data, where 1=2009/Q1 and 44=2019/Q4	
	Distance to metro station	Created	Categorical
		100=0-100m, 200=101-200m, 300=201-300m, 400=301-400m, 500=401-500m, 600=501-600m, 700=601-700m, 800=701-800m, 900=801-900m and 1000=901-1000m from the nearest metro station	
Macroeconomic variables	Euribor 3 months	Bank of Finland	Continuous
	Inflation (CPI)	Statistics Finland	Continuous
	Employment rate 15-64	Statistics Finland	Continuous
Number of observations before data treatment		3431	
Number of observations after data treatment		3204	
Site ownership		*Share of rental site 0,31 %*	*Share of own site 99,69 %*
Area name		Area's share from the data	Number of observations
	Soukka	35,46 %	1136
	Espoonlahti	35,30 %	1131
	Kivenlahti	29,24 %	937
Housetype		Housetype's share from the data	Number of observations
	Terrace house	8,02 %	257
	Apartment building	87,89 %	2816
	House	4,09 %	131

In table 3 the analyzed variables are presented. The presented variables are chosen based on the literature and previous studies. Also, variables that had a lot of missing values were excluded. The presented variables are chosen to be analyzed, but it is possible that some of them are excluded from the actual prediction model that will be presented in chapter 3.2. The dependent variable is the price per square meter as it can be used to analyze the price development. The actual sale price is not suitable for the dependent variable as the aim is

to analyze and predict the housing price development of real estate of different types and sizes. There are 17 (18*) independent variables and they are divided into four categories: apartment, location, variables related to Westmetro, and macroeconomic variables. Some of them are collected directly from the Bank of Finland, HSP, Statistics Finland, and some are modified or created based on the information received from HSP.

In the location group, a variable area is created based on the address information received from HSP as well as the variable "Distance to metro station", which is presented further in this chapter. The dataset is divided into 3 areas, which are presented in the lower part of table 3. The defined area is based on real estate's address by measuring the distance to its closest metro station. Areas were transformed into dummy variables, as according to Laakso (1997), it is possible to reduce heteroscedasticity and multicollinearity problems by using area-level dummy variables. As the dummy variable has more than two values in the model, there should be one dummy less included in the model than there is created. The dummy variables in the model are compared into reference variable. In this dataset the reference area is Soukka.

In the group variables related to Westmetro, there are only created or modified variables. Dummy variables of stages of the metro project are created: dummy variable "official decision made" and dummy variable "under construction". Reference variable is the stage "no official decision". The value of the variable changes always on the next complete quarter as the change happened. For example, the official decision about Westmetro's phase 2 was made in June 2012, so the variable "official decision made" gets a value of 1 from Q3/2012. It was possible to get the date of sale from HSP and those are grouped into a quarter of the years. Distance measuring variables were created by using Google maps. The address information from HSP is used to get the distance to the nearest metro station and city center of Helsinki. Distance to the nearest metro station is classified every 100 meters and the groups are presented in table 3.

Some macroeconomic variables are added to the dataset. Quarterly average values for Euribor 3 months are collected from the Bank of Finland, while Consumer Price Index (CPI) (Tilastokeskus 2020a) and the employment rate for 15-64 years (Tilastokeskus 2020b) are collected from Statistics Finland. CPI is used to measure inflation, as it can be used as a general measure of inflation (Tilastokeskus 2020c). In CPI, the used index is 2005=100.

Research data from phase 2 areas includes 3204 observations after the data treatment. Rows with missing values were deleted as well as the observations that were more than 1000 meters away from the nearest metro station. Outliers of the data were checked by areas and were removed from the dataset. As mentioned above, there are 3 areas, which are presented in the lower part of table 3: Soukka, Espoonlahti, and Kivenlahti. Each of them represents one station area. The area is defined based on real estate's address by measuring the distance to its closest metro station.

When considering the whole dataset, the share of rental land is 0,31 % and own land is 99,69 %. As written in chapter 2.4.1, according to Brotherus (2019), rental land becomes more common in urban areas. Approximately 20 % of the value of an apartment is from the value of the land share and for that reason, the price per square meter is lower for real estate in rental land and a growing number of rental lands skews the statistics as they do not separate the owning type of the land. In this dataset, the share of rental land is relatively low. House type is divided into three categories: "Apartment building", including also small apartment buildings and 'Luhtitalo' -house type, "House", including also semi-detached houses and 'Erillistalo' -house type and "Terrace house". Apartment building is by far the most common option in the vicinity of the station as it has 87,89 % share of all observations in the dataset. House types are also transformed into dummy variables and the reference type is terrace house.

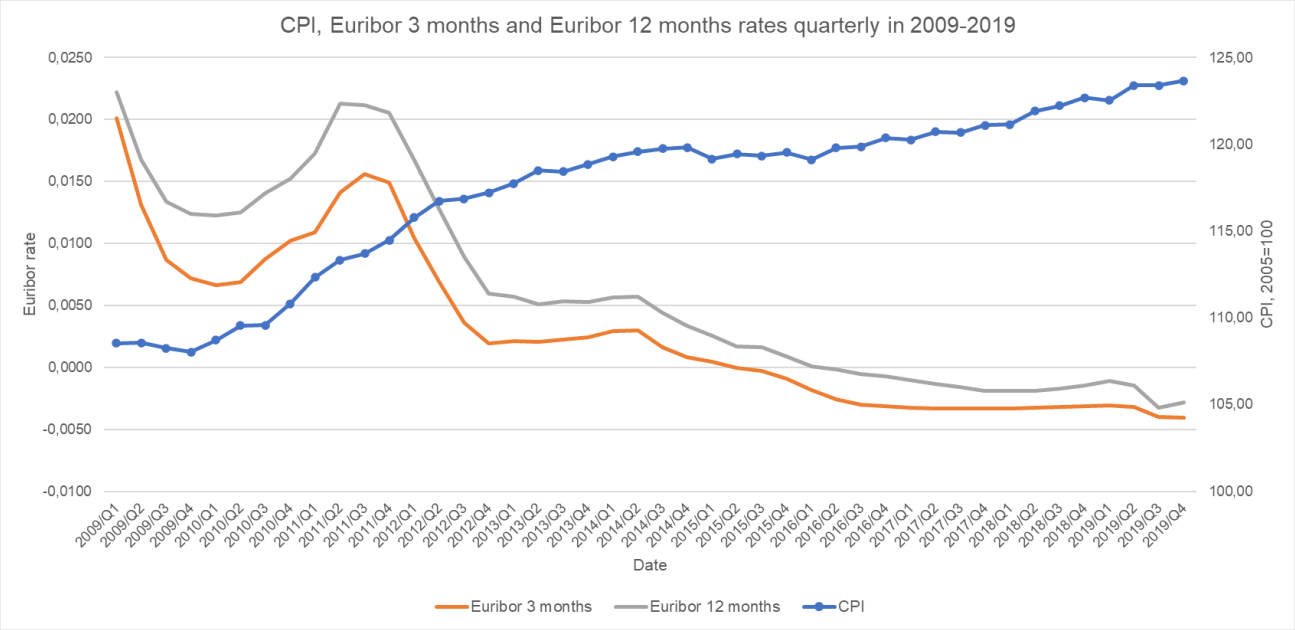

Figure 4 CPI (Tilastokeskus 2020a) and Euribor 3 and 12 months rates in 2009-2019 quarterly (Bank of Finland 2020a)

In figure 4, CPI and rates for Euribor 3 and 12 months in years 2009-2019 are presented quarterly. The level of Euribor decreased significantly during the financial crisis and it has remained low ever since. As the level of Euribor is low, the household benefits from it in a form of low-priced mortgages. As mentioned in chapter 2.2, most of the mortgages in Finland are tied to Euribor 12 months. However, the predictions of the Bank of Finland for 2020-2023 are for Euribor 3 months, and for this reason Euribor 3 months is chosen into the dataset. It can be seen from figure 4 that the rates for Euribor 12 months are a bit higher. However, the difference is not significant. CPI that is used in the dataset, the year 2005 has value 100 and the following years are compared to that. Bank of Finland's predictions of CPI, Employment rate, and Euribor 3 months for 2020-2023 is used in the data for the predictions. CPI should not be used to measure the development of housing prices and Statistics Finland has a separate housing price index (Tilastokeskus 2020d). However, there is no predicted values for the housing price index, and they would be needed to create the predictions in this study. If one wants to find out how much housing prices have changed compared to general price development, CPI can be used to calculate the actual price development (Tilastokeskus 2020e).

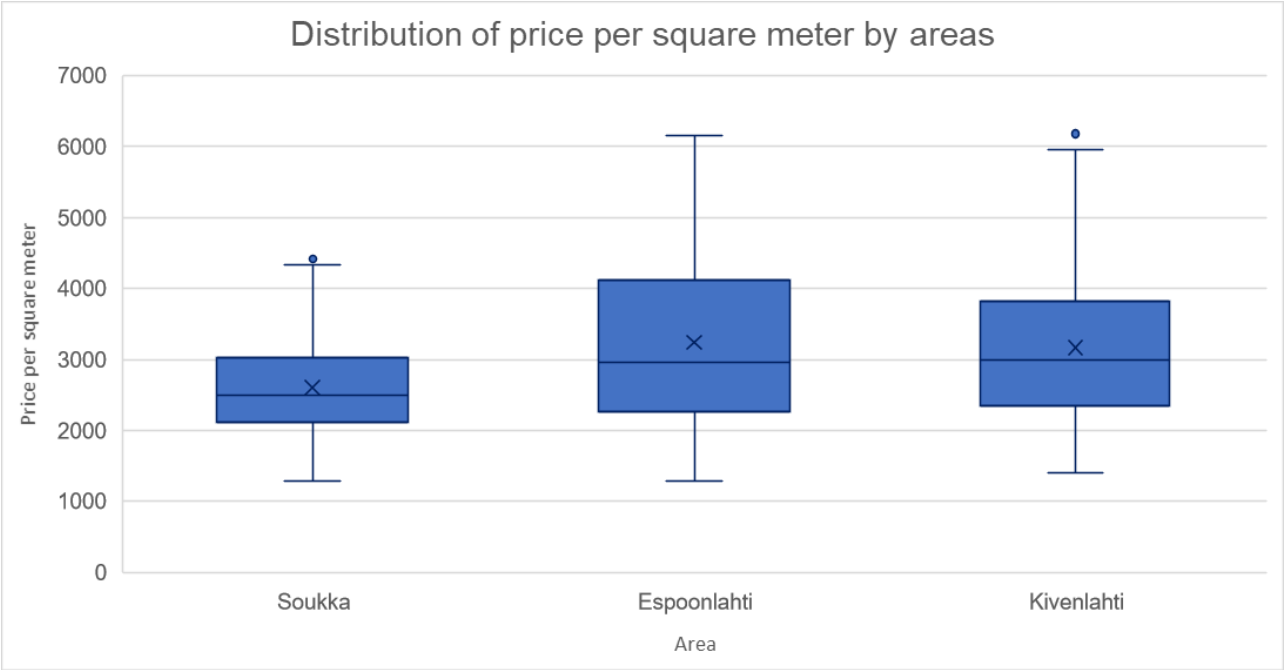

Figure 5 Distribution of price per square meter by areas in 2009-2019

Price per square meter varies a lot between areas. In figure 5 prices per square meter by area are presented in box plots. The mean price per square meter is presented by the x on the box and the median is presented by the line through the box. The top of the box presents the third quarter of the values. And the upper whisker presents the highest quarter of the values. The values are from the whole dataset from years 2009-2019 after removing outliers. There are still few values included, that can be considered as outliers and they are presented in figure 5 by the dot. However, the values are close to the highest quarter of the values, so they are decided to keep in the dataset. In Espoonlahti and Kivenlahti, the top part of the box and the upper whisker have a larger price section compared to Soukka, this means that there is a relatively large increase in price per square meter within these areas. Graphs about price development will be presented further in this chapter. Observations with outlier values were excluded from the dataset.

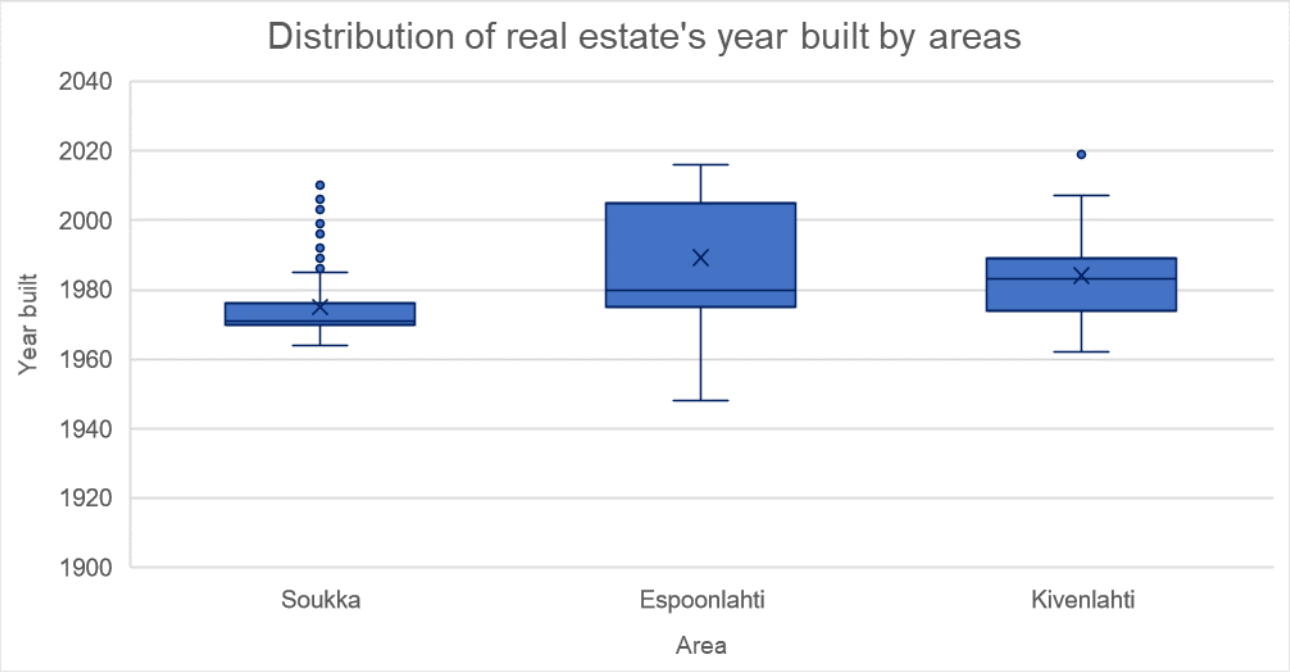

Figure 6 Distribution of real estate's year built by areas

In figure 6, the distribution of real estate's year built by areas is presented in the box plot. It can be seen from figure 6, that Espoonlahti has the widest and most even year built distribution on its real estate. In Soukka, most of the real estate are built from the 1960s to the mid-1980s. In Kivenlahti, the year built range is wider than in Soukka but not as large as in Espoonlahti.

In appendix 1 yearly nominal changes and real changes of average prices per square meter are presented by house types for every area: all house types combined, apartment buildings, terrace houses, and houses. Changes are calculated for the years 2010-2019, and the changes are calculated by comparing yearly average prices to the previous year's. Constructions for Westmetro's phase 1 started in Q4/2009 and from appendix 1, it can be seen that in 2010 and 2011 there are large increases in average yearly changes in most of the phase 1 areas. The official decision about phase 2 is done in Q2/2012 and in 2012 the nominal change in Espoonlahti is 23,85 %, so significantly larger than during the previous years in the area. However, a similar change in other phase 2 areas, Soukka and Kivenlahti cannot be seen. Phase 1 started operating in Q4/2017 and it can be seen that the price effects started to capitalize on housing prices before the operating started as for phase 1 areas Urheilupuisto and Niittykumpu the nominal change in 2017 is 32,94 % and 52,45 %

respectively. Also, in Kivenlahti, which is part of phase 2 areas, there is a significant increase as the nominal average yearly change for 2017 is 10,64 %, which is remarkably larger than the changes in the area during previous years. When considering only apartment buildings, the yearly changes are relatively similar to all house types changes, as majority of the observations are apartment buildings, 87,89 % as mentioned earlier in table 3.

When considering only houses, there are several gaps in phase 1 areas, so the yearly changes could not have been calculated in appendix 1. There are only three phase 1 areas that have yearly changes calculated for 2017 and 2018. In Urheilupuisto and Matinkylä, there is an increase in average prices per square meter in 2017, but the average prices decreased in 2018, as the yearly change for 2018 is negative in both areas. In Niittykumpu, there is no yearly change value for 2017, but the yearly change from 2017 to 2018 is significant, as the nominal yearly change is 38,39 %. As mentioned earlier, phase 1 started operating during Q4/2017. It can be seen that yearly changes of houses in Kivenlahti have large variations from 2013 to 2018 where the nominal yearly changes are varying between -39,87 % to 66,13 %. It also seems that trend of price development in Soukka and Espoonlahti has been decreasing as, throughout the observation period, there are only a few years in those areas when the average yearly change has been positive.

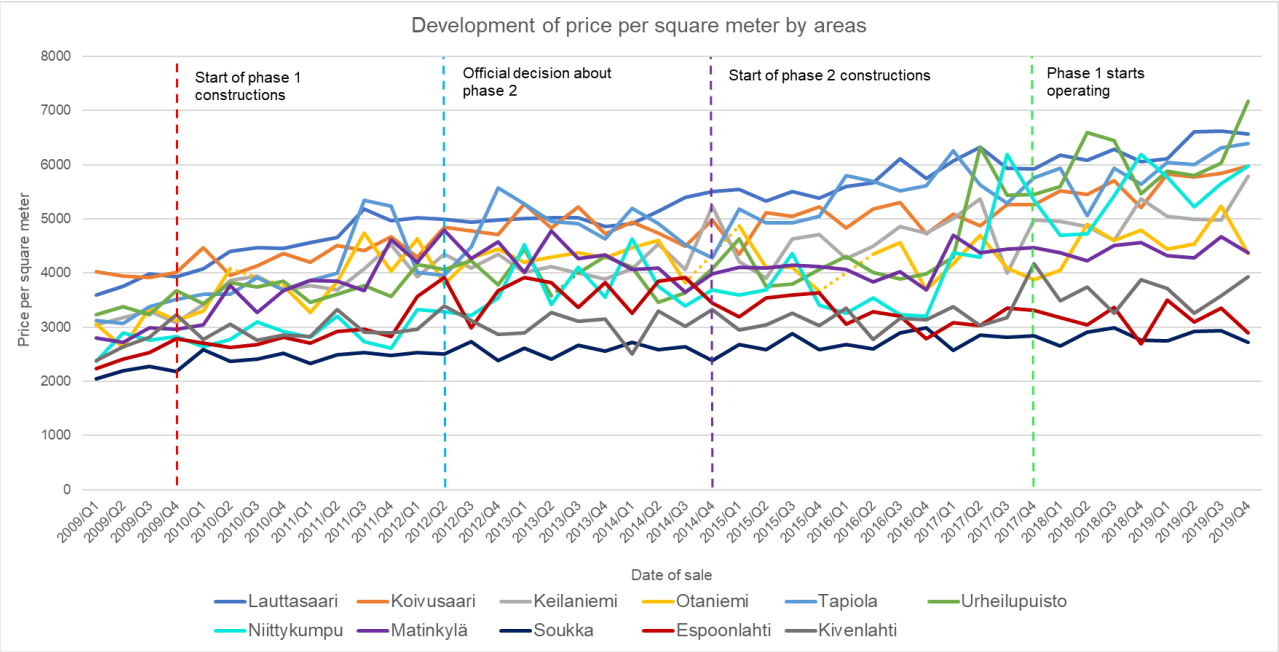

Figure 7 Development of average price per square meter by areas in 2009-2019

In figure 7, the development of housing prices of Westmetro's phase 1 and selected phase 2 areas in 2009-2019 are presented. When observing the phase 1 areas from figure 7, it can be seen that the price development in the areas is different from each other. Urheilupuisto and Niittykympu have significant increase during the observation period 2009-2019, but they also have a lot of high peaks and lows in quarterly presented average prices, for example, during the years 2017-2019 both areas have a significant increase in average prices right before the metro starts operating in Q4/2017 but around the time the operating starts, both areas have decrease in average prices before the prices take an upturn again. As an opposite to varying price development, average prices per square meter develop more consistently in Lauttasaari in 2009-2019 compared to Urheilupuisto and Niittykumpu.

The level of average price per square meter of phase 2 areas Soukka, Espoonlahti, and Kivenlahti is lower than in most of the phase 1 areas even at the beginning of the observation period in 2009. Based on figure 7, areas Niittykumpu and Matinkylä are the most similar with phase 2 areas based on average price per square meter level in 2009. At the beginning of the observation timeframe, areas Niittykumpu, Matinkylä, Soukka, Espoonlahti, and Kivenlahti have had a relatively similar average price per square meter. After the construction for phase 1 started, Matinkylä started to diverge from Niittykumpu and phase 2 areas and at the same time Niittykumpu, Espoonlahti, and Kivenlahti started to diverge from Soukka. During the following years 2010-2016, there is a lot of variation in average prices in Niittykumpu, Espoonlahti, and Kivenlahti, while there are no significant increases or decreases in Soukka during that time. It can be seen that there is a peak in average prices in Espoonlahti and Kivenlahti when the official decision about Westmetro's phase 2 is done in Q2/2012. However, there is no peak in the average prices of Soukka during that time. Approximately one year before phase 1 started operating, Niittykumpu and Matinkylä completely diverged from phase 2 areas.

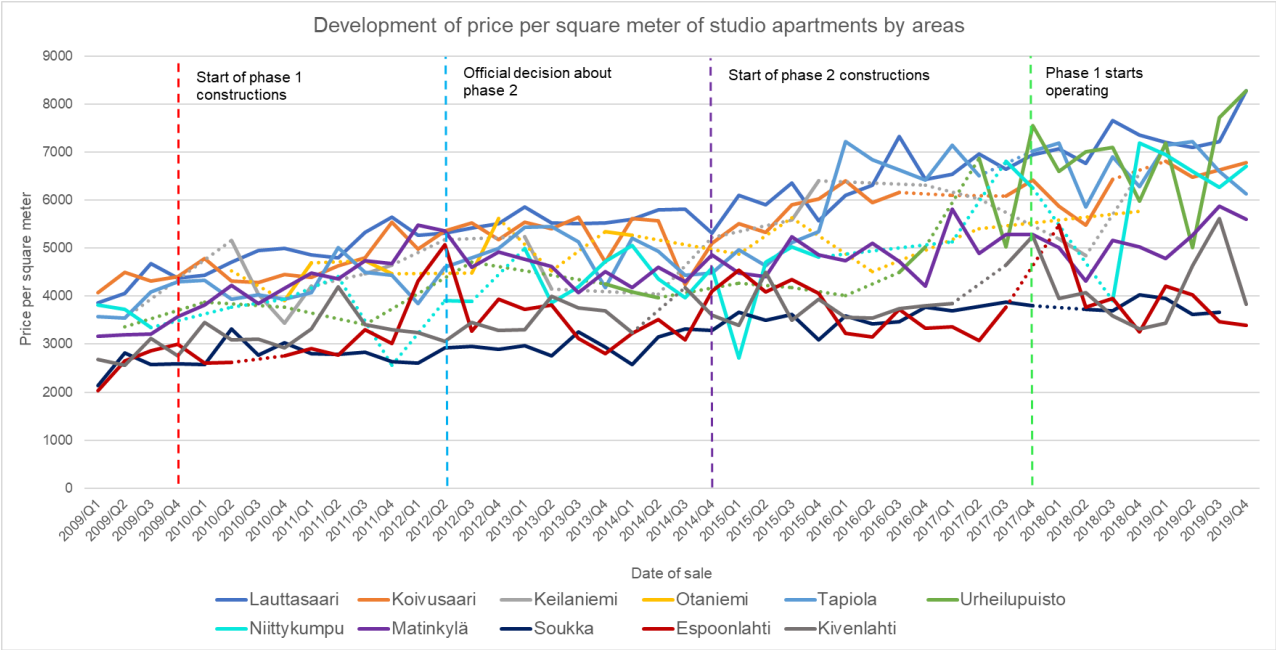

Figure 8 Development of average price per square meter of studio apartments by areas in 2009-2019

In figure 8, the development of average prices per square meter of studio apartments in Westmetro's phase 1 and selected phase 2 areas in 2009-2019 are presented. When considering only studio-sized apartments, there are several quarters in different areas, when there are no sold apartments. These time periods are presented by the dotted line in graphs in figure 8. If there was a single quarter with observation between the gaps, it is also presented by the dotted line as the single quarter with observation does not form a line. For example, in Niittykumpu in Q4/2011, it can be seen that there is a significant decrease presented by the dotted line. If there are no single quarters with observations between the longer breaks, the dotted line is linear, for example, in Niittykumpu from Q4/2017 to Q3/2018. The peak in average prices in Espoonlahti at the time when the official decision about phase 2 is made is higher than in figure 7 when all housing sizes are considered. On the other hand, when considering only studio apartments there is no increasing effect at that time on average prices in Kivenlahti.

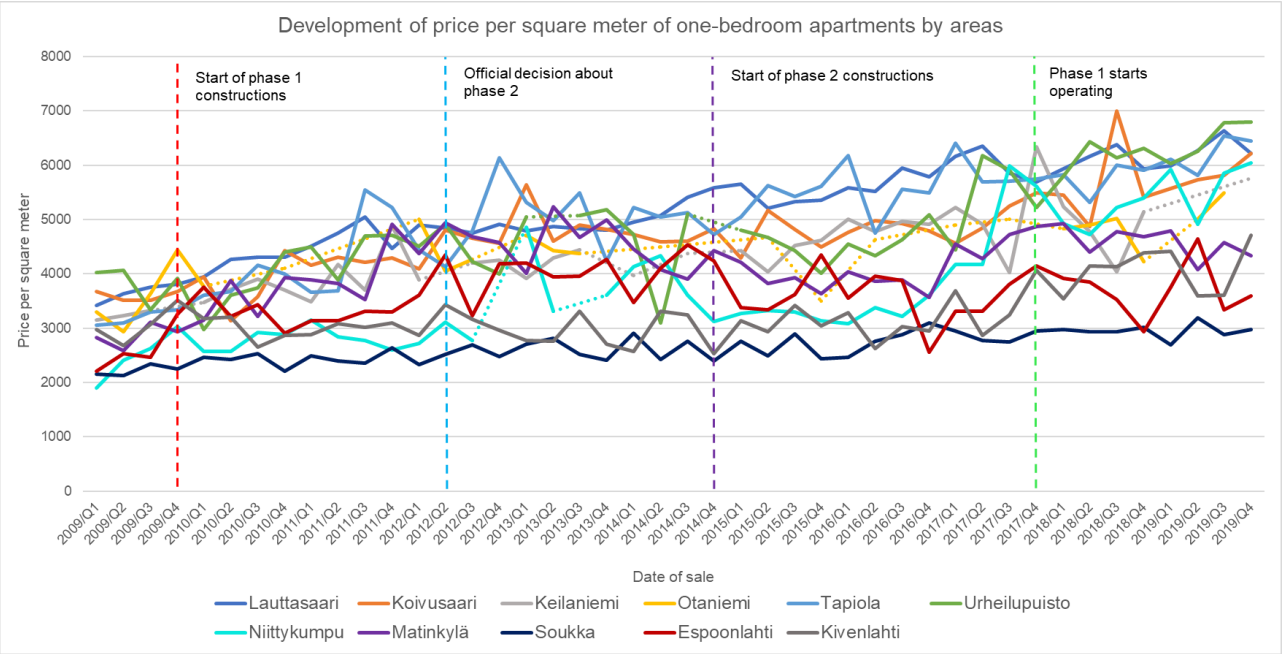

Figure 9 Development of average price per square meter of one-bedroom apartments by areas in 2009-2019

Price development in 2009-2019 of one-bedroom apartments is presented in figure 9. It seems that average prices per square meter in Espoonlahti are higher than in Niittykumpu most of the time in 2009-2016. During 2016, approximately a year before phase 1 starts operating, there is a large increase in average prices in Niittykumpu. Similarly, as in figures 7 and 8, there are no significant high peaks or lows in the housing price development of Soukka. Once again, quarters with no sold apartments are presented by dotted lines.

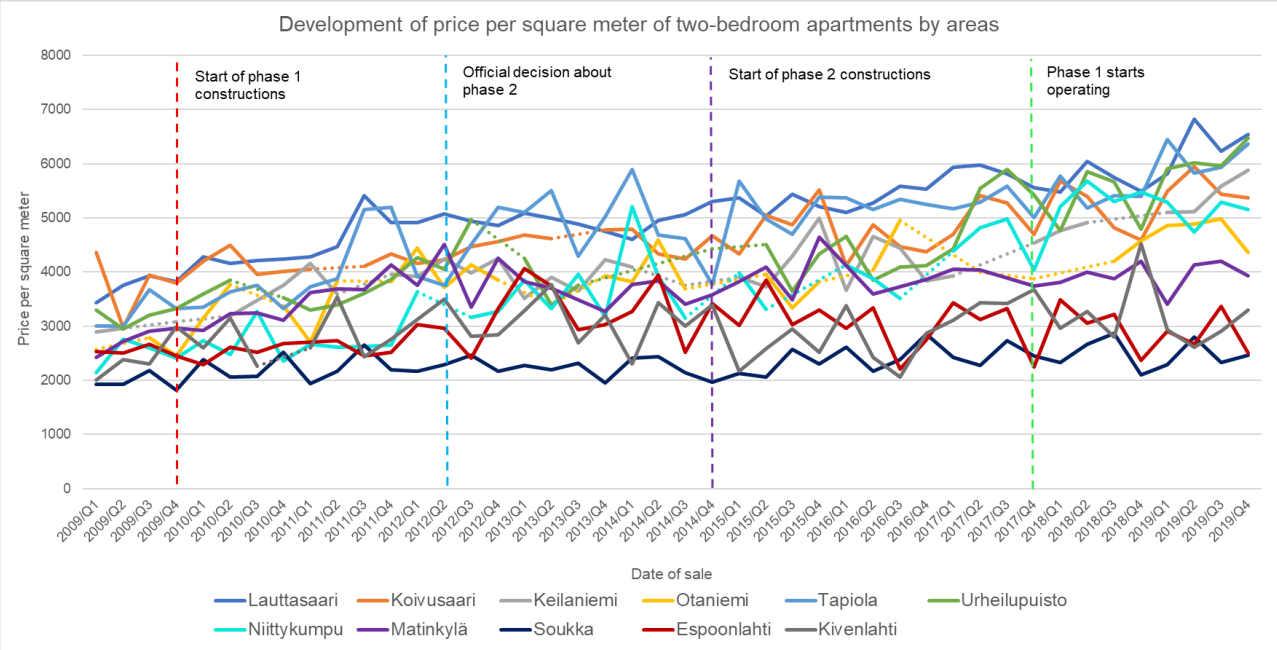

Figure 10 Development of average price per square meter of two-bedroom apartments by areas in 2009-2019

There are decrease in several areas at the time phase 1 starts operating, but in addition to phase 1 areas, there is a decrease in average prices per square meter at that time in Espoonlahti as well, when considering two-bedroom apartments in figure 10. In comparison to the price development of one-bedroom apartments, in figure 9, when considering two-bedroom apartments in figure 10, Niittykumpu starts to diverge from Espoonlahti many years earlier, around 2015-2016.

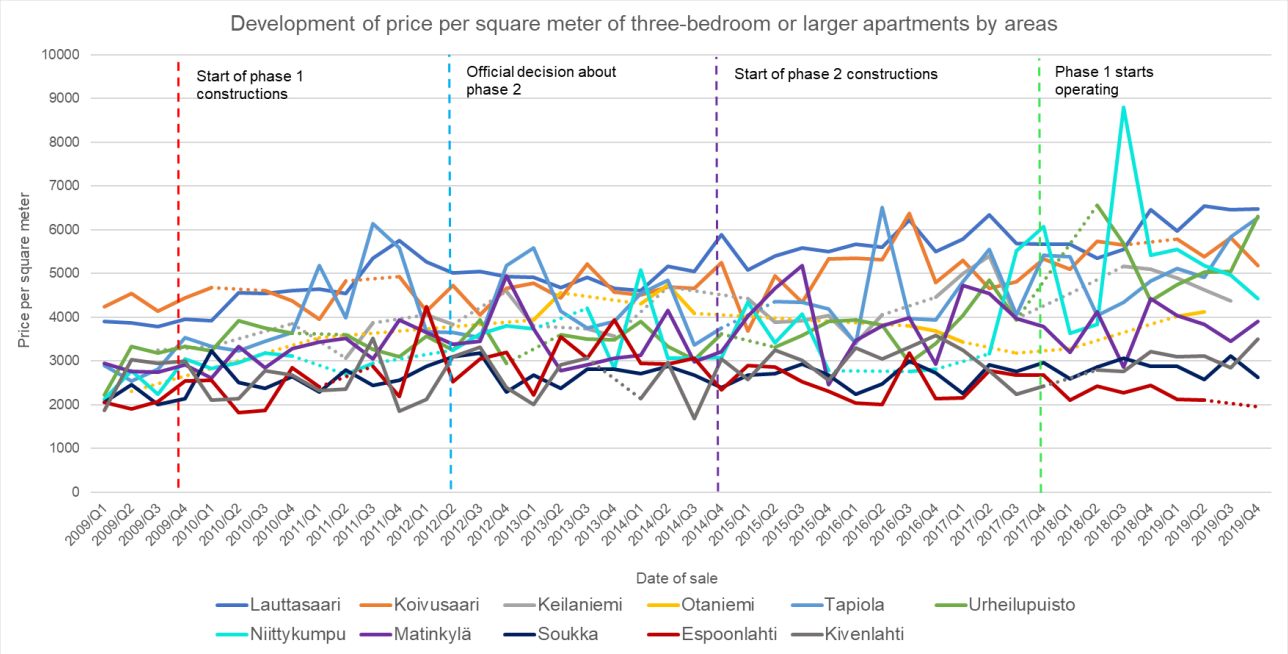

Figure 11 Development of average price per square meter of three-bedroom or larger apartments by areas in 2009-2019

In figure 11 price development in 2009-2019 for real estate of three-bedroom or larger is presented. There is a large quarterly variation in several areas, but the most notable peaks are in Tapiola and Niittykumpu. The high peak in Niittykumpu in Q3/2018 can be explained by the small number of observations: there is only one sold real estate during that quarter in Niittykumpu that has three or more bedrooms, however, this information cannot be seen from figure 11. Outliers were removed from the data, but because different qualities of the apartments such as year built, create variation on housing prices and if the number of observations is small, the specific observation has a significant effect on specific quarter's average prices which happens to be the case in Niittykumpu in figure 11.

3.2 Methodology

In this research quadratic function of OLS regression is used as a method to predict housing prices. A linear function might not be able to explain as accurately the relationship of two variables, so a quadratic function can be used (Brooks 2014, 31). Performances of the linear and quadratic regression models are compared, and the quadratic regression showed better accuracy. For modeling continuous data, especially with a combination of dummy variables and transformed data, OLS regression is a powerful technique. (Hutcheson & Sofroniou 1999, 55-56) As the dataset includes several dummy variables, OLS seemed an appropriate model to work with them as well as with the housing price data.

Data needs to be tested, whether the OLS estimator is BLUE and these test results are analyzed in chapter 3.2.1. In chapter 3.2.2 linear and quadratic regression models for different combination of variables in the datasets are compared to find a model with the best fit for the phase 2 data. Because the model will be used for predicting housing prices, the dataset is split randomly into training and test set, using a ratio 70-30, so that the prediction accuracy of the model can be evaluated. However, as the aim of this study is to predict the housing price development for the time Westmetro’s phase 2 starts operating and the estimated time for Westmetro’s phase 2 to start operating is in 2023, the effect of start of the operating cannot be find out based on data from 2009-2019 from phase 2 areas. For this reason, the data from phase 1 areas is collected from 2007-2019. Because the official decision about Westmetro's phase 1 was made in May 2008, 2007 is used as a first year for data collection so that the period before the official decision is included in the data as well. A separate regression model for phase 1 areas is conducted, to get the effect of the metro's start to operate and it is discussed in chapter 3.2.3. Chapter 3.2.4 presents the observations that are used to create the predictions for the years 2020-2023. Bank of Finland's predictions for macroeconomic variables for 2020-2023 are used to create the predictions. However, because this study is done during the global COVID-19 pandemic, alternative predictions are created by using the Bank of Finland predictions for the macroeconomic variables that were created before the pandemic. Data for the alternative predictions is presented in chapter 3.2.5. In chapter 3.2.6, an alternative regression model for the prediction is created by excluding the information about the Westmetro from the data. Prediction results can be then compared, whether there is a difference in the results when the information about the different stages of the Westmetro is included or not in the data.

3.2.1 Testing the data

Once the data is split into training and test sets, the data needs to be tested whether the OLS estimator is BLUE and the training set is used for this purpose. If these assumptions are violated parameter estimates are biased and their standard errors might be wrong. If the assumptions, presented in chapter 2.6.3, fail, the model can be improved by transforming variables for example into logarithmic or dummy form. (Puumalainen 2019)

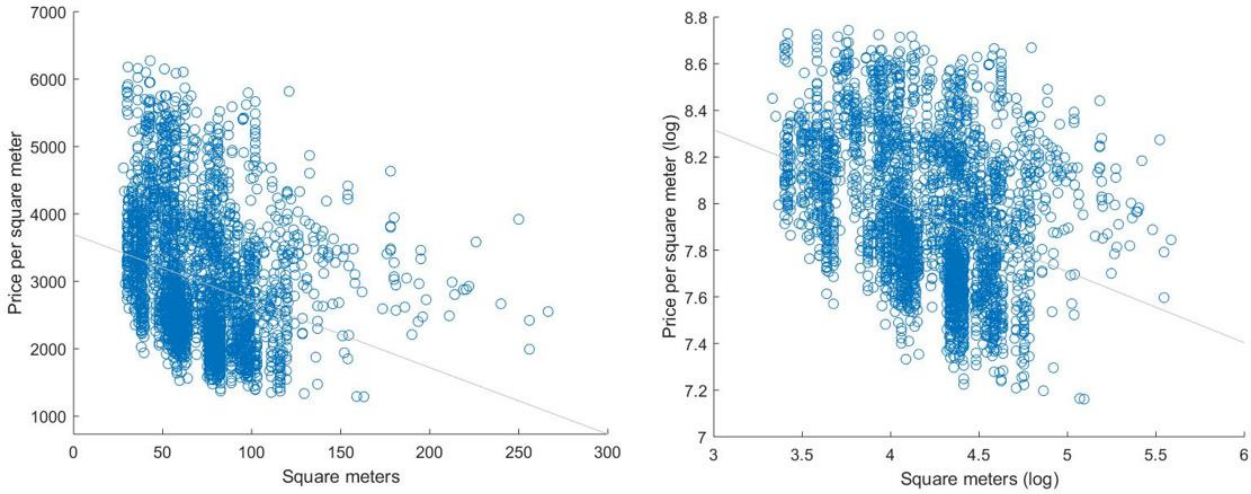

Figure 12 Relationship of price per square meter and square meters

Logarithmic transformations can be useful for variables for at least three reasons: it can help to make the right-skewed distribution closer to a normal distribution as well as help to make the non-linear relationship into a linear one and with the heteroscedasticity problems by making the variance more constant by rescaling the data. (Brooks 2014, 33) To reduce non-linearity, logarithmic transformations are done for most of the continuous variables: Price per square meter (dependent variable) and independent variables: distance to city center, distance to metro station, square meters, floor, year built, date of sale, and CPI. In figure 12, the relationship between price per square meter and square meter is presented in normal and logarithmic form. It can be seen how transforming values into logarithmic form reduces non-linearity.

The dependent variable and residuals should be normally distributed. One of the most common tests applied for normality is Bera-Jarque. The null hypothesis of the test is

normality, and it should be rejected if the data or residuals are significantly skewed, leptokurtic, or platykurtic. (Brooks 2014, 209-210)

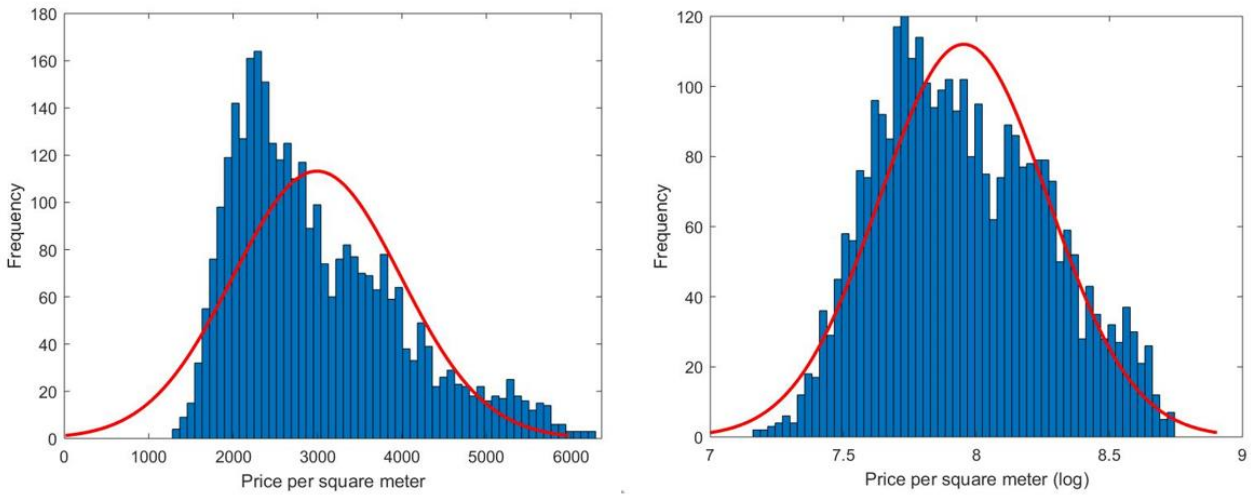

Figure 13 Distribution of the dependent variable

In figure 13, the distribution of the dependent variable, price per square meter, is presented in normal and logarithmic form. In the histogram presented on the left side, the distribution is right-skewed. Once the variable is transformed into logarithmic form, the distribution is closer to be normally distributed. Bera-Jarque test is applied for the logarithmic form of price per square meter and the null hypothesis is rejected, meaning the data is not normally distributed.

Under the BLUE assumptions, there should not be perfect multicollinearity in the data, so the explanatory variables should not correlate strongly with each other. The correlation matrix of the dataset is presented in appendix 2. The highest correlations are mostly between the sale date and other time-related independent variables, such as the stage of metro constructions and with all the macroeconomic variables. The highest correlation is between the sale of date and CPI, 0,95. In addition to that, almost as high correlation is between variables square meters and the number of rooms as they get a 0,94 correlation. Multicollinearity problem might appear if the correlation between explanatory variables is over 0,9 (KvantiMOTV 2003).

Table 4 Variance inflation factors (VIF) of independent variables

Variable	VIF	Variable	VIF
D_apt. buildings	2,04	Rooms	8,98
D_houses	1,55	Log date of sale	18,17
D_espoonlahti	2,48	Log CPI	25,51
D_kivenlahti	5,21	Employment rate	1,88
Log dist. to city center	4,09	Euribor 3 months	10,55
Log dist. To metro station	1,47	D_stage_decision	8,28
Log square meters	9,22	D_stage_construction	17,77
Log year built	1,47	D_site ownership	1,02
Log floor	1,15		

However, it is not always possible to observe multicollinearity problems from correlation coefficients and Variance inflation factors (VIF) can be used to measure multicollinearity (KvantiMOTV 2003). VIF values of independent variables are presented in table 4. In table 4, "D" refers to dummy variable and "Log" refers to that variable is used in logarithmic form. When VIF values were measured for all the independent variables, some of the variables have VIF over 10, which is often considered as limiting value that multicollinearity problems occur. It should be still noted that sometimes removing variables might do more harm than good. (O'Brien 2007, 683-684) Also, if the model is only used for predicting and the coefficients are not interpreted, multicollinearity will not cause large damage (Puumalainen 2019). As the model in this study is used for predicting and the values that have VIF value over 10 are considered essential for the prediction, it is decided that variables are still included in the data.

3.2.2 Linear and quadratic regression models

Once the data is tested, OLS regressions for phase 2 housing price data are conducted by using fitlm function in Matlab. The dataset includes house types: houses, apartment buildings, and terrace houses. Four regression models are conducted: a linear model for the dataset that includes all house types, a quadratic stepwise model for the dataset that includes all house types, a linear model for a dataset that includes only apartment buildings and terrace houses, and a quadratic stepwise model for the dataset that includes only apartment buildings and terrace houses. A separate model for houses is not conducted due to a small number of observations, 131. When all house types are combined into one dataset, an assumption is made that the same factors affect similarly for all house types.

For this reason, models are conducted for different datasets, to see whether there is a difference between how the models fit on data. Logarithmic forms of the specific continuous variables are used, as presented in previous chapter 3.2.1.

As the aim of this study is to predict the development of housing prices for Westmetro's phase 2 areas in 2020-2023, the data is split into training and tests by using a 70-30 ratio. 70-30 ratio is typically used in data splitting. Model is improved through training set and can be tested by making predictions against the test set. The accuracy of the model's predictions can be determined as the test set already has values for the dependent variable that is predicted. (Microsoft 2018)

The quadratic function of OLS regression is conducted and the regression model created by using quadratic function, which has an intercept term, linear and squared terms for each variable, and interaction variables from all products of pairs of the independent variables. Stepwise selection is used to determine the best fit of the quadratic model and the used function in Matlab is stepwiselm. It searches for terms to add to or remove from the model, at each step to determine the best fit of the model. (MathWorks 2020) Matlab code for linear and quadratic stepwise regression for all house types is presented in appendix 3.

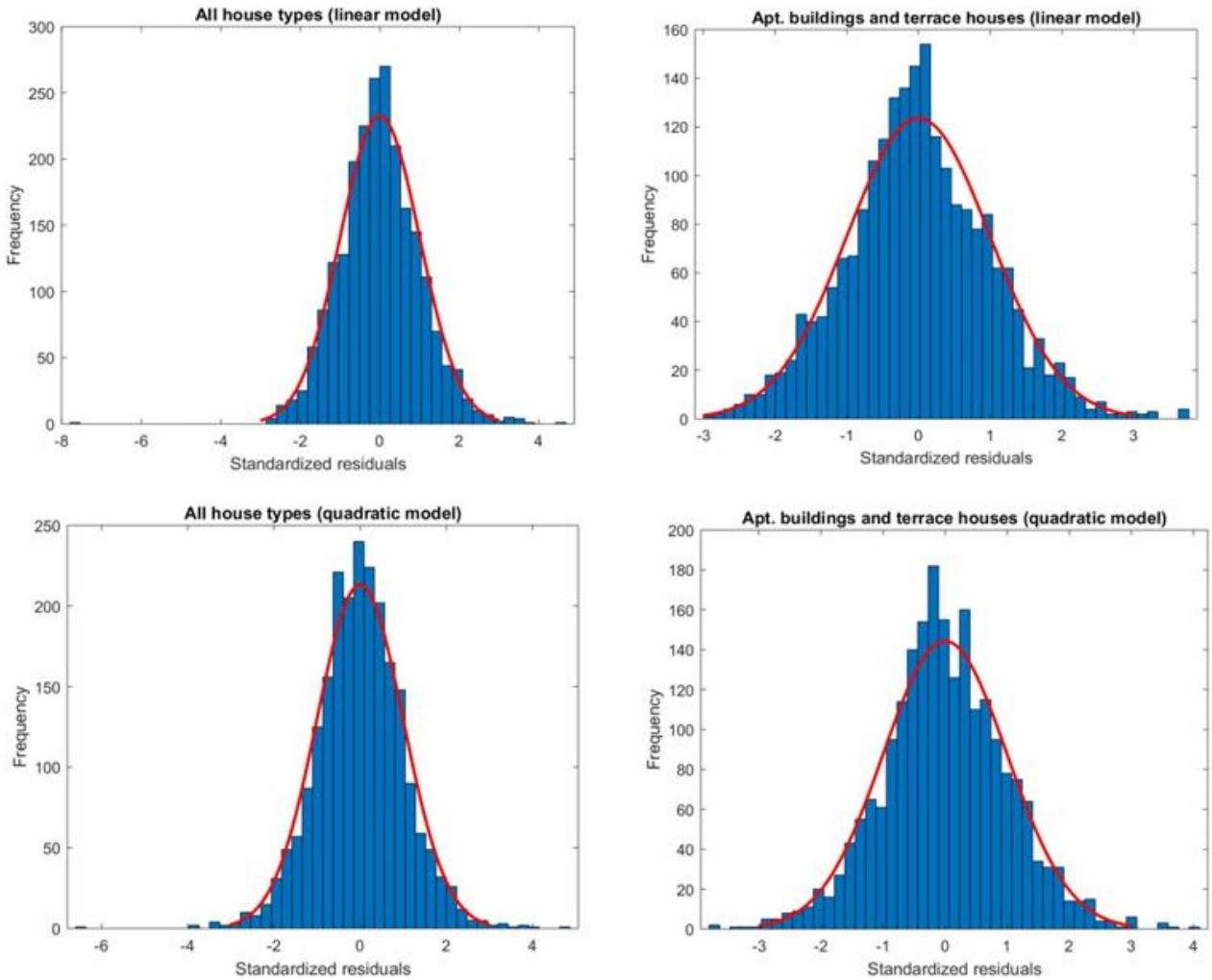

Figure 14 Normal distributions of standardized residuals of linear and quadratic all house types -models and apartment buildings and terrace houses -models

Normal distribution of the residuals is tested for linear and quadratic all house types -models as well as apartment buildings and terrace houses -models with Bera-Jarque test and the null hypothesis is rejected, meaning the data is not normally distributed. However, when observing normal distribution graphs of the standardized residuals of both models in figure 14, the residuals seem to be close to normally distributed, but the normal distribution of all house types model seems a bit more leptokurtic.

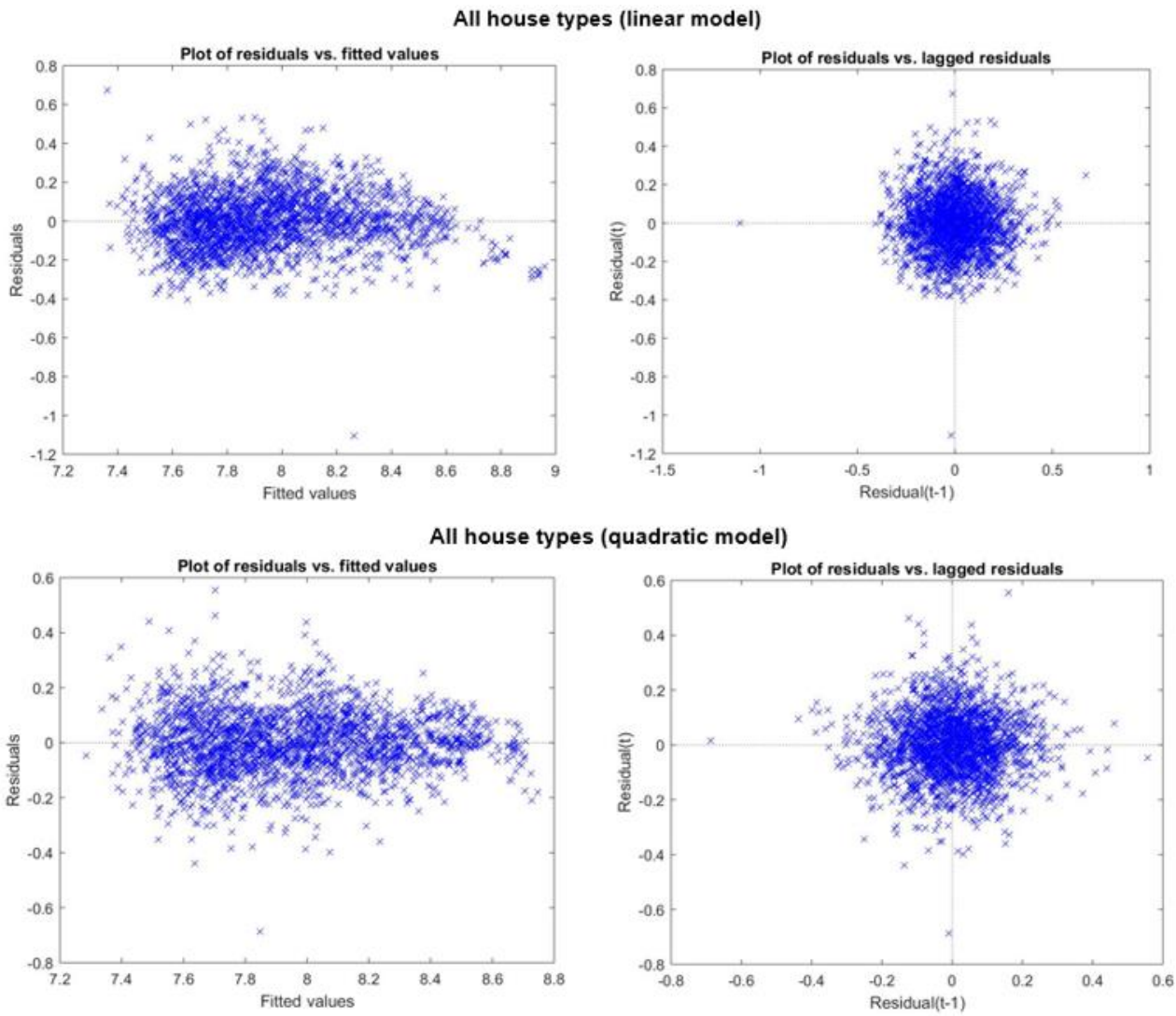

Figure 15 Plots of quadratic stepwise model of all house types: residuals vs fitted values and residuals vs lagged residuals

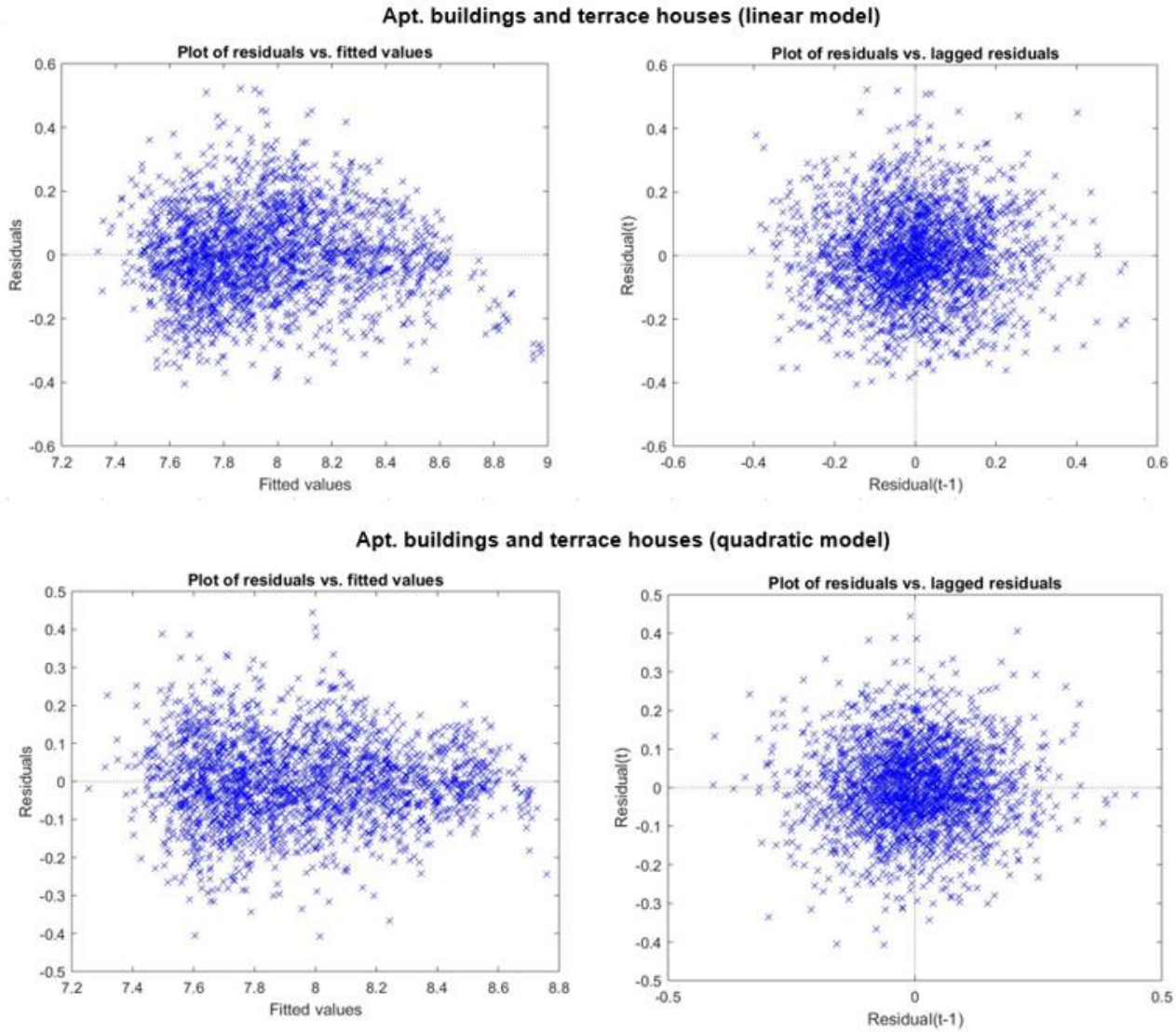

Figure 16 Plots of quadratic stepwise model of apartment buildings and terrace houses: residuals vs fitted values and residuals vs lagged residuals

In figures 15 and 16, on the left side of the figures, residuals are plotted against fitted values. In both figures 15 and 16, in the upper part of the figures are the plots from the linear models and in the bottom part of the figures are the plots from the quadratic model. Based on the left side of plots 15 and 16, the errors are not heteroscedastic. The variance of the errors is not systematically increasing when the fitted values increase. On the right side of the figures, 15 and 16 residuals are plotted against lagged values, and residuals of all models are evenly divided into all four quadrants indicating there is no autocorrelation. Autocorrelation is tested with Durbin-Watson test, which tests for the first-order autocorrelation. Durbin-Watson results are on a scale 0-4, where 0 refers to positive autocorrelation, 2 refers to no autocorrelation, and 4 refers to negative autocorrelation. (Brooks 2014, 194-196) Durbin-

Watson test is conducted for all four models: linear and quadratic models for all house types as well as for apartment buildings and terrace houses and the result for every model is presented below in table 6. For linear and quadratic stepwise models for all house types, results are close to value 2 as Durbin-Watson gets a value of 2,03 for the linear model and 2,01 for the quadratic stepwise model. For apartment buildings and terrace houses, regression models Durbin-Watson gets a value of 1,98 for the linear model and 1,97 for the quadratic stepwise model. Values for all four models presented in table 6 are very close to 2, meaning there is no autocorrelation.

Table 5 Summary of the regression models

	All house types	***Quadratic stepwise model All house types***	***Apt. Buildings and terrace houses***	***Quadratic stepwise model Apt. buildings and terrace houses***
Number of observations	2243	2243	2152	2152
Root Mean Squared Error	0,147	0,116	0,14	0,111
R-squared	0,794	0,873	0,815	0,886
Adjusted R-Squared	0,792	0,869	0,813	0,883
Durbin-Watson	2,03	2,01	1,98	1,97

In table 5, summaries of the phase 2's linear model for all house types, the quadratic stepwise model for all house types, the linear model for apartment buildings and terrace houses along the quadratic stepwise model for apartment buildings and terrace houses are presented. The number of observations refers to the number of observations in the train set. It can be seen that quadratic stepwise models for both datasets have significantly higher adjusted R-squared values and lower RMSEs compared to linear models. Based on these statistics presented in table 5, it seems that the quadratic stepwise model for apartment buildings and terrace houses would have the best fit for the data. Outputs of the linear and quadratic stepwise models for all house types are presented in appendix 4 and for apartment buildings and terrace houses in appendix 5.

Table 6 Accuracy of the models

	All house types	***Quadratic stepwise model All house types***	***Apt. buildings and terrace houses***	***Quadratic stepwise model Apt. buildings and terrace houses***
MSE (log values)	0,0176	0,0130	0,0190	0,0144
RMSE (log values)	0,1327	0,1139	0,1380	0,1201
MSE	174736,33	110458,08	187609,44	114072,02
RMSE	418,01	332,35	433,14	337,75

The performance of all four models is evaluated by creating the valuations by using test sets. The accuracy of the models can be compared as the test set already has values for

the dependent variable that is defined. MSE and RMSE are calculated for logarithmic values of the predictions and for values that are transformed back to functional form. MSE and RMSE values are presented in table 6. As it was mentioned in chapter 2.6.1, MSE values can be compared with those of other models for similar data and prediction timeframe and the model with the lowest error measuring values is the most accurate (Brooks 2014, 292-293). The quadratic stepwise model that includes all housing types shows better performance compared to others as it has the smallest MSE and RMSE values. Especially, when comparing the MSEs of the functional form values, the difference between the linear models and quadratic stepwise models is significant. In table 5 where the goodness of fit statistics are presented, it seems that the quadratic stepwise model for apartment buildings and terrace houses would have a better fit than the quadratic stepwise model for all house types. However, as the quadratic stepwise model for all house types gives the best accuracy, it is decided that the actual predictions are created by using that dataset and model. This way also the price development for houses can be predicted as well. The quadratic stepwise regression model is conducted 10 times for all house types data by using different training and test set to evaluate the consistency of the regression model. For the conducted models, RMSE values variated between: 0,110-0,116, R-Squared: 0,873-0,883 and adjusted R-Squared: 0,869-0,879. Based on this, the model gives relatively similar results for the regression despite the division of training and test sets. The formula for the phase 2's quadratic stepwise regression model for all house types is the following:

$$Ln(Price\ per\ square\ meter) = \beta_0 + \beta_1 * D\_Apartment\ building + \beta_2 * D\_House + \beta_3 * D\_Espoonlahti + \beta_4 * D\_Kivenlahti + \beta_5 * Ln(Distance\ to\ city\ center) + \beta_6 * Ln(Distance\ to\ metro\ station) + \beta_7 * Ln(Square\ meters) + \beta_8 * Ln(Year\ built) + \beta_9 * Ln(Floor) + \beta_{10} * Rooms + \beta_{11} * Ln(Date\ of\ sale) + \beta_{12} * Ln(CPI) + \beta_{13} * Employment\ rate + \beta_{14} * Euribor\ 3\ months + \beta_{15} * D\_stage\_decision + \beta_{16} * D\_stage\_construction + \beta_{17} * Site\ ownership + \beta_{18} * D\_Apartment\ building * D\_Kivenlahti + \cdots + \beta_{66} * Euribor\ 3\ months\text{^}2$$

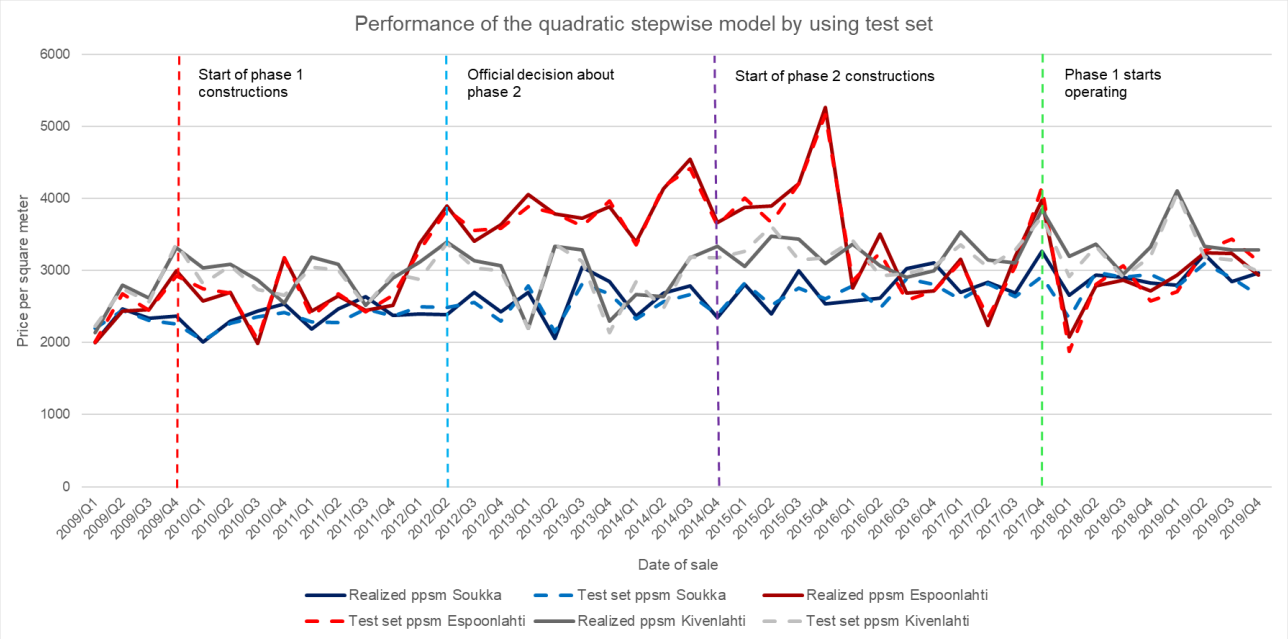

Figure 17 Performance of the quadratic stepwise model by using test set

In figure 17, the performance of the quadratic stepwise model for all house types by using the test set for Espoonlahti, Kivenlahti, and Soukka in 2009-2019 are presented. Output of the quadratic stepwise model for all house types that is used to test the performance of the test set, is presented in appendix 4 and summary above in table 5. Most of the time test set valuations seem to be relatively close to the realized price, but there are few points where there are some small differences on average prices, for example in Soukka around Q4/2017, when the Westmetro's phase 1 started operating.

In the conducted models, the average of quarterly Euribor 3 months values are included in the dataset, as the Bank of Finland's predictions for years 2020-2023 are for Euribor 3 months and they will be used to create the predictions further in this study. However, it is more common that a mortgage is tied into Euribor 12 months. For this reason, OLS regression is conducted by using the quadratic stepwise model for all house types dataset that includes an average of quarterly Euribor 12 months values, just to compare the models with different interest rates. As presented in figure 4 in chapter 3.1, Euribor 12 months rate is higher than Euribor 3 months, but there is no significant difference between them. The output of the Euribor 12 months model is presented in appendix 6. This model cannot be used in the actual predictions as there are no predicted values for Euribor 12 months.

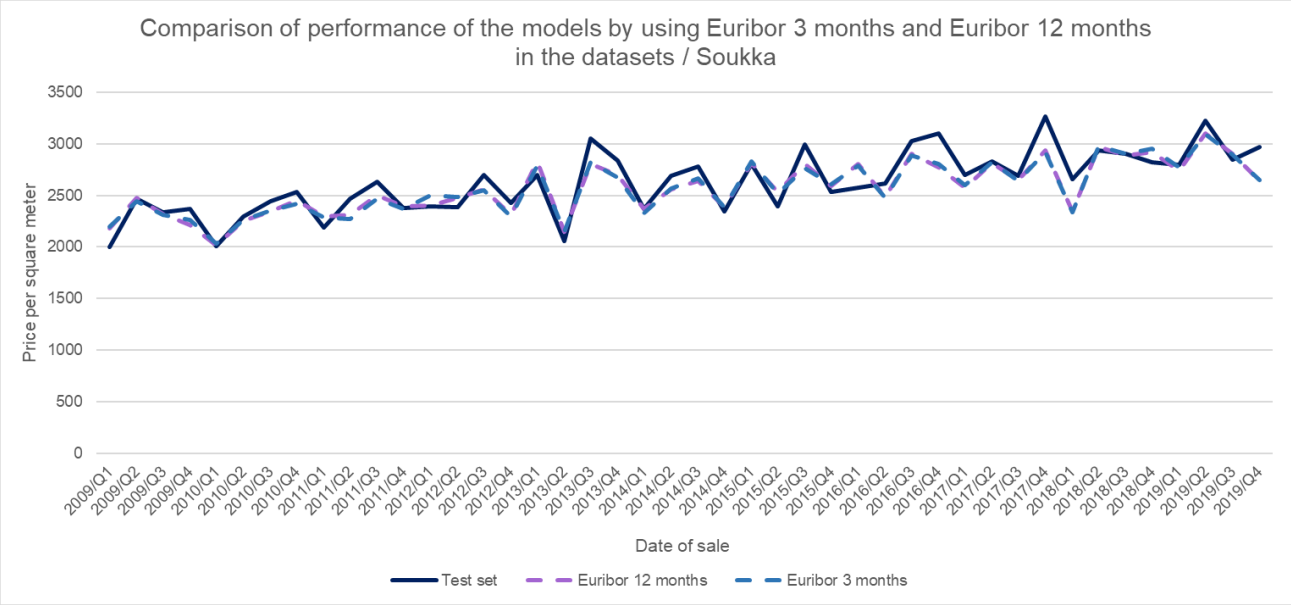

Figure 18 Performances of the regression models in Soukka by using Euribor 3 months and Euribor 12 months in the datasets

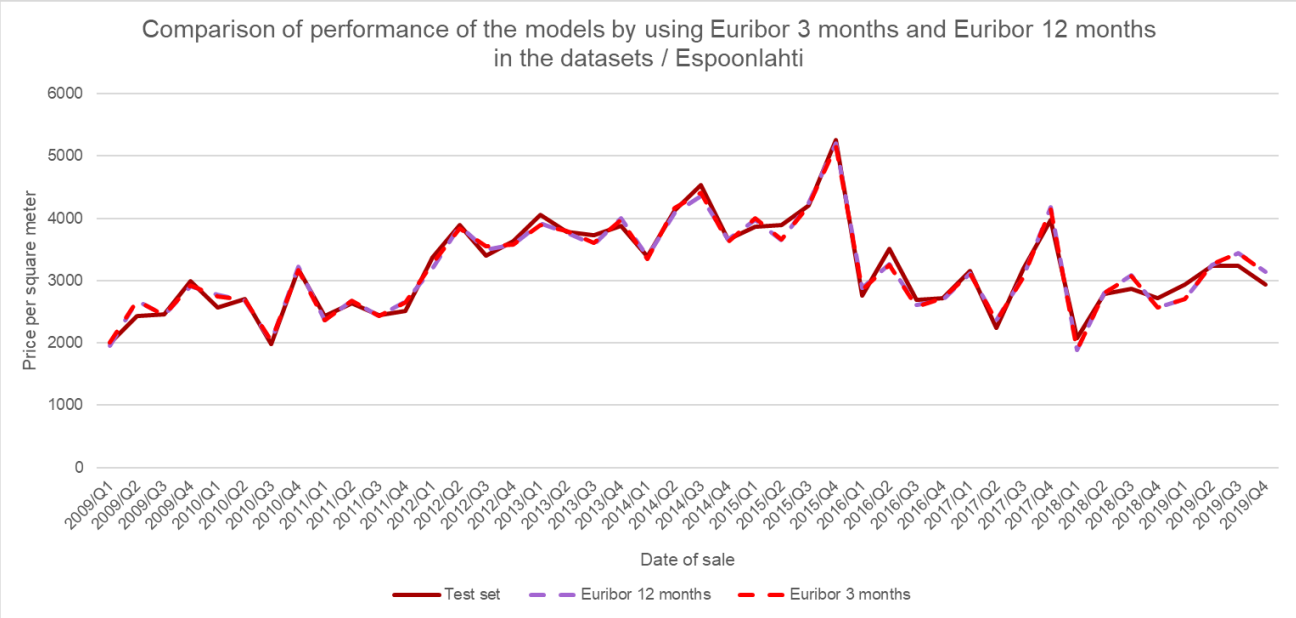

Figure 19 Performances of the regression models in Espoonlahti by using Euribor 3 months and Euribor 12 months in the datasets

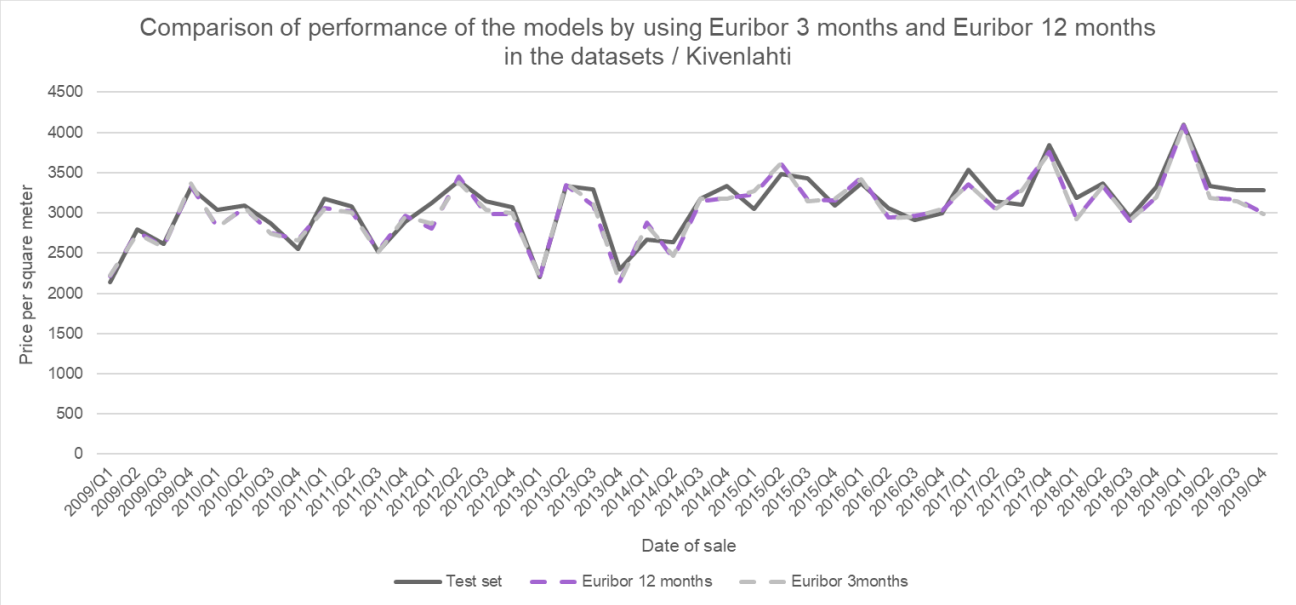

Figure 20 Performances of the regression models in Kivenlahti by using Euribor 3 months and Euribor 12 months in the datasets

In figures 18, 19, and 20 are presented the test set values of price per square meter as well as the defined price per square meter values by using variable Euribor 3 months and Euribor 12 months in the datasets for all house types model. The train and test sets include the same observations so that models can be compared. The defined values are again presented by dash lines and quarterly average values of prices per square meter are used in the graphs. Based on the graphs, it seems that both models, Euribor 3 months and 12 months are giving similar prediction results when using test set data as the results, presented by dash lines, are on top of each other in the graphs.

Table 7 Accuracy of the quadratic stepwise model by using Euribor 12 months

	Quadratic stepwise model Euribor 12 months
MSE (log values)	0,0130
RMSE (log values)	0,1140
MSE	110574,38
RMSE	332,53

As presented in table 7 MSE for Euribor 12 months model in functional form is 110574,38 and RMSE 332,53. Responding values for Euribor 3 months model are 110458,08 and 332,35, respectively. This indicates that even though Euribor 12 months is the more commonly used rate in mortgages and there is some difference in interest rates, using

Euribor 3 months in the predictions does not affect the reliability of the prediction model based on historical data, and based on MSE and RMSE values, the regression model using Euribor 3 months seems to perform a bit better than the one using Euribor 12 months, however, there is no large difference between the performances.

3.2.3 Equation for the prediction

Because the aim of this study is to predict the housing price development for the time Westmetro's phase 2 starts operating and the estimated time for Westmetro's phase 2 to start operating is in 2023, the effect of the start of the operating cannot find out based on data from 2009-2019 from Soukka, Espoonlahti, and Kivenlahti, which was used for conducting the quadratic stepwise model for phase 2 areas. A quadratic regression model for phase 1 areas is conducted, to get the effect of the metro's start to operate. Based on data description and presentation in chapter 3.1, areas Niittykumpu and Matinkylä are the most similar areas with phase 2 areas based on the price level, before the effects of Westmetro's phase 1 are capitalized on housing prices. Because the official decision about Westmetro's phase 1 was made in May 2008, the data from phase 1 areas is collected from 2007-2019. This way, similar variables of the stages of the Westmetro project can be created for phase 1 data as there is for phase 2 data. In addition to variables presented in chapter 3.1 in table 3, phase 1 data gets a dummy variable "stage of metro: operating". Westmetro's phase 1 started first operating in December 2017. The quadratic stepwise regression model is conducted for phase 1 data.

The output of the quadratic stepwise model for phase 1 areas is presented in appendix 7. Coefficients for interaction between variables x8 Niittykumpu : x23 Stage of metro: operating and x9 Matinkylä : x23 Stage of metro: operating are presented in appendix 7. As mentioned earlier, Niittykumpu and Matinkylä were considered as most similar areas to phase 2 areas based on the price level and also, based on distance to the city center. For this reason, coefficients of these interaction variables will be used for phase 2 predictions for the year 2023, when is the expected start of operating of Westmetro's phase 2. The coefficient that is used in the predictions for phase 2 areas is an average of coefficients of interaction variables x8 Niittykumpu : x23 Stage of metro: operating and x9 Matinkylä : x23 Stage of metro: operating.

Shopping center Lippulaiva is under construction in Espoonlahti in the phase 2 area and the estimated opening time for the shopping center is in 2022 (Citycon 2020). For this reason, the effect of a shopping center on a specific area is observed from the output of the quadratic stepwise model for phase 1. Dummy variable "shopping center" is created for phase 1 dataset, where 1=there is a shopping center in the proximity of the metro station, 0= no shopping center. From phase 1 areas, Matinkylä has value 1 in the shopping center variable throughout the observation time and Tapiola gets a value 1 in Q2/2017. All the other areas have a value 0. However, interaction variable x7 Tapiola : x16 Shopping center is excluded from the output of the quadratic stepwise regression model as insignificant. When the similar model is conducted without the stepwise selection, the interaction variable x7 Tapiola : x16 Shopping center does not get a numeric value. For this reason, there is no coefficient for the shopping center variable, and the variable is not included in the prediction dataset.

The effect of the metro's start of operating should be added into the chosen prediction model. As mentioned above, the coefficient that is used in the predictions for phase 2 areas is an average of coefficients of interaction variables x8 Niittykumpu : x23 Stage of metro: operating and x9 Matinkylä : x23 Stage of metro: operating. This average value is added in the equation of the model, which is chosen for the prediction in chapter 3.2.2 and is presented in red in the equation. Finally, the equation that is used for creating the predictions, which is a combination of quadratic stepwise model for all house types in phase 2, the whole output presented in appendix 4, and quadratic stepwise model for all house types in phase 1 is the following:

$$Ln(Price\ per\ square\ meter) = \beta_0 + \beta_1 * D\_Apartment\ building + \beta_2 * D\_House + \beta_3 * D\_Espoonlahti + \beta_4 * D\_Kivenlahti + \beta_5 * Ln(Distance\ to\ city\ center) + \beta_6 * Ln(Distance\ to\ metro\ station) + \beta_7 * Ln(Square\ meters) + \beta_8 * Ln(Year\ built) + \beta_9 * Ln(Floor) + \beta_{10} * Rooms + \beta_{11} * Ln(Date\ of\ sale) + \beta_{12} * Ln(CPI) + \beta_{13} * Employment\ rate + \beta_{14} * Euribor\ 3\ months + \beta_{15} * D\_stage\_decision + \beta_{16} * D\_stage\_construction + \beta_{17} * Site\ ownership + \beta_{18} * D\_Apartment\ building * D\_Kivenlahti + \cdots + \beta_{66} * Euribor\ 3\ months\text{^}2 + \beta_{67} * D\_stage\_operating$$

3.2.4 Observations used for the prediction

Once the best fitting model is chosen, the predictions are created. The predicted variable is the dependent variable, price per square meter. Based on the updated project plan, phase 2 will start operating in 2023 (Länsimetro 2020c). So, the predictions are created for the years 2020-2023 for phase 2 areas Soukka, Espoonlahti, and Kivenlahti.

The observations for the prediction are created by modifying test set observations and by using sold real estate during Q1-Q3/2020 in the observation areas in Soukka, Espoonlahti, and Kivenlahti as observations. All time and price-related information are removed from the observations and variables Date of sale, CPI, Employment, and Euribor 3 months variables get updated values for the prediction as well as all the stage of metro variables. The total number of observations for the prediction is 3493. Test set observations and Q1-Q3/2020 data are used as a basis for observations used for predictions so that the predictions are created for observations that represent the existing real estate in the area. For example, if some area does not have real estate built in the 1960s or does not have a 20-floor apartment building there are no predictions created for that type of observation for the specific area as the predictions are based on characteristics of an existing real estate in the selected area. So, all the physical features of observation real estate are from the test set and Q1-Q3/2020 data, and only the time-related variables are updated to have future values.

For the Q1-Q3/2020 prediction period, only the sold real estate observations are included, so that it is possible to compare the predicted and realized housing prices and to get the accuracy of the predictions for that time frame. The prediction dataset for Q4/2020-Q4/2023 includes observations from the test set and Q1-Q3/2020 period. By creating a relatively large dataset for the predictions, it is possible to get diversity on the dataset so that there are observations with different apartment qualities for every area and every quarter in the years 2020-2023. This way results can be analyzed in more detail, for example by housing type, apartment size, or year built.

As mentioned above, each observation gets updated value for Date of sale, CPI, Employment, Euribor 3 months, and for the stage of metro variables. Quarterly date of sale values from Q1/2020 to Q4/2023 are given for the observations and every date of sale gets a corresponding date of sale code that is on scale 45-60, corresponding codes for Q1/2009-

Q4/2019 were 1-44. Once each observation gets an updated date of sale value, values of macroeconomic variables are updated to be corresponding to the date of sale. For predicting the development of price per square meter, the Bank of Finland's predictions for macroeconomic variables are used. As the expected start of the metro to operate is in 2023, dummy variable "stage of metro: operating" is added into prediction dataset and variable gets value 1=operating from Q1/2023 onwards. Before Q1/2023 variable gets value 0=not operating. During Q1/2020-Q4/2022 observations have value 1 in dummy variable "stage of metro: under construction".

Table 8 Predictions for macroeconomic variables in 2020-2023 in quarterly values. Quarterly values calculated based on Bank of Finland's (2020b) predictions

	CPI	Employment	Euribor 3 months
Q1/2020	123,66	71,60 %	-0,00400
Q2/2020	123,85	71,52 %	-0,00440
Q3/2020	124,07	71,54 %	-0,00469
Q4/2020	124,33	71,64 %	-0,00488
Q1/2021	124,61	71,80 %	-0,00500
Q2/2021	124,92	72,00 %	-0,00505
Q3/2021	125,27	72,23 %	-0,00506
Q4/2021	125,65	72,47 %	-0,00504
Q1/2022	126,05	72,70 %	-0,00500
Q2/2022	126,48	72,91 %	-0,00496
Q3/2022	126,95	73,07 %	-0,00494
Q4/2022	127,44	73,17 %	-0,00495
Q1/2023	127,96	73,20 %	-0,00500
Q2/2023	128,51	73,13 %	-0,00512
Q3/2023	129,08	72,96 %	-0,00531
Q4/2023	129,69	72,65 %	-0,00560

The Bank of Finland's (2020b) predictions for macroeconomic variables are yearly predictions. In the used dataset macroeconomic variables have quarterly average values and the date of sales are grouped quarterly as well. Cubic spline interpolation is used to get yearly predictions of CPI, Employment rate 15-64 years, and Euribor 3 months into quarterly. By using this method, it is possible to fit a series of cubic polynomials between each of the data points, with the insistence that the created curve is continuous and appears smooth. To determine rates of change and cumulative change over time, cubic splines can be used. (McKinley & Levine n.d.) Predictions for 2020-2023 in quarterly values for the macroeconomic variables are presented in table 8 and Matlab code for the process is presented in appendix 8. In the Bank of Finland's predictions CPI is determined on level

2015=100, so the CPI values are transformed into 2005=100 level, which is used level in the dataset for the quadratic stepwise model as well. These transformed 2005=100 prediction values are presented in table 8.

3.2.5 Consideration of COVID-19 on housing price prediction

Bank of Finland's predictions for macroeconomic variables for 2020-2023, which were presented in chapter 3.2.4, is from December 2020, so the predictions are created during the global COVID-19 pandemic. Challenges created by COVID-19 on predicting are mentioned on specifications of the Bank of Finland's predictions. Due to COVID-19 restrictions, there have been unexpected changes in the employment rate and the long-term effects of COVID-19 on the economy are unclear. For this reason, predictions for phase 2 areas in 2020-2023 are also created by using the Bank of Finland's predictions from December 2019. The predictions for housing prices are created by using an otherwise similar dataset, but the macroeconomic variables have different values. Bank of Finland created predictions also in December 2019, so the predictions were created right before the start of the COVID-19 pandemic. This way it can be seen how predictions for housing prices differ from each other, when the predicted effects of COVID-19 on the economy are included on macroeconomic variables and when they are not.

December 2019's predictions are only for the years 2020-2022, so one more predicted year is needed. Similarly, as in December 2020's predictions, the predicted values are yearly predictions, and they need to be transformed into quarterly values. Cubic spline interpolation is once again used to transform the yearly predictions into quarterly, but also to get predicted values of CPI, Employment rate 15-64 years, and Euribor 3-months for the year 2023. By using this method, it is possible to fit a series of cubic polynomials between each of the data points, with the insistence that the created curve is continuous and appears smooth. To determine rates of change and cumulative change over time, cubic splines can be used. (McKinley & Levine n.d.) Predictions for 2020-2023 in quarterly values for the macroeconomic variables are presented in table 9 and Matlab code for the process is presented in appendix 8. In the Bank of Finland's predictions CPI is determined on level 2015=100, so the CPI values are again transformed into 2005=100 level, which is used level in the dataset for the quadratic stepwise model as well. These transformed 2005=100 prediction values are presented in table 9.

*Table 9 Predictions for macroeconomic variables in 2020-2023 before Covid-19 in quarterly values. Quarterly values and *referred values are calculated based on Bank of Finland's (2019) predictions*

	CPI	Employment	Euribor 3 months
Q1/2020	124,85	72,70 %	-0,0040
Q2/2020	125,25	72,77 %	-0,0041
Q3/2020	125,66	72,84 %	-0,0041
Q4/2020	126,08	72,92 %	-0,0041
Q1/2021	126,52	73,00 %	-0,0040
Q2/2021	126,98	73,09 %	-0,0038
Q3/2021	127,45	73,19 %	-0,0036
Q4/2021	127,93	73,29 %	-0,0033
Q1/2022	128,43	73,40 %	-0,0030
Q2/2022	128,95	73,52 %	-0,0026
Q3/2022	129,48	73,64 %	-0,0021
Q4/2022	130,02	73,77 %	-0,0016
Q1/2023*	130,58	73,90 %	-0,0010
Q2/2023*	131,16	74,04 %	-0,0003
Q3/2023*	131,74	74,19 %	0,0004
Q4/2023*	132,35	74,34 %	0,0012

There are differences between the Bank of Finland's predictions created in December 2019, before the COVID-19, presented in table 9, and predictions created during the COVID-19 pandemic in December 2020, presented in table 8, in chapter 3.2.4. In December 2019 predictions, inflation is predicted to increase faster than in December 2020 predictions. For the employment rate, in December 2020 it was predicted that the employment rate would decrease from 2019 during 2020 and it would overcome the 2019 rate in 2022. In December 2019's predictions employment rate is predicted to increase from 2019 in 2020-2023. In December 2019's predictions Euribor 3 months is predicted to stay low but still to start increasing, while in December 2020's predictions in table 8, Euribor 3 months is predicted to continue decreasing. Comparison of the prediction results by using different values on macroeconomic variables are presented further in the results.

3.2.6 Prediction model without the information about the Westmetro

An alternative prediction model is also created, as the regression model is conducted for the dataset that does not have any stage of metro variables. Model is conducted by using otherwise similar dataset as the quadratic regression model for all house types, which output is presented in appendix 4, but it does not have stage of metro variables in the dataset.

Table 10 Summary and accuracy of the regression model with no stage of metro variables

	Quadratic stepwise model no "stage of metro" -variables
Number of observations	2243
Root Mean Squared Error	0,117
R-squared	0,871
Adjusted R-Squared	0,867
Durbin-Watson	2,01
MSE (log values)	0,0135
RMSE (log values)	0,1162
MSE	121748,55
RMSE	348,92

The summary of the conducted regression model is presented in table 10. In comparison to the quadratic regression model for all house types, which includes the stage of metro variables, and which will be used for the actual predictions, this model has higher MSE and RMSE values, meaning the model with the stage of metro variables has better accuracy as it has MSE and RMSE values of 110458,08 and 332,35, respectively. Predictions for 2020-2023 will be created by using the no stage of metro -variables model as well and a comparison of the prediction results of these two models is presented further in the results chapter.

4 RESULTS

In this chapter, the prediction results for Soukka, Espoonlahti, and Kivenlahti areas for the years 2020-2023 are presented. Predictions are created by using the modified quadratic model for all house types as the coefficient of dummy variable "stage of metro: operating" is added into the model's equation. The coefficient of dummy variable "stage of metro: operating" is average of from interaction variables x8 Niittykumpu : x23 Stage of metro: operating and x9 Matinkylä : x23 Stage of metro: operating from phase 1's quadratic stepwise model for all house types, which is presented in appendix 7. Logarithmic forms of the dependent variable and independent variables are transformed back to functional form so that the results are easier to interpret. Because the study is completed when part of the prediction period is already realized, predictions for Q1-Q3/2020 are created only to sold real estate in Soukka, Espoonlahti, and Kivenlahti, so that it is possible to compare the predicted and realized housing prices and to get the accuracy of the predictions for that time frame.

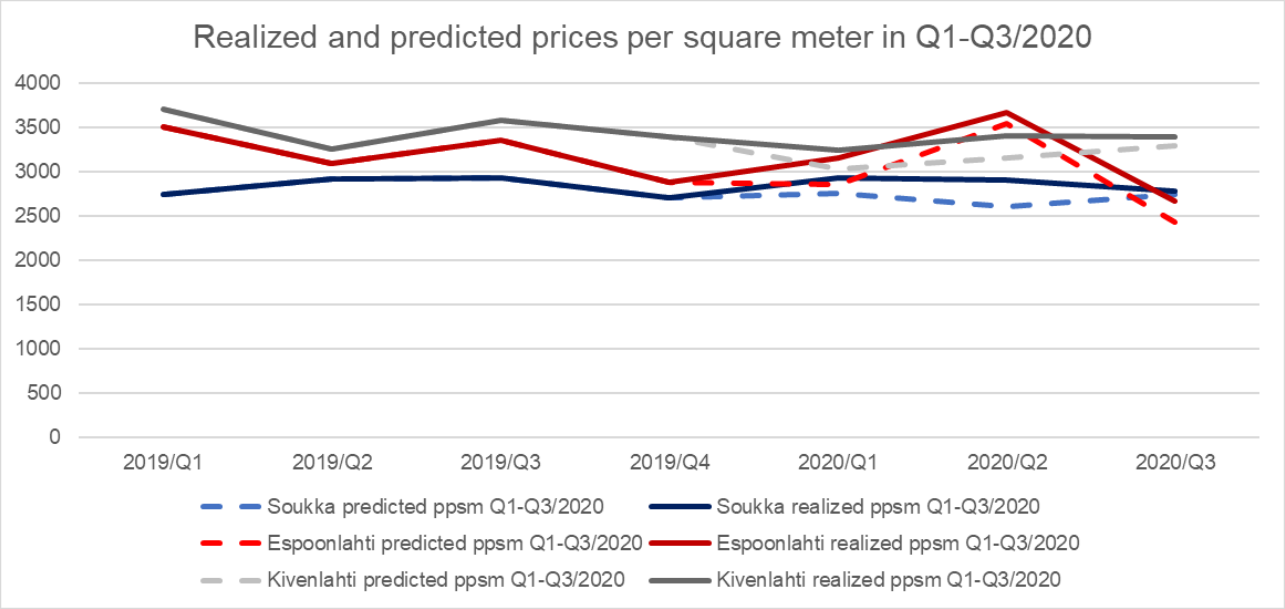

Figure 21 Comparison of realized and predicted average prices per square meter in Q1-Q3/2020 by areas

In figure 21 realized and predicted average prices per square meter in Q1-Q3/2020 are presented. As the housing price data from Q1-Q3/2020 sales was already available, it is compared to the prediction results of this study. In the graphs in figure 21, predictions are presented by the dash lines. As mentioned in chapter 3.2.4, observations used in the

prediction for Q1-Q3/2020 are the same as the realized real estate sales during that time frame, so that predictions can be compared to realized prices. Based on figure 21, it seems that predictions are relatively accurate for Espoonlahti and Soukka, where there are only small differences in average prices between predicted and realized prices. However, in Espoonlahti there is a bit larger misprediction in Q1/2020 and Soukka in Q2/2020. Based on the graph in figure 21, predictions for Kivenlahti are not as accurate as for the two other areas as there seem to be larger mispredictions in two quarters Q1/2020 and Q2/2020.

Table 11 Accuracy of the Q1-Q3/2020 predictions

	Soukka	***Espoonlahti***	***Kivenlahti***
All house types			
MSE (log values)	0,0301	0,0418	0,0268
RMSE (log values)	0,1735	0,2045	0,1636
MSE	300689,88	275099,88	312311,83
RMSE	548,35	524,50	558,85
Apartment buildings			
MSE (log values)	0,0242	0,0433	0,0242
RMSE (log values)	0,1555	0,2082	0,1554
MSE	167585,91	284902,80	281801,05
RMSE	409,37	533,76	530,85
Terrace houses			
MSE (log values)	0,0089	0,0056	0,0656
RMSE (log values)	0,0942	0,0747	0,2561
MSE	77405,95	39829,71	758673,93
RMSE	278,22	199,57	871,02
Houses			
MSE (log values)	0,1249	-	0,0314
RMSE (log values)	0,3533	-	0,1772
MSE	2174462,99	-	399320,53
RMSE	1474,61	-	631,92

In table 11 accuracy of the Q1-Q3/2020 prediction results are presented as the MSE and RMSE values are calculated and presented by areas and house types. In table 6 in chapter 3.2.2, the accuracy of the quadratic model for all house types by using the test set was presented. The accuracy of the model by using the test set was significantly better compared to Q1-Q3/2020 prediction accuracy. However, it has to be noted, that number of observations and timeframe are different. For example, in Kivenlahti there is only one sold house in Q1-Q3/2020 period, so the MSE and RMSE values are based on that observation only. The poorest accuracy compared to others is on houses in Soukka, which has significantly higher MSE and RMSE values than other areas or house types. By observing

prediction results, it can be seen that the model has predicted significantly lower prices per square meter for houses that were built in the 2010s compared to realized prices. In figure 6 in chapter 3.1, the distribution of real estate's year built by areas was presented and from the box plot, it can be seen that there is not many real estate built in the 2000s or 2010s in Soukka. This explains the poor prediction performance of the prediction model for houses built in the 2010s in Soukka.

Below in figures 22-24, realized priced development from 2009-2019 and predicted price development for the whole prediction timeframe 2020-2023 is presented separately for each area: Soukka, Espoonlahti, and Kivenlahti. By combining the realized price development from 2009-2019 and the predicted price development for 2020-2023 it is easier to interpret the prediction results and the predicted price development. The predicted prices for 2020-2023 are presented by dash lines. In the graphs presented in figures 22-24, the price development is presented for all housing types and separately for apartment buildings, terrace houses, and houses.

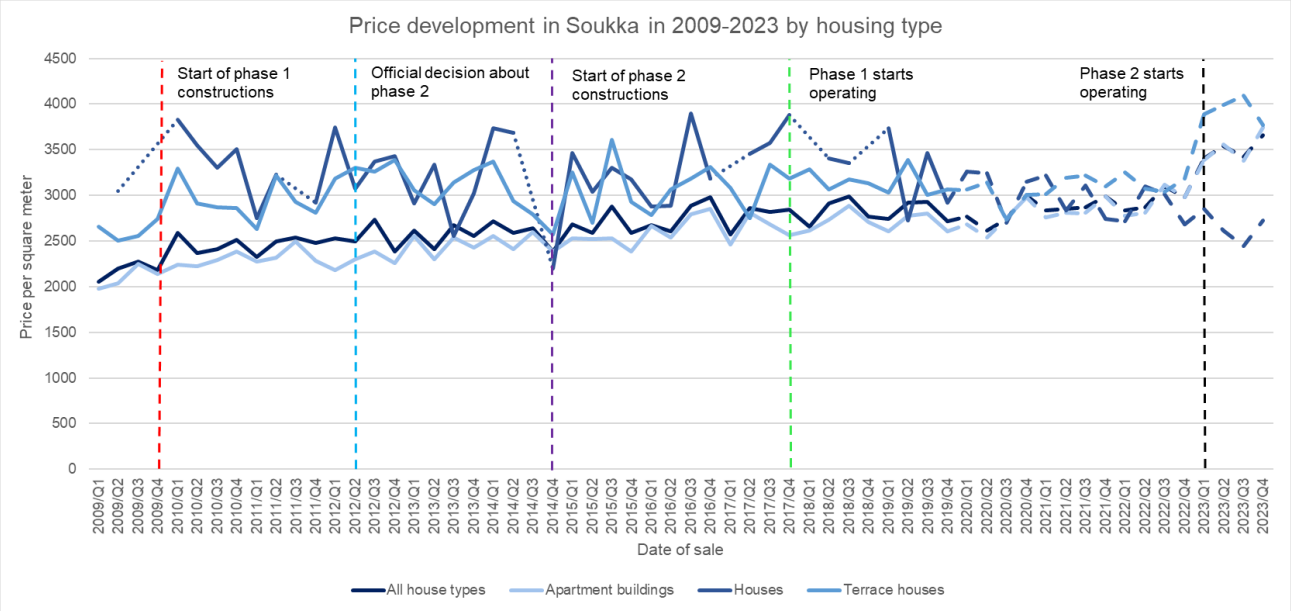

Figure 22 Realized and predicted price development in Soukka

In figure 22 realized and predicted price development in Soukka are presented. Realized average price per square meter is from 2009-2019 and the predicted timeframe is 2020-2023. Similarly, as in chapter 3.1, if there are quarters when there are no sold real estate, these time periods are presented by the dotted line in graphs in figure 22. Based on figure

22, average prices per square meter are increasing during the prediction time frame 2020-2023 when considering all house types, apartment buildings, and terrace houses. The predicted price development for houses seems to be decreasing. However, the predicted decrease in price development for houses does not seem to affect significantly the development of all house types due to the small number of observations from the houses category. In table 3 in chapter 3.1 it was presented that only 4,09 % of the observations are houses when considering all three areas, Soukka, Espoonlahti, and Kivenlahti, while apartment buildings have 87,89 % share of the total number of observations and terrace houses have 8,02 % share. The small number of observations and variation in average prices per square meter in 2009-2019 can explain the prediction results for houses. As the number of observations of houses is small, every observation and its qualities have significant effect on predictions. In appendix 1, the yearly changes in 2010-2019 are presented for each house type. As mentioned in chapter 3.1, it was mentioned that in some of the phase 1 areas, there were a decrease in housing prices after the metro's phase 1 started operating in Q4/2017. Based on the predictions, there is a decrease in average prices of houses after the phase 2 starts operating. For apartment buildings and terrace houses, there is a higher peak in average prices predicted around Q1/2023 when Westmetro's phase 2 is expected to start operating.

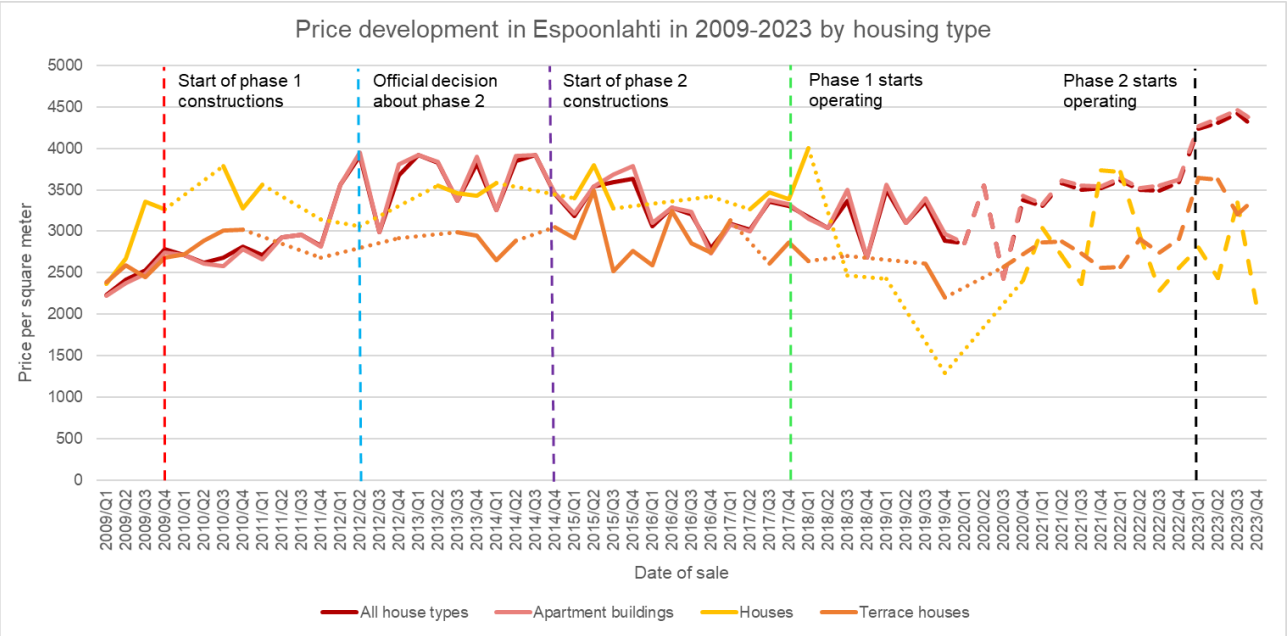

Figure 23 Realized and predicted price development in Espoonlahti

Realized and predicted price development in Espoonlahti in 2009-2023 is presented in figure 23. It can be seen that there are large variations in average prices per square meter for terrace houses and houses during the years 2009-2019, which is the timeframe for the data collection. Once again, if there are quarters when there are no sold real estate, these time periods are presented by the dotted line in graphs in figure 23. If there was a single quarter with observation between the gaps, it is also presented by the dotted line as the single quarter with observation does not form a line.

From the dotted line in figure 23, it can be seen that there is significant decrease in average prices per square meter of houses in 2018 and 2019. As it was mentioned, the number of observations for houses is small and when considering only Espoonlahti, the share of houses is only 3,18 %. This means every observation has a significant effect on price development of houses due to small number of observations. As presented in appendix 1, the overall price development of houses in Espoonlahti is decreasing, as the yearly changes are negative except in 2010, 2012 and 2013. Variative price behavior and small number of observations affect the predictions as well. Also predicted price development for houses seems to have large variations, as there is a high peak in prices predicted for Q4/2021-Q1/2022. The average prices of all house types follow quite closely apartment building average prices, as most of the observations are from apartment buildings and as it was mentioned, there are several quarters when there are no observations from terrace houses or houses. Similarly, as in Soukka, also in Espoonlahti, there is a larger increase in average prices predicted around the time Westmetro's phase 2 is expected to start operating for apartment buildings and terrace houses. For houses, the predicted price development for that time continues as variative, as there are decreases and increases predicted for the quarters in 2023.

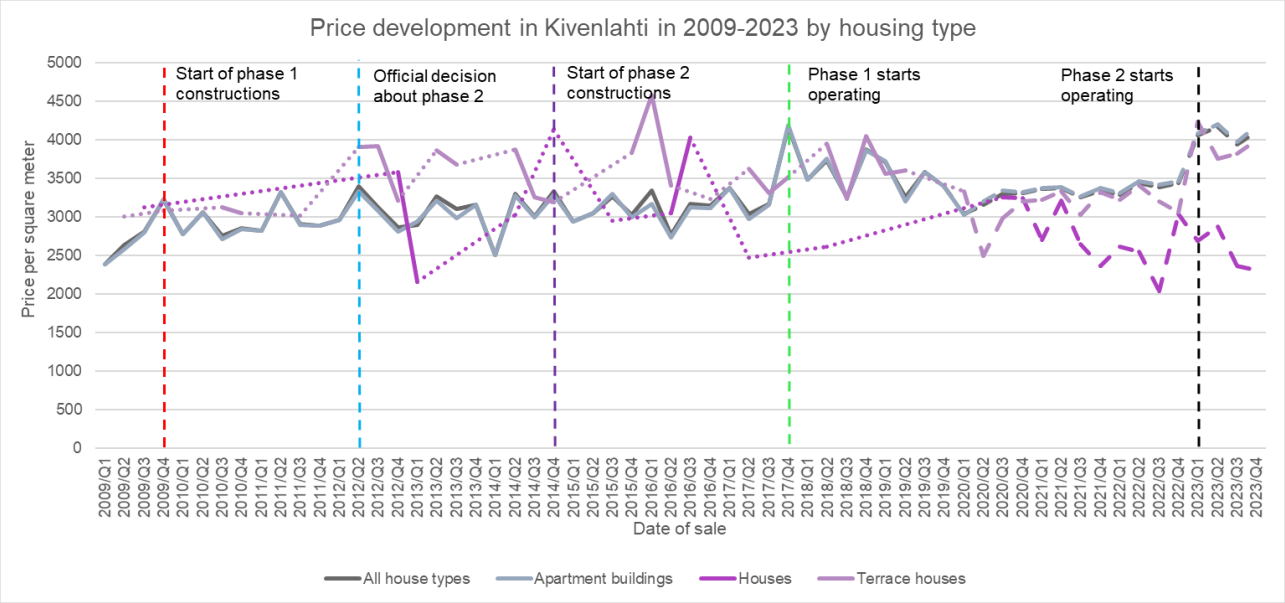

Figure 24 Realized and predicted price development in Kivenlahti

In figure 24 price development by housing types in Kivenlahti is presented. Similarly, as in Espoonlahti, presented above in figure 23, also in Kivenlahti, there are several quarters and years when there were no sold terrace house apartments or houses during the observed time frame 2009-2019 and these are presented by the dotted line. When considering only Kivenlahti, the share of houses is only 1,17 %, which is very low. This causes large variations in average prices per square meter for that category, as due to the low number of observations per quarter, average prices might vary a lot depending on specific sold real estate's qualities. For example, in 2012/Q4 average price for houses is 3584,07 € but in the next quarter 2013/Q1 the average price decreases to 2155,17 €. Overall, the price development for houses is decreasing and increasing throughout the whole observation period 2009-2019, which is also presented in appendix 1 and this variation in price behavior can be seen also in the prediction results for houses in 2020-2023. Once again, the price development of all house types is following the predicted development of apartment buildings and there is a significant increase predicted in average prices for apartment buildings and terrace houses around phase 2's expected start of operating in Q1/2023.

Table 12 Predicted price development by house types

All house types	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2778,20 €	-1,75 %	-2,33 %	3056,14 €	-4,86 %	-5,42 %	3198,77 €	-8,30 %	-8,84 %
2021	2886,91 €	3,91 %	2,97 %	3475,76 €	13,73 %	12,70 %	3334,90 €	4,26 %	3,31 %
2022	2945,99 €	2,05 %	0,74 %	3547,12 €	2,05 %	0,75 %	3390,61 €	1,67 %	0,37 %
2023	3506,91 €	19,04 %	17,12 %	4288,38 €	20,90 %	18,95 %	4068,01 €	19,98 %	18,04 %
Yearly change from 2019 to 2023		**5,53 %**	**4,37 %**		**7,49 %**	**6,31 %**		**3,18 %**	**2,05 %**
Apartment buildings	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2735,26 €	1,50 %	0,90 %	3065,10 €	-5,94 %	-6,50 %	3221,47 €	-7,41 %	-7,95 %
2021	2843,66 €	3,96 %	3,02 %	3508,29 €	14,46 %	13,42 %	3352,91 €	4,08 %	3,14 %
2022	2920,21 €	2,69 %	1,38 %	3584,37 €	2,17 %	0,86 %	3414,67 €	1,84 %	0,54 %
2023	3517,05 €	20,44 %	18,49 %	4348,82 €	21,33 %	19,37 %	4110,78 €	20,39 %	18,44 %
Yearly change from 2019 to 2023		**6,88 %**	**5,71 %**		**7,48 %**	**6,30 %**		**3,42 %**	**2,29 %**
Terrace houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2987,53 €	-4,28 %	-4,85 %	2701,80 €	12,31 %	11,65 %	3018,99 €	-15,63 %	-16,13 %
2021	3131,34 €	4,81 %	3,86 %	2760,27 €	2,16 %	1,24 %	3226,74 €	6,88 %	5,91 %
2022	3136,93 €	0,18 %	-1,10 %	2779,70 €	0,70 %	-0,58 %	3225,36 €	-0,04 %	-1,32 %
2023	3933,36 €	25,39 %	23,36 %	3475,21 €	25,02 %	23,00 %	3950,00 €	22,47 %	20,49 %
Yearly change from 2019 to 2023		**5,95 %**	**4,79 %**		**9,63 %**	**8,43 %**		**1,33 %**	**0,22 %**
Houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3069,71 €	-4,43 %	-4,99 %	2916,13 €	57,02 %	56,10 %	2989,02 €	-	-
2021	2975,45 €	-3,07 %	-3,95 %	2868,75 €	-1,62 %	-2,52 %	2731,87 €	-8,60 %	-9,43 %
2022	2877,08 €	-3,31 %	-4,54 %	2742,35 €	-4,41 %	-5,63 %	2558,93 €	-6,33 %	-7,53 %
2023	2660,27 €	-7,54 %	-9,03 %	2631,89 €	-4,03 %	-5,58 %	2559,50 €	0,02 %	-1,59 %
Yearly change from 2019 (2018*) to 2023		**-4,60 %**	**-5,65 %**		**9,11 %**	**7,91 %**		**-0,57 %***	**-1,53 %***

In table 12, predicted yearly changes for all apartment sizes are presented by areas and by house types. Predicted yearly percentual changes in the areas are presented in both nominal forms as well as in form of real changes, that consider inflation. CPI values that are used to calculate the real changes are average of the quarterly CPI values that are presented in table 8 in chapter 3.2.4. Average of the specific year's quarterly values are used for each year. In the top part of table 12, the results are presented combined: all house types together by areas. Below all house types -section are three categories separately: apartment buildings, terrace houses, and houses. In every category, yearly changes are presented by comparing the considered year's average prices to the previous year. At the bottom of every section yearly change for the prediction time frame, 2020-2023 is presented. Compounding interest was calculated to get the yearly change from 2019 realized average prices to 2023 predicted average prices. If there was no housing price data from 2019 in the category, 2018 average prices were used instead, and these values are presented by red text in table 12 and in the following tables in this chapter. If there was an even longer break in the observations in the selected category, the yearly change is not presented. Ppsm refers to average price per square meter of the year in table 12 and in the following tables in this chapter.

Based on the prediction results presented in table 12, the predicted price development for 2020 is negative for most of the areas and house types. An explanation for this could be the used macroeconomic variables. As mentioned in chapter 3.2.4, the macroeconomic variables used in the prediction dataset are the Bank of Finland's predictions from December 2020. There was a significant decrease in the employment rate in the predicted values for 2020, which affects the predicted price development. As it can be seen from table 12 and from figures 22-24 above, the average prices are predicted to develop differently depending on housing type. For this reason, the results are presented in three housing type categories further in chapters 4.1-4.3.

To get an extensive data collection from all areas, an assumption was made that different factors affect similarly to old apartments and new-build apartments. From the predicted areas, Espoonlahti has apartment buildings and terrace houses built quite constantly from the 1970s to new-built apartments in the 2010s. In Soukka, most of the apartment buildings and terrace houses are built in the 1970s and the 1980s, which might affect the predicted price development as there might be large renovations such as plumping or facade renovation in the near future. In Kivenlahti, the year-built variates between 1970 to 2010s. However, the majority of the apartment buildings and terrace houses in the area are built in 1970-1990, but as there are also several apartment buildings built in the 2010s, it probably creates large variations for both realized and predicted average prices per square meter in the area. For this reason, the predicted price development is also analyzed based on the decade of the year built, to get better insight into whether there are large variations in how the prices develop, and the results are presented in chapters 4.1-4.3. In addition to this, the predicted price development is also separately presented based on the number of rooms and distance to the metro station.

4.1 Apartment buildings

In this chapter predicted price development for apartment buildings is presented. Results in this chapter are divided into three categories: apartment size based on the number of rooms, distance to metro station, and year built.

Table 13 Predicted price development for apartment buildings by number of rooms

Apartment buildings	**Soukka**			**Espoonlahti**			**Kivenlahti**		
Studio apartments	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	4079,82 €	8,92 %	8,28 %	3843,85 €	1,78 %	1,19 %	4222,64 €	-3,65 %	-4,21 %
2021	3983,54 €	-2,36 %	-3,25 %	4142,37 €	7,77 %	6,79 %	4145,50 €	-1,83 %	-2,72 %
2022	4133,84 €	3,77 %	2,45 %	4301,55 €	3,84 %	2,52 %	4257,28 €	2,70 %	1,39 %
2023	4891,62 €	18,33 %	16,42 %	5152,33 €	19,78 %	17,84 %	5040,15 €	18,39 %	16,48 %
Yearly change from 2019 to 2023		**6,90 %**	**5,73 %**		**8,08 %**	**6,89 %**		**3,56 %**	**2,42 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
One-bedroom	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2904,09 €	1,20 %	0,61 %	3663,06 €	-4,41 %	-4,97 %	3245,79 €	-15,26 %	-15,76 %
2021	2984,29 €	2,76 %	1,83 %	3944,58 €	7,69 %	6,71 %	3398,83 €	4,72 %	3,77 %
2022	3073,98 €	3,01 %	1,69 %	4043,49 €	2,51 %	1,20 %	3452,35 €	1,57 %	0,28 %
2023	3696,50 €	20,25 %	18,31 %	4885,77 €	20,83 %	18,88 %	4151,39 €	20,25 %	18,31 %
Yearly change from 2019 to 2023		**6,54 %**	**5,37 %**		**6,26 %**	**5,10 %**		**2,03 %**	**0,91 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Two-bedrooms	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2493,85 €	9,84 %	9,20 %	2829,88 €	-2,06 %	-2,63 %	2836,99 €	1,11 %	0,51 %
2021	2450,24 €	-1,75 %	-2,64 %	3217,75 €	13,71 %	12,67 %	2982,22 €	5,12 %	4,17 %
2022	2504,94 €	2,23 %	0,93 %	3218,33 €	0,02 %	-1,26 %	3026,08 €	1,47 %	0,18 %
2023	3029,34 €	20,93 %	18,98 %	3936,93 €	22,33 %	20,35 %	3689,53 €	21,92 %	19,96 %
Yearly change from 2019 to 2023		**7,48 %**	**6,30 %**		**8,04 %**	**6,86 %**		**7,08 %**	**5,91 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Three-bedrooms or larger	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2249,07 €	-8,75 %	-9,28 %	2064,96 €	-0,57 %	-1,16 %	2690,29 €	-10,09 %	-10,62 %
2021	2258,70 €	0,43 %	-0,48 %	2391,35 €	15,81 %	14,76 %	2717,53 €	1,01 %	0,10 %
2022	2295,79 €	1,64 %	0,34 %	2406,03 €	0,61 %	-0,67 %	2729,96 €	0,46 %	-0,82 %
2023	2825,33 €	23,07 %	21,08 %	2984,77 €	24,05 %	22,05 %	3345,54 €	22,55 %	20,57 %
Yearly change from 2019 to 2023		**3,47 %**	**2,34 %**		**9,49 %**	**8,29 %**		**2,83 %**	**1,70 %**

In table 13, predicted price development based on the number of rooms is presented. Based on the predictions, there is a significant increase in every apartment size category in average prices in 2023, when is the expected start of operating of Westmetro's phase 2. The largest yearly change from 2019 to 2023 is predicted to be in Espoonlahti in all other apartment sizes but one-bedroom, however also the predicted yearly change for one-bedroom apartments in Espoonlahti is significant. In Espoonlahti the predicted yearly nominal change variates between 6,26-9,49 % and real change 5,10-8,29 %. The smallest increase is predicted to be in Kivenlahti and for one-bedroom apartments, for which the nominal yearly change is predicted to be 2,03 % and real change 0,91 %. A potential reason for the relatively low predicted yearly change in Kivenlahti and for the decrease in average price in 2020 is that there were more new-built apartments sold in 2019 than in previous years, which increased the average prices per square meter in 2019 and can be seen in the predictions as a price decrease as the dataset used for predictions includes apartment buildings with different kind of qualities (year built, size, distance to metro station, etc.).

Table 14 Predicted price development for apartment buildings by distance to metro station

Apartment buildings	**Soukka**			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 0-200m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2879,41 €	0,80 %	0,21 %	5239,62 €	-5,31 %	-5,87 %	-	-	-
2021	2890,77 €	0,39 %	-0,52 %	5525,86 €	5,46 %	4,51 %	-	-	-
2022	2971,48 €	2,79 %	1,48 %	5637,31 €	2,02 %	0,71 %	-	-	-
2023	3468,39 €	16,72 %	14,84 %	6601,82 €	17,11 %	15,22 %	-	-	-
Yearly change from 2019 to 2023		**4,97 %**	**3,82 %**		**4,51 %**	**3,37 %**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 201-400m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2574,32 €	-16,04 %	-16,53 %	4668,39 €	2,23 %	1,63 %	3595,81 €	12,01 %	11,35 %
2021	2764,52 €	7,39 %	6,41 %	4416,04 €	-5,41 %	-6,26 %	3597,27 €	0,04 %	-0,87 %
2022	2846,77 €	2,98 %	1,66 %	4473,73 €	1,31 %	0,01 %	3649,85 €	1,46 %	0,17 %
2023	3415,84 €	19,99 %	18,05 %	5358,29 €	19,77 %	17,84 %	4293,82 €	17,64 %	15,74 %
Yearly change from 2019 to 2023		**2,74 %**	**1,61 %**		**4,08 %**	**2,94 %**		**7,54 %**	**6,36 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 401-600m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2743,09 €	2,21 %	1,61 %	4599,69 €	2,20 %	1,60 %	2969,71 €	-1,21 %	-1,79 %
2021	2841,95 €	3,60 %	2,66 %	4855,28 €	5,56 %	4,60 %	3092,04 €	4,12 %	3,17 %
2022	2920,38 €	2,76 %	1,45 %	4934,09 €	1,62 %	0,33 %	3156,13 €	2,07 %	0,77 %
2023	3538,36 €	21,16 %	19,20 %	5982,39 €	21,25 %	19,29 %	3791,98 €	20,15 %	18,21 %
Yearly change from 2019 to 2023		**7,16 %**	**5,98 %**		**7,37 %**	**6,20 %**		**5,98 %**	**4,82 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 601-800m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2522,69 €	9,38 %	8,74 %	2873,88 €	-3,80 %	-4,37 %	3767,13 €	-13,86 %	-14,37 %
2021	2723,04 €	7,94 %	6,96 %	3086,92 €	7,41 %	6,44 %	3931,73 €	4,37 %	3,42 %
2022	2817,50 €	3,47 %	2,15 %	3152,06 €	2,11 %	0,81 %	3986,75 €	1,40 %	0,11 %
2023	3464,00 €	22,95 %	20,96 %	3858,88 €	22,42 %	20,45 %	4898,32 €	22,86 %	20,88 %
Yearly change from 2019 to 2023		**10,70 %**	**9,49 %**		**6,61 %**	**5,44 %**		**2,87 %**	**1,75 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 801-1000m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2842,64 €	21,91 %	21,19 %	2419,76 €	-10,14 %	-10,67 %	2924,70 €	-2,07 %	-2,65 %
2021	2948,25 €	3,72 %	2,77 %	2492,48 €	3,00 %	2,07 %	2944,21 €	0,67 %	-0,25 %
2022	3026,56 €	2,66 %	1,35 %	2554,68 €	2,50 %	1,19 %	3001,07 €	1,93 %	0,63 %
2023	3735,67 €	23,43 %	21,44 %	3169,17 €	24,05 %	22,05 %	3694,91 €	23,12 %	21,13 %
Yearly change from 2019 to 2023		**12,50 %**	**11,27 %**		**4,16 %**	**3,01 %**		**5,46 %**	**4,31 %**

In table 14 predicted price development is presented in categories based on the distance to the metro station every 200 meters. In Kivenlahti, there were no sold apartments in a 200-meter radius from the metro station in the timeframe of data collection 2009-2019, so there are no data for the predictions in this category either. Interesting observations from table 14 is that there is a high increase in prices predicted for every category and area for 2023, when is the expected start of phase 2's operating. Also, based on table 14 it seems that in every area, the predicted increase from 2022 to 2023 gets a bit higher the further from the metro station it is, and for every area, the predicted change from 2022 to 2023 is at its highest in the 801-1000 meter from the metro station group.

Table 15 Predicted price development for apartment buildings by year built

Apartment buildings	**Soukka**			**Espoonlahti**			**Kivenlahti**		
Year built 1960s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2384,57 €	-15,18 %	-15,68 %	-	-	-	-	-	-
2021	2606,23 €	9,30 %	8,30 %	-	-	-	-	-	-
2022	2646,38 €	1,54 %	0,24 %	-	-	-	-	-	-
2023	3132,62 €	18,37 %	16,46 %	-	-	-	-	-	-
Yearly change from 2019 to 2023		**2,74 %**	**1,61 %**		**-**	**-**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1970s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2653,40 €	1,21 %	0,62 %	2369,84 €	-8,36 %	-8,90 %	2651,29 €	5,94 %	5,32 %
2021	2793,43 €	5,28 %	4,32 %	2514,36 €	6,10 %	5,14 %	2780,36 €	4,87 %	3,92 %
2022	2877,29 €	3,00 %	1,69 %	2567,26 €	2,10 %	0,80 %	2843,06 €	2,25 %	0,95 %
2023	3478,19 €	20,88 %	18,93 %	3161,84 €	23,16 %	21,17 %	3438,59 €	20,95 %	18,99 %
Yearly change from 2019 to 2023		**7,32 %**	**6,15 %**		**5,15 %**	**4,00 %**		**8,27 %**	**7,08 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1980s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3397,23 €	-17,11 %	-17,59 %	2991,98 €	0,63 %	0,04 %	3335,78 €	-4,65 %	-5,21 %
2021	3610,03 €	6,26 %	5,30 %	2965,19 €	-0,90 %	-1,79 %	3500,27 €	4,93 %	3,98 %
2022	3655,41 €	1,26 %	-0,04 %	3043,50 €	2,64 %	1,33 %	3558,17 €	1,65 %	0,36 %
2023	4469,22 €	22,26 %	20,29 %	3741,63 €	22,94 %	20,95 %	4272,00 €	20,06 %	18,12 %
Yearly change from 2019 to 2023		**2,19 %**	**1,07 %**		**5,92 %**	**4,75 %**		**5,12 %**	**3,97 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1990s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3481,26 €	-	-	-	-	-	3970,50 €	-6,26 %	-6,81 %
2021	3513,64 €	0,93 %	0,01 %	-	-	-	4068,34 €	2,46 %	1,53 %
2022	3471,08 €	-1,21 %	-2,47 %	-	-	-	4149,28 €	1,99 %	0,69 %
2023	4258,42 €	22,68 %	20,70 %	-	-	-	4971,35 €	19,81 %	17,88 %
Yearly change from 2019 (2018*) to 2023		**4,60 %***	**3,47 %***		**-**	**-**		**4,09 %**	**2,94 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 2000s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	4588,58 €	16,24 %	15,56 %	4608,11 €	5,86 %	5,23 %	4382,83 €	-7,63 %	-8,17 %
2021	4610,20 €	0,47 %	-0,44 %	4451,04 €	-3,41 %	-4,28 %	4376,35 €	-0,15 %	-1,05 %
2022	4710,15 €	2,17 %	0,86 %	4527,67 €	1,72 %	0,42 %	4406,71 €	0,69 %	-0,59 %
2023	5524,72 €	17,29 %	15,40 %	5501,71 €	21,51 %	19,55 %	5284,26 €	19,91 %	17,98 %
Yearly change from 2019 to 2023		**8,77 %**	**7,57 %**		**6,03 %**	**4,87 %**		**2,73 %**	**1,60 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 2010s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	-	-	-	5220,80 €	-1,61 %	-2,19 %	5832,23 €	1,36 %	0,76 %
2021	-	-	-	5425,61 €	3,92 %	2,98 %	5934,28 €	1,75 %	0,83 %
2022	-	-	-	5530,27 €	1,93 %	0,63 %	6016,47 €	1,39 %	0,09 %
2023	-	-	-	6550,65 €	18,45 %	16,54 %	7239,81 €	20,33 %	18,39 %
Yearly change from 2019 to 2023		**-**	**-**		**5,41 %**	**4,25 %**		**5,91 %**	**4,75 %**

In table 15, predicted price development for apartment buildings based on the year built is presented by areas. Differences between the areas can be seen from table 15. Soukka is the only area from the three observed areas that have real estate which is built in the 1960s, however, Soukka is also the only area that does not have apartment buildings built in the 2010s. Based on the results presented in table 15, there are significant yearly increases predicted in 2023 when Westmetro's phase 2 starts operating but compared to table 14, there seems to be more variation in the predicted increase between the areas in the same category. For example, when considering category year built in the 2000s, the predicted nominal increase for 2023 in Soukka is 17,29 % but for Espoonlahti it is 21,51 %. However, when observing results based on year built, other qualities of the apartment are not analyzed separately such as the number of rooms, so when comparing the previous year's average prices to the following year, one reason for yearly variation within the same year built

category might be the differences of average prices per square meter depending on the number of rooms.

As it was mentioned in chapter 2.1, plumping renovations should be done every 40-60 years and it might affect the price per square meter 850 euros. Based on this information, plumping renovations might take a place in real estate which are built in the 1960s-1980s. This could explain variation in predicted price development for those categories as there is no information in the collected dataset whether the renovation has been done or not, it might affect the predicted price development, as in part of the training set data observations with year built from the 1960s-1980s might have a higher price due to done renovations and part of them might have a lower price per square meter due to future renovation cost.

4.2 Terrace houses

In this chapter predicted price development for terrace house apartments is presented. Results in this chapter are also divided into three categories: apartment size based on the number of rooms, distance to metro station, and year built.

Table 16 Predicted price development for terrace houses by number of rooms

Terrace houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
One-bedroom	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3939,08 €	-3,31 %	-3,88 %	-	-	-	-	-	-
2021	3770,10 €	-4,29 %	-4,85 %	-	-	-	-	-	-
2022	3833,12 €	1,67 %	1,07 %	-	-	-	-	-	-
2023	4766,69 €	24,36 %	23,63 %	-	-	-	-	-	-
Yearly change from 2019 to 2023		**4,00 %**	**2,86 %**		**-**	**-**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Two-bedrooms	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3300,86 €	3,36 %	2,75 %	3100,77 €	28,89 %	28,13 %	3442,57 €	-7,37 %	-7,92 %
2021	3318,17 €	0,52 %	-0,39 %	2992,98 €	-3,48 %	-4,35 %	3576,33 €	3,89 %	2,94 %
2022	3358,38 €	1,21 %	-0,08 %	3035,76 €	1,43 %	0,13 %	3555,03 €	-0,60 %	-1,86 %
2023	4218,26 €	25,60 %	23,58 %	3792,01 €	24,91 %	22,89 %	4335,09 €	21,94 %	19,97 %
Yearly change from 2019 to 2023		**7,20 %**	**6,03 %**		**12,05 %**	**10,82 %**		**3,92 %**	**2,78 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Three-bedrooms or larger	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2890,33 €	-5,06 %	-5,61 %	2569,20 €	-	-	2866,33 €	-18,82 %	-19,30 %
2021	2963,17 €	2,52 %	1,59 %	2675,60 €	4,14 %	3,20 %	3070,54 €	7,12 %	6,15 %
2022	2975,36 €	0,41 %	-0,87 %	2684,10 €	0,32 %	-0,96 %	3010,95 €	-1,94 %	-3,19 %
2023	3734,91 €	25,53 %	23,50 %	3356,02 €	25,03 %	23,01 %	3694,72 €	22,71 %	20,73 %
Yearly change from 2019 to 2023		**5,24 %**	**4,09 %**		**4,65 %**	**3,51 %**		**1,14 %**	**0,03 %**

Predicted price development for different size terrace house apartments is presented in table 16. There are no studio apartments when considering only terrace houses. When observing all the three areas, predicted price development is largest in the two-bedroom category.

However, when considering areas separately, the largest increase in housing prices for terrace houses is in Espoonlahti. For all three areas for all apartment sizes, there is a large increase predicted for 2023 when the metro is expected to start operating.

Table 17 Predicted price development for terrace houses by distance to metro station

Terrace houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 201-400m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2712,80 €	-	-	-	-	-	-	-	-
2021	2675,30 €	-1,38 %	-2,28 %	-	-	-	-	-	-
2022	2676,29 €	0,04 %	-0,87 %	-	-	-	-	-	-
2023	3285,79 €	22,77 %	21,66 %	-	-	-	-	-	-
Yearly change from 2019 to 2023		**-**	**-**		**-**	**-**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 401-600m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2645,92 €	-	-	-	-	-	2898,82 €	-20,29 %	-21,01 %
2021	2560,65 €	-3,22 %	-4,10 %	-	-	-	3223,71 €	11,21 %	10,20 %
2022	2559,99 €	-0,03 %	-1,30 %	-	-	-	3239,69 €	0,50 %	-0,79 %
2023	3179,90 €	24,22 %	22,21 %	-	-	-	3946,83 €	21,83 %	19,86 %
Yearly change from 2019 (2018*) to 2023		**8,67 %***	**6,31 %***		**-**	**-**		**2,07 %**	**0,95 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 601-800m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3140,20 €	-8,07 %	-8,90 %	-	-	-	3250,73 €	-	-
2021	3204,70 €	2,05 %	1,13 %	-	-	-	3284,79 €	1,05 %	0,13 %
2022	3264,40 €	1,86 %	0,94 %	-	-	-	3248,18 €	-1,11 %	-2,01 %
2023	4095,25 €	25,45 %	24,31 %	-	-	-	4024,24 €	23,89 %	22,77 %
Yearly change from 2019 (2018*) to 2023		**3,33 %**	**2,20 %**		**-**	**-**		**-0,18 %***	**-1,23 %***
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 801-1000m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3033,07 €	-1,27 %	-1,85 %	2701,80 €	12,31 %	11,65 %	3134,07 €	-6,90 %	-7,45 %
2021	3206,22 €	5,71 %	4,75 %	2760,27 €	2,16 %	1,24 %	3127,39 €	-0,21 %	-1,12 %
2022	3233,26 €	0,84 %	-0,44 %	2779,70 €	0,70 %	-0,58 %	3099,45 €	-0,89 %	-2,16 %
2023	4064,38 €	25,71 %	23,68 %	3475,21 €	25,02 %	23,00 %	3848,90 €	24,18 %	22,18 %
Yearly change from 2019 to 2023		**7,25 %**	**6,07 %**		**9,63 %**	**8,43 %**		**3,40 %**	**2,27 %**

There are no terrace houses in a 0–200-meter radius from the metro station and Soukka is the only area that has terrace houses in a 201–400-meter radius from the metro station. Soukka did not have sold terrace house apartments in the 401-600m category in 2019 and Kivenlahti did not have sold terrace house apartments in 601-800m category in 2019, so 2018 average prices were used instead to calculate the yearly change from 2018 to 2023 and they are presented in red in table 17. There were no terrace house apartments sold in 2019 or 2018 in the 201-400m category, which could have been used to calculate the yearly change in Soukka from 2019 (2018*) to 2023. As it was mentioned earlier in chapter 4, it was decided that if there were observations only from the time period 2009-2017, they are not used to calculate the yearly changes. Espoonlahti only has terrace houses in 801–1000-meter radius from the metro station. Despite the distance from the metro station, it seems that there is a large increase in the predicted yearly changes in 2023 for all three areas in the 801–1000-meter category.

Table 18 Predicted price development for terrace houses by year built

Terrace houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
Year built 1970s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2735,79 €	5,77 %	5,15 %	2551,58 €	15,98 %	15,30 %	2731,80 €	-	-
2021	2872,16 €	4,98 %	4,03 %	2605,64 €	2,12 %	1,19 %	2733,29 €	0,05 %	-0,85 %
2022	2815,40 €	-1,98 %	-3,23 %	2627,65 €	0,84 %	-0,44 %	2729,99 €	-0,12 %	-1,40 %
2023	3532,86 €	25,48 %	23,46 %	3287,56 €	25,11 %	23,09 %	3337,80 €	22,26 %	20,29 %
Yearly average change from 2019 to 2023		**8,11 %**	**6,92 %**		**10,56 %**	**9,35 %**		**1,84 %**	**0,74 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1980s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3176,13 €	-5,46 %	-6,01 %	2702,47 €	3,48 %	2,88 %	3035,07 €	-14,04 %	-14,54 %
2021	3257,41 €	2,56 %	1,63 %	2819,50 €	4,33 %	3,38 %	3129,96 €	3,13 %	2,19 %
2022	3289,19 €	0,98 %	-0,31 %	2831,50 €	0,43 %	-0,86 %	3130,46 €	0,02 %	-1,26 %
2023	4123,63 €	25,37 %	23,34 %	3529,11 €	24,64 %	22,62 %	3882,57 €	24,03 %	22,02 %
Yearly average change from 2019 to 2023		**5,26 %**	**4,10 %**		**7,82 %**	**6,64 %**		**2,40 %**	**1,28 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1990s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2945,66 €	-	-	3172,31 €	-	-	3702,30 €	-0,38 %	-0,97 %
2021	3096,32 €	5,11 %	4,50 %	3096,12 €	-2,40 %	-2,97 %	3619,63 €	-2,23 %	-2,81 %
2022	3075,11 €	-0,69 %	-1,27 %	3122,27 €	0,84 %	0,25 %	3631,74 €	0,33 %	-0,25 %
2023	3870,32 €	25,86 %	25,12 %	3902,80 €	25,00 %	24,27 %	4432,50 €	22,05 %	21,33 %
Yearly average change from 2019 (2018*) to 2023		**2,53 %***	**1,42 %***		**-**	**-**		**4,50 %**	**3,36 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 2000s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2717,24 €	0,25 %	-0,34 %	-	-	-	3541,96 €	-	-
2021	2936,99 €	8,09 %	7,45 %	-	-	-	3526,35 €	-0,44 %	-1,03 %
2022	2942,83 €	0,20 %	-0,39 %	-	-	-	3483,13 €	-1,23 %	-1,81 %
2023	3702,45 €	25,81 %	25,07 %	-	-	-	4338,65 €	24,56 %	23,83 %
Yearly average change from 2019 to 2023		**8,11 %**	**6,92 %**		**-**	**-**		**-**	**-**

In table 18 predicted price development of terrace houses based on the year built is presented. There are no terrace houses built in any of the three areas in the 1960s or the 2010s. In every year built category for each area, there is a significant increase predicted for 2023 when Westmetro's phase 2 is expected to start operating.

4.3 Houses

This chapter presents the predicted price development for houses. Similarly, as in chapters 4.1 and 4.2, for apartment buildings and terrace houses, results are divided into three categories: apartment size based on the number of rooms, distance to metro station, and year built. As presented in table 3 in chapter 3.1, only 4,09 % of the observations are houses. When the observation time is relatively long 2009-2019 and the observations are from three different areas, it means that there were not many observations per quarter in the data, and there might have been several quarters when there were no houses sold. The lack of observations might cause mispredictions for the predicted price development of houses. However, as mentioned in chapter 3.2.2, because the accuracy of the all house types -model was better than apartment buildings and terrace houses model's, it was decided to use the all house types model and to create the predictions for houses as well.

Table 19 Predicted price development for houses by number of rooms

Houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
One-bedroom	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3201,38 €	-11,86 %	-12,38 %	-	-	-	-	-	-
2021	3246,03 €	1,39 %	0,47 %	-	-	-	-	-	-
2022	3200,88 €	-1,39 %	-2,65 %	-	-	-	-	-	-
2023	2955,60 €	-7,66 %	-9,15 %	-	-	-	-	-	-
Yearly change from 2019 to 2023		**-5,02 %**	**-6,07 %**		-	-		-	-
	Soukka			**Espoonlahti**			**Kivenlahti**		
Two-bedrooms	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2331,53 €	-	-	3799,89 €	-	-	-	-	-
2021	2299,00 €	-1,40 %	-2,29 %	3737,02 €	-1,65 %	-2,55 %	-	-	-
2022	2253,47 €	-1,98 %	-2,87 %	3718,21 €	-0,50 %	-1,41 %	-	-	-
2023	2115,04 €	-6,14 %	-6,99 %	3347,57 €	-9,97 %	-10,78 %	-	-	-
Yearly change from 2019 (2018*) to 2023		**-12,77 %***	**-13,71 %***		**-1,11 %***	**-2,18 %***		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Three-bedrooms or larger	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3222,59 €	0,78 %	0,19 %	2621,55 €	41,16 %	40,33 %	2989,02 €	-	-
2021	2996,68 €	-7,01 %	-7,85 %	2775,76 €	5,88 %	4,92 %	2731,87 €	-8,60 %	-9,43 %
2022	2886,33 €	-3,68 %	-4,91 %	2607,63 €	-6,06 %	-7,26 %	2558,93 €	-6,33 %	-7,53 %
2023	2675,03 €	-7,32 %	-8,82 %	2550,32 €	-2,20 %	-3,78 %	2559,50 €	0,02 %	-1,59 %
Yearly change from 2019 (2018*) to 2023		**-4,36 %**	**-5,41 %**		**8,25 %**	**7,06 %**		**-0,45 %***	**-1,53 %***

In table 19, predicted price development for houses is presented based on the number of rooms. Soukka is the only area to have one-bedroom houses. Some categories that did not have sold houses in 2019, so 2018 average prices were used instead to calculate the yearly change from 2018 to 2023 and they are presented in red in table 19. An interesting observation from table 19 is that, even though the yearly changes in three-bedroom or larger houses in Kivenlahti are negative and relatively large per year, the average yearly change from 2018 to 2023 is not as large as the yearly changes in 2021-2023. When considering all three areas, price development is predicted to be significantly higher in Espoonlahti than in Soukka or Kivenlahti.

Table 20 Predicted price development for houses by distance to metro station

Houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 401-600m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2787,92 €	-	-	-	-	-	-	-	-
2021	2753,85 €	-1,22 %	-1,80 %	-	-	-	-	-	-
2022	2688,32 €	-2,38 %	-2,95 %	-	-	-	-	-	-
2023	2426,98 €	-9,72 %	-10,25 %	-	-	-	-	-	-
Yearly change from 2019 to 2023		**-**	**-**		**-**	**-**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 601-800m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3184,06 €	6,52 %	5,89 %	-	-	-	-	-	-
2021	2858,19 €	-10,23 %	-10,76 %	-	-	-	-	-	-
2022	2768,82 €	-3,13 %	-3,70 %	-	-	-	-	-	-
2023	2555,44 €	-7,71 %	-8,25 %	-	-	-	-	-	-
Yearly change from 2019 to 2023		**-3,84 %**	**-4,90 %**		**-**	**-**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Distance to metro station 801-1000m	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3112,17 €	-5,51 %	-6,72 %	2916,13 €	57,02 %	55,02 %	2989,02 €	-	-
2021	3019,37 €	-2,98 %	-3,55 %	2868,75 €	-1,62 %	-2,20 %	2731,87 €	-8,60 %	-9,14 %
2022	2922,53 €	-3,21 %	-3,78 %	2742,35 €	-4,41 %	-4,97 %	2558,93 €	-6,33 %	-6,88 %
2023	2713,35 €	-7,16 %	-8,00 %	2631,89 €	-4,03 %	-4,90 %	2559,50 €	0,02 %	-0,89 %
Yearly change from 2019 (2018*) to 2023		**-4,73 %**	**-5,78 %**		**9,10 %**	**7,91 %**		**-0,45 %***	**-1,53 %***

Predicted price development for houses based on the distance to the metro station is presented in table 20. There are no houses in any of the three areas in a 0–400-meter radius from the metro stations and Soukka is the only area to have houses in a 401–800-meter radius from the metro station. Espoonlahti and Kivenlahti have house only in 801-1000m category. Similarly, as in table 19, also based on the prediction results presented in table 20, Espoonlahti is the only area to have positive yearly change. It seems that the predicted change in every area for 2023 is negative. The overall change for the prediction time frame is negative in most of the areas, so it seems that the start of the metro's operating does not alone affect negatively the price development of the houses.

Table 21 Predicted price development for houses by year built

Houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
Year built 1960s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3185,73 €	-	-	-	-	-	-	-	-
2021	3151,05 €	-1,09 %	-1,99 %	-	-	-	-	-	-
2022	3071,93 €	-2,51 %	-3,76 %	-	-	-	-	-	-
2023	2769,25 €	-9,85 %	-11,31 %	-	-	-	-	-	-
Yearly change from 2019 to 2023		**-**	**-**		**-**	**-**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1970s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2461,33 €	2,41 %	1,81 %	-	-	-	2185,95 €	-	-
2021	2427,71 €	-1,37 %	-2,26 %	-	-	-	2117,12 €	-3,15 %	-4,03 %
2022	2347,89 €	-3,29 %	-4,52 %	-	-	-	2065,47 €	-2,44 %	-3,68 %
2023	2140,34 €	-8,84 %	-10,31 %	-	-	-	1881,67 €	-8,90 %	-10,37 %
Yearly change from 2019 to 2023		**-2,86 %**	**-3,92 %**		**-**	**-**		**-**	**-**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1980s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2871,83 €	-9,89 %	-10,42 %	2869,30 €	18,47 %	17,78 %	3251,43 €	-	-
2021	3013,26 €	4,92 %	3,97 %	2934,74 €	2,28 %	1,35 %	3207,62 €	-1,35 %	-2,24 %
2022	2904,22 €	-3,62 %	-4,85 %	2672,96 €	-8,92 %	-10,08 %	3086,18 €	-3,79 %	-5,01 %
2023	2711,12 €	-6,65 %	-8,16 %	2509,80 €	-6,10 %	-7,62 %	2847,04 €	-7,75 %	-9,24 %
Yearly change from 2019 (2018*) to 2023		**-3,96 %**	**-5,02 %**		**0,90 %**	**-0,21 %**		**1,69 %***	**0,59 %***
	Soukka			**Espoonlahti**			**Kivenlahti**		
Year built 1990s	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2548,39 €	-	-	3056,62 €	-	-	-	-	-
2021	2727,24 €	7,02 %	6,05 %	3017,92 €	-1,27 %	-2,16 %	-	-	-
2022	2631,66 €	-3,50 %	-4,74 %	2950,54 €	-2,23 %	-3,48 %	-	-	-
2023	2456,39 €	-6,66 %	-8,17 %	2728,93 €	-7,51 %	-9,00 %	-	-	-
Yearly change from 2019 (2018*) to 2023		**-3,84 %***	**-4,88 %***		**-**	**-**		**-**	**-**
		Soukka			**Espoonlahti**			**Kivenlahti**	
Year built 2000s		Nominal change	Real change		Nominal change	Real change		Nominal change	Real change
2020	3061,35 €	-11,37 %	-12,80 %	-	-	-	2726,36 €	-	-
2021	3026,57 €	-1,14 %	-2,73 %	-	-	-	2656,66 €	-2,56 %	-4,13 %
2022	2934,97 €	-3,03 %	-4,59 %	-	-	-	2548,08 €	-4,09 %	-5,64 %
2023	2721,98 €	-7,26 %	-8,75 %	-	-	-	2448,39 €	-3,91 %	-5,46 %
Yearly change from 2019 to 2023		**-5,78 %**	**-6,81 %**		**-**	**-**		**-**	**-**
		Soukka			**Espoonlahti**			**Kivenlahti**	
Year built 2010s		Nominal change	Real change		Nominal change	Real change		Nominal change	Real change
2020	3367,91 €	-	-	-	-	-	-	-	-
2021	3278,21 €	-2,66 %	-3,55 %	-	-	-	-	-	-
2022	3202,55 €	-2,31 %	-3,55 %	-	-	-	-	-	-
2023	2903,18 €	-9,35 %	-10,81 %	-	-	-	-	-	-
Yearly change from 2019 (2018*) to 2023		**-4,57 %***	**-5,61 %***		**-**	**-**		**-**	**-**

Table 21 presents the predicted price development of houses based on the decade of year built. An interesting observation from table 21 is that even though all the yearly predictions

in Kivenlahti in the year built in the 1980s -category is negative, the yearly average change from 2018 to 2023 is still positive. Despite the predicted decreasing price development for 2021-2023, the predicted average prices in 2023 are still higher than the average price in 2018 was, which creates a positive yearly change from 2018 to 2023. In opposite to tables 15 and 18, where there were no apartment buildings or terrace houses built in the 2010s in Soukka, there are houses built in that decade in Soukka. Soukka is also the only one of the three areas that have houses built every decade from the 1960s to 2010s when the area is limited to a 1-kilometer radius from the metro station. However, it seems that the price development for houses in Soukka is decreasing.

When considering tables 19-21, it is interesting that not in any category there is a predicted increase for 2023, when the metro is expected to start operating. It is the opposite to results of apartment buildings and terrace houses, presented in chapters 4.1 and 4.2, respectively. Also, the overall development of prices for houses is decreasing in most of the categories presented above. So, it seems that the start of the metro's operating does not alone affect decreasingly as the overall price development is predicted to be decreasing. However, as it was mentioned earlier in this chapter, the lack of observations might cause mispredictions for the predicted price development of houses. The small number of observations of this house type category causes large variations in average prices per square meter during the observation time frame 2009-2019, which is used to conduct the model for prediction. The small number of observations was also divided randomly into training and test sets, which decreased the number of house observations used to conduct the model. Due to the low number of observations per quarter, average prices might vary a lot depending on specific sold real estate's qualities.

4.4 Predictions by using information before COVID-19

Predictions for 2020-2023 were also created by using a prediction dataset that included the Bank of Finland's predictions from December 2019 for macroeconomic variables. Those Bank of Finland's predictions were created right before the global COVID-19 pandemic. There were differences between the Bank of Finland's December 2019 and December 2020 predictions. In December 2019 predictions inflation was predicted to increase faster, the employment rate would grow throughout the prediction period and Euribor 3 months would remain low but start to increase. In December 2020 predictions, when the COVID-19

pandemic was going on, inflation was predicted to be slower, the employment rate would have a temporary drop and achieve 2019 level in 2022 and Euribor 3 months would decrease.

Table 22 Predicted price development by using Bank of Finland's predictions before COVID-19

All house types	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2823,59 €	-0,15 %	-1,91 %	3110,88 €	-3,16 %	-4,86 %	3240,73 €	-7,10 %	-8,73 %
2021	2915,68 €	3,26 %	1,83 %	3530,23 €	13,48 %	11,91 %	3362,83 €	3,77 %	2,33 %
2022	2946,65 €	1,06 %	-0,50 %	3587,17 €	1,61 %	0,04 %	3393,35 €	0,91 %	-0,65 %
2023	3291,20 €	11,69 %	9,79 %	4074,04 €	13,57 %	11,64 %	3814,86 €	12,42 %	10,51 %
Yearly change from 2019 to 2023		**3,87 %**	**2,21 %**		**6,12 %**	**4,42 %**		**2,26 %**	**0,63 %**
Apartment buildings	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2797,49 €	3,81 %	1,98 %	3123,80 €	-4,14 %	-5,83 %	3272,48 €	-5,94 %	-7,60 %
2021	2898,79 €	3,62 %	2,19 %	3568,30 €	14,23 %	12,65 %	3390,32 €	3,60 %	2,17 %
2022	2947,07 €	1,67 %	0,09 %	3629,72 €	1,72 %	0,15 %	3426,28 €	1,06 %	-0,50 %
2023	3323,56 €	12,78 %	10,86 %	4135,47 €	13,93 %	11,99 %	3863,54 €	12,76 %	10,84 %
Yearly change from 2019 to 2023		**5,38 %**	**3,70 %**		**6,14 %**	**4,44 %**		**2,65 %**	**1,01 %**
Terrace houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2985,53 €	-4,35 %	-6,03 %	2693,10 €	11,95 %	9,97 %	2992,72 €	-16,36 %	-17,84 %
2021	3116,04 €	4,37 %	2,93 %	2755,23 €	2,31 %	0,89 %	3175,53 €	6,11 %	4,64 %
2022	3099,57 €	-0,53 %	-2,07 %	2770,58 €	0,56 %	-1,00 %	3156,43 €	-0,60 %	-2,14 %
2023	3658,86 €	18,04 %	16,03 %	3275,15 €	18,21 %	16,20 %	3644,87 €	15,47 %	13,51 %
Yearly change from 2019 to 2023		**4,05 %**	**2,39 %**		**8,02 %**	**6,29 %**		**0,46 %**	**-1,14 %**
Houses	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All apartment sizes	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2978,19 €	-7,27 %	-8,91 %	2804,87 €	51,03 %	48,37 %	2847,16 €	-	-
2021	2794,42 €	-6,17 %	-7,47 %	2724,31 €	-2,87 %	-4,22 %	2524,45 €	-11,33 %	-12,56 %
2022	2655,90 €	-4,96 %	-6,43 %	2578,24 €	-5,36 %	-6,83 %	2333,74 €	-7,55 %	-8,99 %
2023	2301,80 €	-13,33 %	-14,81 %	2343,22 €	-9,12 %	-10,66 %	2219,95 €	-4,88 %	-6,49 %
Yearly change from 2019 (2018*) to 2023		**-7,99 %**	**-9,69 %**		**5,98 %**	**4,02 %**		**-3,25 %***	**-4,68 %***

In table 22, predictions for 2020-2023 that were created by using the Bank of Finland's December 2019 predictions in the prediction dataset are presented. The most notable change in results predicted in table 22, compared to results in table 12, where the values from December 2020's predictions were used, is that predicted changes for the year 2023 are significantly different. For all house types, apartment buildings, and terrace houses the increase for the year 2023 is significantly lower in table 22 than in table 12. For houses, the predicted decreases for 2023 are larger than in table 12. As mentioned above, in December 2019 predictions Euribor was predicted to start increase and inflation to grow faster. However, the employment rate is predicted to be higher in December 2019 predictions, but alone its increasing effects are not enough to change the development of these predictions. The yearly changes are predicted to be smaller when using December 2019's predictions for macroeconomic variables.

4.5 Predictions by excluding the information about Westmetro

Predictions for 2020-2023 were also created by conducting a quadratic stepwise regression model by using a dataset, that does not have the stage of metro variables. This means that even though the metro is already under construction during the timeframe 2009-2019 and it might affect price behavior, there is no variable in the used data that indicates the stages of the metro project. Also, because there is no stage of metro variables, there is no "stage of metro: operating" variable in the predictions, so the predictions are created by using the same variables as there is in the conducted regression model, so there is no need to include any information from phase 1 data to this prediction model.

Table 23 Predicted price development by using the dataset without "stage of metro" -variables

All apartment sizes	**Soukka**			**Espoonlahti**			**Kivenlahti**		
All house types	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2724,35 €	-3,66 %	-4,22 %	3039,24 €	-5,39 %	-5,94 %	3175,72 €	-8,96 %	-9,49 %
2021	2805,60 €	2,98 %	2,05 %	3427,89 €	12,79 %	11,76 %	3289,73 €	3,59 %	2,65 %
2022	2841,27 €	1,27 %	-0,02 %	3483,45 €	1,62 %	0,32 %	3329,15 €	1,20 %	-0,09 %
2023	2829,25 €	-0,42 %	-2,03 %	3503,28 €	0,57 %	-1,05 %	3322,27 €	-0,21 %	-1,82 %
Yearly change from 2019 to 2023		**0,01 %**	**-1,08 %**		**2,19 %**	**1,07 %**		**-1,12 %**	**-2,30 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Apartment buildings	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2670,05 €	-0,92 %	-1,50 %	3046,78 €	-6,51 %	-7,05 %	3194,95 €	-8,17 %	-8,71 %
2021	2750,92 €	3,03 %	2,09 %	3456,19 €	13,44 %	12,41 %	3305,41 €	3,46 %	2,52 %
2022	2802,92 €	1,89 %	0,59 %	3515,60 €	1,72 %	0,42 %	3350,66 €	1,37 %	0,08 %
2023	2807,65 €	0,17 %	-1,45 %	3538,13 €	0,64 %	-0,98 %	3350,66 €	0,00 %	-1,61 %
Yearly change from 2019 to 2023		**1,03 %**	**0,08 %**		**2,08 %**	**0,96 %**		**-0,94 %**	**-2,02 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Terrace houses	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	2921,92 €	-6,39 %	-6,94 %	2659,61 €	10,55 %	9,90 %	2984,82 €	-16,58 %	-17,07 %
2021	2999,85 €	2,67 %	1,74 %	2704,24 €	1,68 %	0,76 %	3160,97 €	5,90 %	4,94 %
2022	2968,14 €	-1,06 %	-2,32 %	2705,75 €	0,06 %	-1,22 %	3134,33 €	-0,84 %	-2,11 %
2023	2913,32 €	-1,85 %	-3,43 %	2681,76 €	-0,89 %	-2,49 %	3051,74 €	-2,63 %	-4,21 %
Yearly change from 2019 to 2023		**-1,71 %**	**-2,79 %**		**2,75 %**	**1,63 %**		**-3,90 %**	**-4,96 %**
	Soukka			**Espoonlahti**			**Kivenlahti**		
Houses	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change	Ppsm	Nominal change	Real change
2020	3165,55 €	-1,44 %	-2,02 %	3075,30 €	65,59 %	64,62 %	3171,35 €	-	-
2021	3092,03 €	-2,32 %	-3,21 %	3049,80 €	-0,83 %	-1,73 %	2907,56 €	-8,32 %	-9,15 %
2022	3016,34 €	-2,45 %	-3,69 %	2965,42 €	-2,77 %	-4,01 %	2771,45 €	-4,68 %	-5,90 %
2023	2894,51 €	-4,04 %	-5,59 %	3014,93 €	1,67 %	0,03 %	2874,86 €	3,73 %	2,06 %
Yearly change from 2019 (2018*) to 2023		**-2,57 %**	**-3,64 %**		**12,88 %**	**11,64 %**		**1,89 %***	**0,79 %***

In table 23, when considering all house types, apartment buildings, and terrace houses, the yearly average changes from 2019 to 2023 are significantly lower, compared to predictions that include the stage of metro variables and are presented in table 12. An interesting observation is that for houses, the average yearly change from 2019 (2018*) to 2023 is predicted to be higher when the stage of metro variables are not included in the prediction. For Espoonlahti and Kivenlahti the average yearly changes for houses are significantly higher and for Soukka, the predicted yearly change is still decreasing but the decrease is

not as large as in table 12. For years 2020-2022, there are some differences in the predictions presented in tables 12 and 23, but they are relatively small. The most significant, and large differences are predicted for 2023, when is the expected start for Westmetro's phase 2 to start operating. In table 23, where the stage of metro variables are not included in the regression model, there is only a small increase or decrease predicted for 2023 and as an opposite, in table 12, where the stage of metro variables are included in the regression model, there is relatively large increase predicted at housing prices for 2023, except for houses.

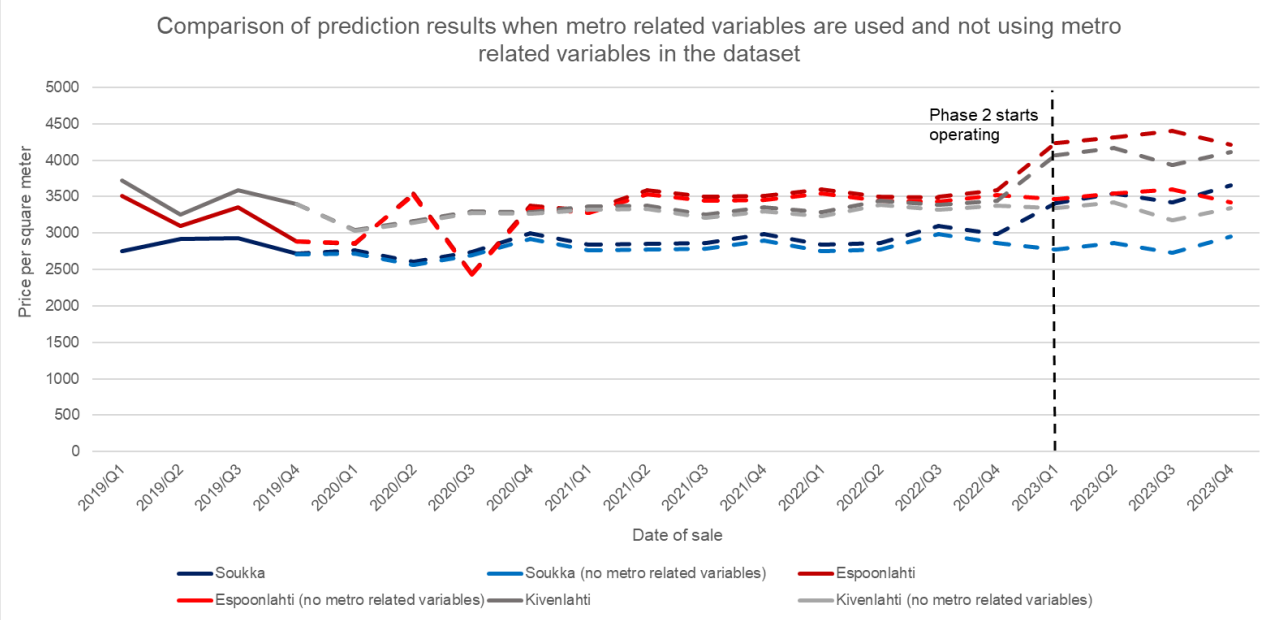

Figure 25 Comparison of predictions results: metro related variables included and excluded in the dataset

In figure 25 comparison of prediction results when stages of metro variables are included in the model and when they are not, is presented. The regression model, that is used for the actual predictions in this study, includes the stage of metro variables. A comparative regression model is created, otherwise identical dataset, but the stage of metro variables are removed from the dataset. Predictions are presented by the dash lines on the graph in figure 25. It can be seen from the graph, that the prediction results for years 2020-2022 are similar by using both models, the one including the stages of metro variables and the one not using them. The most notable difference during 2020-2022 is in Soukka, where the small difference can be seen as the model that uses the metro-related variables gives a bit higher values for prediction. However, in 2023, the difference is significant for every area between the two models. As mentioned earlier, the expected start of Westmetro's phase 2 to start

operating is in 2023. This can be seen in the prediction results of the two different models as the model that includes the stage of metro variables, gives 600-800 euros higher predictions for 2023 than the model that does not include the stage of metro variables, when considering all house types.

5 DISCUSSION AND CONCLUSIONS

In this chapter main findings and conclusions of the study are presented, also results are reflected in previous studies. The reliability and limitations of the study are evaluated and finally, recommendations for further research are given.

5.1 Main findings and conclusions

The aim of this study was to predict the price development for Westmetro's phase 2 areas in the years 2020-2023. The main findings of the prediction results are presented in this chapter.

Table 24 Predicted yearly changes for apartment buildings

Apartment buildings	**Soukka**		**Espoonlahti**		**Kivenlahti**	
Yearly average change from 2019 (2018*) to 2023	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
Studio apartments	6,90 %	5,73 %	8,08 %	6,89 %	3,56 %	2,42 %
One-bedroom	6,54 %	5,37 %	6,26 %	5,10 %	2,03 %	0,91 %
Two-bedrooms	7,48 %	6,30 %	8,04 %	6,86 %	7,08 %	5,91 %
Three-bedrooms or larger	3,47 %	2,34 %	9,49 %	8,29 %	2,83 %	1,70 %
Distance to metro station 0-200m	4,97 %	3,82 %	4,51 %	3,37 %	-	-
Distance to metro station 201-400m	2,74 %	1,61 %	4,08 %	2,94 %	7,54 %	6,36 %
Distance to metro station 401-600m	7,16 %	5,98 %	7,37 %	6,20 %	5,98 %	4,82 %
Distance to metro station 601-800m	10,70 %	9,49 %	6,61 %	5,44 %	2,87 %	1,75 %
Distance to metro station 801-1000m	12,50 %	11,27 %	4,16 %	3,01 %	5,46 %	4,31 %
Year built 1960s	2,74 %	1,61 %	-	-	-	-
Year built 1970s	7,32 %	6,15 %	5,15 %	4,00 %	8,27 %	7,08 %
Year built 1980s	2,19 %	1,07 %	5,92 %	4,75 %	5,12 %	3,97 %
Year built 1990s	4,60 %*	3,47 %*	-	-	4,09 %	2,94 %
Year built 2000s	1,59 %	0,48 %	6,03 %	4,87 %	1,98 %	0,86 %
Year built 2010s	-	-	5,41 %	4,25 %	5,91 %	4,75 %

In table 24, the predicted yearly changes from 2019 (2018*) to 2023 of apartment buildings are presented. Overall, the development of apartment buildings' prices is predicted to be increasing. The price development variates in every area, for example in the distance to metro station categories. There is no category that would exceed all the others in every three areas. Environmental qualities affect the development as well, for example, the seaside. For example, in Soukka in 601-800m and 801-1000m distance to metro station categories, the predicted price development is significantly larger compared to other categories of the same area. One possible explanation for this could be the proximity of the seaside in part of the 1 km radius fringe area. As mentioned in chapter 2.1, if the sold apartment has a seaside view, it might increase the price per square meter even by 2000 euros. This kind of difference

between the observation apartments might create significant changes in average price development. However, there is no way to get the information about the seaside view from the collected data, which creates variation in prices between otherwise similar apartments.

As mentioned in chapter 2.1, plumping renovations should be done every 40-60 years and it might affect the price per square meter even 850 euros. Based on this information, plumping renovations might take a place in real estate which are built in the 1960s-1980s. Because there is no way to find out whether the renovation has been done or not, it might affect the predicted price development. Soukka is the only area to have apartment buildings with year built 1960s and the predicted yearly nominal change for that category is 2,74 %. This could be considered relatively low compared to other categories of Soukka. An interesting observation from table 24 is that the lowest price increase is predicted to be in Soukka and Kivenlahti for the year built 2000s category, as both areas have less than 2 % predicted yearly change.

Table 25 Predicted yearly changes for terrace houses

Terrace houses	**Soukka**		**Espoonlahti**		**Kivenlahti**	
Yearly average change from 2019 (2018*) to 2023	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
One-bedroom	4,00 %	2,86 %	-	-	-	-
Two-bedrooms	7,20 %	6,03 %	12,05 %	10,82 %	3,92 %	2,78 %
Three-bedrooms or larger	5,24 %	4,09 %	4,65 %	3,51 %	1,14 %	0,03 %
Distance to metro station 401-600m	8,67 %*	6,31 %*	-	-	2,07 %	0,95 %
Distance to metro station 601-800m	3,33 %	2,20 %	-	-	-0,18 %*	-1,23 %*
Distance to metro station 801-1000m	7,25 %	6,07 %	9,63 %	8,43 %	3,40 %	2,27 %
Year built 1970s	8,11 %	6,92 %	10,56 %	9,35 %	1,84 %	0,74 %
Year built 1980s	5,26 %	4,10 %	7,82 %	6,64 %	2,40 %	1,28 %
Year built 1990s	2,53 %*	1,42 %*	-	-	4,50 %	3,36 %
Year built 2000s	8,11 %	6,92 %	-	-	-	-

The predicted yearly changes from 2019 (2018*) to 2023 of terrace houses are presented in table 25. The largest predicted increases from the three areas are in Espoonlahti, based on table 25. There are no terrace houses in a 0–400-meter radius from the metro stations in any of the three areas. However, price development seems to be significantly increasing in 801–1000-meter radius from the metro station, especially in Soukka and Espoonlahti, but also in Kivenlahti, compared to other categories' development when considering terrace houses.

Table 26 Predicted yearly changes for houses

Houses	**Soukka**		**Espoonlahti**		**Kivenlahti**	
Yearly average change from 2019 (2018*) to 2023	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
One-bedroom	-5,02 %	-6,07 %	-	-	-	-
Two-bedrooms	-12,77 %*	-13,71 %*	-1,11 %*	-2,18 %*	-	-
Three-bedrooms or larger	-4,36 %	-5,41 %	8,25 %	7,06 %	-0,45 %*	-1,53 %*
Distance to metro station 601-800m	-3,84 %	-4,90 %	-	-	-	-
Distance to metro station 801-1000m	-4,73 %	-5,78 %	9,10 %	7,91 %	-0,45 %*	-1,53 %*
Year built 1970s	-2,86 %	-3,92 %	-	-	-	-
Year built 1980s	-3,96 %	-5,02 %	0,90 %	-0,21 %	1,69 % *	0,59 %*
Year built 1990s	-3,84 %*	-4,88 %*	-	-	-	-
Year built 2000s	-5,78 %	-6,81 %	-	-	-	-
Year built 2010s	-4,57 %*	-5,61 %*	-	-	-	-

In table 26, the predicted yearly changes from 2019 (2018*) to 2023 of houses are presented. Predicted development seems to be increasing only in Espoonlahti and in addition, in Kivenlahti for houses built in the 1980s. Price development in Soukka is predicted to de decreasing in every category. However, as it was presented in table 3 in chapter 3.1, only 4,09 % of the observations in the dataset are houses and these were divided into training and test sets. In figures 22-24 in chapter 4, the price development of houses for 2009-2019 is presented and it can be seen that there are large variations in the price development and relatively long breaks when there were no observations. In appendix 1, the yearly changes from 2010 to 2019 by house type are presented and the price development of houses in all three areas has been variative and most of the time decreasing. Because the number of observations for houses is low, each observation has a significant effect on the predicted price development. Because the training and test sets are divided randomly, the observations might variate from each other a lot and this might cause mispredictions. Due to the low number of observations of houses, the prediction results for this house type should be reviewed skeptically. As it was presented in chapter 4 in table 11, when evaluating the accuracy of the predictions for Q1-Q3/2020, the poorest accuracy compared to others is on houses in Soukka, which has significantly higher MSE and RMSE values than other areas or house types. By observing prediction results, it can be seen that the model has predicted significantly lower prices per square meter for houses that were built in the 2010s compared to realized prices. In figure 6 in chapter 3.1, the distribution of real estate's year built by areas was presented and from the box plot, it can be seen that there is not many real estate built in the 2000s or 2010s in Soukka. This explains the poor prediction performance of the prediction model for houses built in the 2010s in Soukka.

The main research question for the study was:

What model should be used to predict housing prices for Westmetro's phase 2 areas?

In the literature several methods are used to evaluate the effect of transport infrastructure on land values and housing prices. Many factors are affecting housing prices such as apartment and neighborhood characteristics beside the accessibility factor. Increase in housing price and land value cannot be valued without considering the other influencing factors. (Mulley & Tsai 2016) Based on academic literature and previous studies, the most common way in housing market related studies is to use hedonic price model, which considers a real estate as a bundle of different attributes. According to Chin & Chau (2003) the hedonic price model can be seen as a straightforward model as it only needs to have specific information for instance the housing price, the group of housing attributes, and appropriate specification of the functional relationships. No information about the housing buyers or sellers is needed.

The aim of forecasting is to create a prediction about the future values of the data and one way to achieve it is regression (Prabhu et al. 2019, 200). Regression analysis can be conducted with available information on housing price data, where the real estate price is the dependent variable and other characteristics are independent variables. (Banister 2007, 17) Based on previous studies, various methods are used for predicting housing prices, including OLS regression, ML regression, WLS regression as well as ANN-model. As the number of observations, time frame, and variables also varied a lot between studies, there was no model that would have performed always better than the others. However, the use of hedonic price models seems to be common in predictions as well. OLS regression is powerful technique for modeling continuous data, especially with a combination of dummy variables and transformed data (Hutcheson & Sofroniou 1999, 56).

In addition to linear regression, quadratic function of OLS regression is conducted. The regression model created by using quadratic function has an intercept term, linear and squared terms for each variable, and interaction variables from all products of pairs of the independent variables (MathWorks 2020). Stepwise selection is used to determine the best fit of the quadratic model. Linear and quadratic regression models are created by using the

collected data from the Westmetro's phase 2 areas. As a result, quadratic regression model gives the best accuracy, it is decided that the actual predictions are created by using the quadratic model.

This study aimed to predict the housing price development for the time Westmetro's phase 2 starts operating and the estimated time for Westmetro's phase 2 to start operating is in 2023. The problem was that the effect of start of the operating cannot be concluded based on data from 2009-2019 from phase 2 areas. As a result, housing price data from phase 1 station areas was collected from a timeframe 2007-2019 including time before the official decision about Westmetro's phase 1, time after the official decision, construction time and time after phase 1 started operating. 2007 had to be used as start year for phase 1 data collection, so that similar variables of the stages of Westmetro project can be created for phase 1 data as there is for phase 2 data. Phase 1 data was introduced and analyzed together with phase 2 data in chapter 3.1. Areas Niittykumpu and Matinkylä seemed to be the most similar areas with phase 2 areas based on price level, before the effects of Westmetro's phase 1 capitalized on housing prices. Niittykumpu and Matinkylä are also the two last stations on Westmetro's phase 1, meaning distance to city center from those areas is also the most similar to distance to city center that phase 2 areas have, when comparing the distances with phase 1 areas.

A quadratic stepwise model for all house types in the phase 1 area was created. From the output, presented in appendix 7, of the quadratic stepwise model for all house types in phase 1, it was possible to get the coefficients of interaction variables of x8 Niittykumpu : x23 Stage of metro: operating and x9 Matinkylä : x23 Stage of metro: operating. The coefficient that is used in the predictions for phase 2 areas is an average of coefficients of interaction variables x8 Niittykumpu : x23 Stage of metro: operating and x9 Matinkylä : x23 Stage of metro: operating. Dummy variable stage of metro: operating is added into prediction dataset and variable gets value 1=operating from Q1/2023 onwards. Before Q1/2023 variable gets value 0=not operating.

Finally, the predictions for phase 2 areas Soukka, Espoonlahti, and Kivenlahti are created by using the following equation, which is a combination of quadratic stepwise model for all house types in phase 2, the whole output presented in appendix 4, and quadratic stepwise model for all house types in phase 1:

$$Ln(Price\ per\ square\ meter) = \beta_0 + \beta_1 * D\_Apartment\ building + \beta_2 * D\_House + \beta_3 * D\_Espoonlahti + \beta_4 * D\_Kivenlahti + \beta_5 * Ln(Distance\ to\ city\ center) + \beta_6 * Ln(Distance\ to\ metro\ station) + \beta_7 * Ln(Square\ meters) + \beta_8 * Ln(Year\ built) + \beta_9 * Ln(Floor) + \beta_{10} * Rooms + \beta_{11} * Ln(Date\ of\ sale) + \beta_{12} * Ln(CPI) + \beta_{13} * Employment\ rate + \beta_{14} * Euribor\ 3\ months + \beta_{15} * D\_stage\_decision + \beta_{16} * D\_stage\_construction + \beta_{17} * Site\ ownership + \beta_{18} * D\_Apartment\ building * D\_Kivenlahti + \cdots + \beta_{66} * Euribor\ 3\ months\text{^}2 + \beta_{67} * D\_stage\_operating$$

There were three sub-questions for the study:

1. *What variables should be chosen to predict housing prices?*

In addition to choosing the method for predicting, the variables for the prediction model were chosen. Based on academic literature and previous research, the variables for the regression model were chosen. There are 17 independent variables, and they are divided into four categories: apartment, location, variables related to Westmetro, and macroeconomic variables. Some of them are collected directly from the Bank of Finland, HSP, Statistics Finland, and some are modified or created based on the information received from HSP.

Price per square meter is the dependent variable as it can be used to analyze the price development of apartments of different type and size. Date of sale information was modified into quarterly form.

According to Lönnqvist (2015, 28), it can be considered that the housing price formats from two parts: the value of its' physical features and land value. The housing price of apartments with otherwise similar features might vary a lot according to location. Apartment’s physical features include size, type, quality, and structural features. As there are different features in apartments, also the apartment buyers differ from each other as they have different requirements. When the buyer chooses the apartment, it also means choosing the environment, access to public transportation, services, and many other things that are

depending on the location. These features are strongly affecting on the choice of the apartment and the housing price. (Laakso & Loikkanen 2004, 241). The dataset is divided into 3 areas, and they were transformed into dummy variables, as according to Laakso (1997), it is possible to reduce heteroscedasticity and multicollinearity problems by using area-level dummy variables. Some of the physical features of the apartment such as sauna and balcony were excluded from the dataset due to low number of observations. Physical features of the apartment that were included into dataset were: year built, number of rooms, floor, square meter, house type and site ownership.

According to Abelson et al. (2005, 1) there are macroeconomic factors that effect on housing prices such as inflation, income, interest rates, stock markets and unemployment rate. Real estate markets are connected to capital markets and macroeconomic developments through debt financing as one often needs it to buy an apartment (Lönnqvist 2015, 27). As the aim of the study was to predict housing prices for the years 2020-2023, in addition to literature and previous studies, the available information for the prediction time frame affected as well. The Bank of Finland's predictions for the years 2020-2023 included CPI, Euribor 3 months and Employment rate and for this reason, they were selected to be the macroeconomic variables of the prediction model.

According to Laakso et al. (2016, 431) housing prices are expected to increase in the area when the accessibility improves and is relatively larger in the areas that have moderate travel distance to business and service concentrations. Lifetime of a transport infrastructure construction project is relatively long as planning, design and construction normally take time and it is probable that the effects of the project can be seen before the project is completed. The buyers will consider the real estate prices based on the information available, including the expected improvements in accessibility in the future. (Banister & Thurstain-Goodwin 2011; Yiu & Wong 2005) Based on previous studies, if the distance to station is at the most 800 meters it seems to have significantly increasing effect on housing prices. (Laakso 1987; McDonald & Osuji 1995; Peltomäki 2017; Harjunen 2018) For this reason, the stage of metro dummy variables was created: "official decision made" and dummy variable "under construction", while reference variable is the stage "no official decision". Also, distance measuring variables were created as variables "distance to city center" and "distance to metro station" were created. Distance to city center variable refers

to distance to Helsinki city center from the real estate, while distance to metro station refers to distance to closest metro station from the real estate in the accuracy of hundred meters.

This study aimed to predict the housing price development for the time Westmetro's phase 2 starts operating and the estimated time for Westmetro's phase 2 to start operating is in 2023, meaning the effect of the start of the operating cannot find out based on data from 2009-2019 from Soukka, Espoonlahti, and Kivenlahti. For this reason, a separate regression model for phase 1 areas is conducted, to get the effect of the metro's start to operate. Otherwise, that model includes the same variables but in addition to them, phase 1 regression model gets a dummy variable "stage of metro: operating" as Westmetro's phase 1 started first operating in December 2017. This “stage of metro: operating” is used for the phase 2 predictions as well by adding the coefficient from the phase 1 regression model into phase 2 regression model’s equation, meaning there are in total 18 independent variables used to create the predictions for the years 2020-2023 in Soukka, Espoonlahti and Kivenlahti.

2. *Does the used interest rate in the dataset have an effect on prediction accuracy based on historical data?*

Bank of Finland's predictions for macro-economic variables were used to create the actual predictions. Most of the mortgages in Finland are tied to Euribor 12 months. As the available predictions were for Euribor 3 months, it had to be tested whether the use of Euribor 3 months values in the dataset affects the reliability of the regression model, by comparing the accuracies of two regression models. When conducting the regression models for comparison, the other one uses the dataset that includes Euribor 3 months values and the other one that includes Euribor 12 months values otherwise in similar datasets. Once the best fitting model, the quadratic stepwise model for all house types, was selected as it had the best performance for valuations by using the test set, comparison of valuation results for datasets using Euribor 3 months and Euribor 12 months by using the same model was created. Comparisons of the Euribor 3 months and Euribor 12 months models for prediction areas Soukka, Espoonlahti, and Kivenlahti are presented in figures 18-20 respectively. MSE and RMSE values were lower for Euribor 3 months model, indicating that even though, Euribor 12 months is a more commonly used rate in mortgages, using Euribor 3 months in

the actual predictions do not have a negative effect on the accuracy and reliability of the prediction model.

3. *Does COVID-19 pandemic affect housing price predictions?*

Table 27 Comparison of predicted yearly changes by using different values for macroeconomic variables

Comparison of predicted yearly changes from 2019 (2018*) to 2023 by using different values for macroeconomic variables						
Predictions using Bank of Finland's predictions from December 2020:						
	Soukka		**Espoonlahti**		**Kivenlahti**	
All apartment sizes	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
All house types	5,53 %	4,37 %	7,49 %	6,31 %	3,18 %	2,05 %
Apartment buildings	6,88 %	5,71 %	7,48 %	6,30 %	3,42 %	2,29 %
Terrace houses	5,95 %	4,79 %	9,63 %	8,43 %	1,33 %	0,22 %
Houses	-4,60 %	-5,65 %	9,11 %	7,91 %	-0,57 %*	-1,53 %*
Predictions using Bank of Finland's predictions from December 2019:						
	Soukka		**Espoonlahti**		**Kivenlahti**	
All apartment sizes	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
All house types	3,87 %	2,21 %	6,12 %	4,42 %	2,26 %	0,63 %
Apartment buildings	5,38 %	3,70 %	6,14 %	4,44 %	2,65 %	1,01 %
Terrace houses	4,05 %	2,39 %	8,02 %	6,29 %	0,46 %	-1,14 %
Houses	-7,99 %	-9,69 %	5,98 %	4,02 %	-3,25 %*	-4,68 %*

In table 27 the predicted yearly changes from 2019 to 2023 by housing type are presented, changes calculated from 2018 are presented in red and with *, similarly as in tables in chapter 4. In table 27 there are two sets of predictions presented: the upper one used Bank of Finland's predictions of macroeconomic variables from December 2020 in the prediction dataset and in the lower section of the table Bank of Finland's predictions of macroeconomic variables from December 2019 were used in the prediction dataset. As mentioned in chapter 3.2.5, in December 2019 predictions Euribor was predicted to start increase and inflation to grow faster, while the employment rate is predicted to be higher than in December 2020 predictions. The yearly changes are predicted to be smaller when using December 2019's predictions for macroeconomic variables and it can be seen from the presented results how the interest rates and changes in inflation affect housing prices. Despite the growing rate of employment, when CPI is growing faster and Euribor 3 months starts to increase, the predicted price development is also lower. It has to be taken into consideration, that predictions for macroeconomic variables from December 2019 are no longer relevant, but also the predictions from December 2020 are created under exceptional circumstances.

Comparison of these results is an example, how the predicted development might vary depending on the changes in the economy during and after the COVID-19 pandemic.

In addition to predictions created for analyzing whether COVID-19 affects housing price predictions, another set of alternative predictions was created by conducting a prediction model by using an otherwise similar dataset, but it does not have the stage of metro variables. So, even though the metro is already under construction during the timeframe 2009-2019 and it might affect price behavior, there is no variables in the used data that indicate the stages of the metro project. When considering all house types, apartment buildings, and terrace houses, the yearly average changes from 2019 to 2023 are significantly lower by using this model for predictions, compared to predictions that include the stage of metro variables in the data. An interesting observation is that for houses, the average yearly change from 2019 (2018*) to 2023 is predicted to be higher when the stage of metro variables is not included in the prediction. For Espoonlahti and Kivenlahti the average yearly changes for houses are significantly higher and for Soukka, the predicted yearly change is still decreasing but the decrease is not as large as when the stage of metro variables is included in the model. For the years 2020-2022, there are some differences in the predictions, but they are relatively small. The most significant, and large differences are predicted for 2023, when is the expected start for Westmetro's phase 2 to start operating. In table 23 in chapter 4.5, where the stage of metro variables is not included in the regression model, there is only a small increase or decrease predicted for 2023 and as an opposite, in table 12 in chapter 4, where the stage of metro variables is included in the regression model, there is relatively large increase predicted at housing prices for 2023, except for houses.

The data analysis and results of this study give detailed information about the housing market in the observed areas for operators in the real estate business, individuals, who are interested to buy or sell real estate, and for future research of a similar field. The results of this study can be used in the evaluation of the value of the real estate and to support decision-making in investments, as well as in municipal zoning decisions. Overall, the accuracy and fit of the model are relatively good and the model could be adapted for other studies for housing price prediction such as in the station areas of tram line in Tampere and Jokeri light rail in Helsinki and Espoo, for the latter one with modifications for distance to city center variable as the Jokeri light rail is a transversal route between the cities and does not go through the city center.

5.2 Comparison to previous research

In previous studies, the results are varying how the better accessibility affected housing prices. Previous studies showed that, if the distance to station is at the most 800 meters it seems to have significantly increasing effect on housing prices. (Laakso 1987; McDonald & Osuji 1995; Peltomäki 2017; Harjunen 2018) The results from this study indicate, that when considering apartment buildings and terrace houses, the predicted increase in housing price around the expected start of metro's operating is significant in all distance groups within 0–1000-meter radius from the metro station. However, the predicted increase is not only based on the effect of metro, but all the considered variables in the equation that are used for the prediction. When considering only houses, the predicted price development is mostly decreasing, before and after the expected start of metro's operating. This variates significantly from the previous studies, but as it was mentioned in chapter 4.3, the lack of observations might cause mispredictions for predicted price development of houses as the small number of observations of houses causes large variations in average prices per square meter during the observation time frame that is used to conduct the model for prediction.

There is variation how the fit of the model and prediction accuracy are reported in previous studies that have used OLS regression as a method. In Dubin's (1998) study adjusted R-squared for the prediction model was 0,731 and sum of squared errors 81116,89 while the total number of observations in the study was 1493 and used splitting ratio was 66-33, so test set included 493 observations. In the study of Ottensman et al. (2008) there is no mention that data would have been split and R-squared for the best fit model is 0,8616. Based on the results from previous studies that used OLS regression for predicting housing prices, performance of the model in this study overcomes them as the adjusted R-squared of the prediction model is 0,869. MSE for test set predictions in functional form is 110458,08, while the number of observations in the test set is 961. However, MSE values between the studies cannot be compared as the data and prediction timeframe are not the same.

5.3 Limitations and recommendations for further research

When evaluating the reliability of the research, the used data should be considered. As it was mentioned in chapter 3.1 the data is entered in HSP manually, so there is a chance for misspelling or missing values as not all the information is mandatory to enter. The collected data was reviewed carefully, and some small defects were corrected from the data. Some variables had missing values more often, for example, variable maintenance fee, sauna, or balcony, so the whole variable was excluded from the dataset. Rows with missing values were deleted, usually, the missing information was year built, number of rooms, or site ownership. However, data from HSP is the best available and is suitable for studies about the real estate market.

The number of observations for terrace houses and houses is relatively low, which affects the reliability of the research. When there are several quarters or years with no observations from the specific housing type in the area and when there are large variations in the realized prices, it affects the prediction accuracy of the created model. However, houses were decided to keep in the dataset and regression model used for prediction, because as it was presented in chapter 3.2.2, the quadratic stepwise model for all house types and quadratic stepwise model for apartment buildings and terrace houses did not significantly differ from each other based on performance. However, when the accuracy of the prediction results was analyzed for Q1-Q3/2020, as there already was the data from the realized sales from that time, it was possible to see how the distribution of real estate's year built in the area affects the predictions. As presented in figure 6 in chapter 3.1, there are not many real estate in Soukka, which are built in 2000s or 2010s. As the data used for training of the prediction model, does not have many observations of houses in Soukka, that would have been built in 2000s or 2010s, there are large mispredictions in Q1-Q3/2020 for houses in Soukka which were built in 2010s and it weakens the accuracy of the model. Also, as the observations that are used for prediction, are modified based on existing data, there are no new-built apartments or houses for 2020-2023. It should be noted when analyzing the prediction results. But as it can be seen already in the case prediction accuracy of houses built in the 2010s in Soukka, it would not have been well-grounded to add observations with year built values of 2020s in the prediction dataset, as the mispredictions would have probably been even larger as there would not have been any observations with year built in the 2020s in the training data that was used for conducting the prediction model.

Not all of the BLUE assumptions in the model are fulfilled. The residuals are not normally distributed based on Bera-Jarque test and there is a multicollinearity problem with some of the independent variables as they have VIF values over 10. However, the model is used to predict the housing prices, instead of interpreting the coefficients of the variables, multicollinearity does not cause large damage.

In chapter 2.1 it was mentioned that the seaside view might increase the price per square meter even 2000 euros compared to the similar apartment without the seaside view. Because all three areas Soukka, Espoonlahti, and Kivenlahti are nearby the coast and a 1-kilometer radius from the metro station runs partially to the seaside, a conclusion can be drawn that the seaside view creates a variation on housing prices in these areas. This might also create large differences in prices per square meter within the same apartment building if the other apartment has a seaside view and the other does not. The problem with the accuracy of the model is that there is no information on whether there is a seaside view from the sold apartment or not. Another problem is that there is no information in the data available if large renovations are coming or if they have been done on the sold real estate. In chapter 2.1 it was mentioned that if the plumping renovation has been done in the apartment building it might increase the price per square meter of the apartment by 850 euros. However, information about the seaside view or the renovations was not available and as it was mentioned above, the used data from HSP is the best available.

It would have been ideal to get a traveling time from the observed real estate to the specific destination before and after the improved accessibility. However, it was not possible to get that information for the time before phase 1 started operating from sources available.

It should be considered that the predictions of the Bank of Finland for CPI, Employment rate, and Euribor 3 months for years 2020-2023 are from December 2020 and created during the global COVID-19 pandemic. Prediction results should be considered with condition that the factors that affect the prediction and which have been used for the prediction, might change but especially during this exceptional time when it is very unclear what is going to be the effect of the COVID-19 pandemic on the economy in the following years.

Based on the findings of this study, recommendations for further research are presented. The prediction model of this study could be improved by adding environmental variables such as a seaside view or if possible, by adding traveling time measuring variable to see whether it gives better accuracy for the prediction model. A natural further study for this research would be few years after phase 2 starts operating to study the capitalized effect of phase 2 on housing prices in the station areas. A similar study could be conducted for office and business premises. If there are more studies predicting housing price development for areas benefitting from improved accessibility, a comparative study could be conducted, to find out which model has the best accuracy. Due to the COVID-19 pandemic, remote work has increased significantly, which raises the question of whether accessibility to central business districts will have a significant part on housing prices and land values in the future. Due to remote work, people might want more space in their apartments which might affect the demand and price development of the different sized real estate.

LIST OF REFERENCES

Abelson, P., Joyeux, R., Milunovich, G. and Chung, D. (2005). House Prices in Australia: 1970 to 2003 facts and explanations. [Online]. [Accessed at 27.6.2020]. Available at: https://www.proptechnow.com.au/wp-content/uploads/uploads/houseprices-1.pdf

Agostini, C. and Palmucci, G. (2008). The Anticipated Capitalization Effect of a New Metro Line on Housing Prices. Fiscal studies 29(2), 233–256.

Bae, C-H. C., Jun, M-J. & Park, H. (2003). The impact of Seoul's subway Line 5 on residential property values. Transport Policy. Vol 10:2. s. 85-94.

Banister, D. (2007). Quantification of the non-transport benefits resulting from rail investment. Working paper N. 1029. Transport Studies Unit Oxford University Centre for the Environment. [Online]. [Accessed at 6.8.2020] Available at: https://www.tsu.ox.ac.uk/pubs/1029-banister.pdf

Banister, D. & Thurstain-Goodwin, M. (2011). Quantification of the non-transport benefits resulting from rail investment. Journal of Transport Geography 19 (2011) 212–223.

Bank of Finland. (2019). Ennuste vuosille 2019–2022 (joulukuu 2019). [Online]. [Accessed at 27.2.2021]. Available at: https://helda.helsinki.fi/bof/bitstream/handle/123456789/16529/eurotalous_2019_5_ennustetaulukot.pdf?sequence=1&isAllowed=y

Bank of Finland. (2020a). Euribor*-korot ja Eonia*-korko, kuukauden keskiarvo. [Online]. [Accessed at 26.03.2020]. Available at: https://www.suomenpankki.fi/fi/Tilastot/korot/taulukot2/korot_taulukot/euribor_korot_long_fi/

Bank of Finland. (2020b). Forecast tables for 2020–2023 (December 2020). [Online]. [Accessed at 27.2.2021]. Available at: https://www.bofbulletin.fi/en/2020/6/forecast-tables-for-2020-2023-december-2020/

Brandt, S. & Maennig, W. (2011). The impact of rail access on condominium prices in Hamburg. Transportation (2012). Vol 39. 997–1017.

Brooks, C. 2014. Introductory Econometrics for Finance. Third edition. Cambridge. Cambridge University Press.

Brotherus, J. (2019). Hypon asuntomarkkinakatsaus Q3/2019. [Online]. [Accessed at 11.7.2020]. Available at: http://www.hypo.fi/wp-content/uploads/2019/09/Hypo_Asuntomarkkinakatsaus_syyskuu2019.pdf

Chin, T. L. & Chau, K. W. (2003). A critical review of literature on the hedonic price model, International Journal for Housing and Its Applications 27 (2), 145-165.

Citycon. (2020). Lippulaiva. [Online]. [Accessed at 7.10.2020]. Available at: https://www.citycon.com/fi/kauppakeskukset/kehityshankkeet/lippulaiva

Debrezion, G., Pels, E. and Rietveld, P. (2007). The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-analysis. Journal of Real Estate Finance & Economics 35(2), 161–180.

Dubé, J.,Thériault, M. & Des Rosiers, F. (2013). Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009. Transportation Research Part A 54 (2013) 49-66.

Dubin, R. (1998). Predicting House Prices Using Multiple Listings Data. Journal of Real Estate Finance & Economics Vol 17:1, 36-39.

Dunse, N. & Jones, C. (1998). A hedonic price model of office rents. Journal of Property Valuation & Investment, Vol. 16 No. 3, 1998, pp. 297-312.

Espoo. (2012). Valtuusto kannattaa Länsimetron jatkamista Matinkylästä Kivenlahteen. [Online]. [Accessed at 2.10.2020]. Available at: https://www.espoo.fi/fi-FI/Valtuusto_kannattaa_Lansimetron_jatkamis(19466)

Espoo. (2014). Länsimetro Kivenlahteen. [Online]. [Accessed at 2.10.2020]. Available at: https://www.espoo.fi/fi-FI/Asuminen_ja_ymparisto/Kaavoitus/Asemakaava/Asemakaavoituskohteet/Espoonlahti/Kivenlahden_metrokeskus_412500/Lansimetro_Kivenlahteen(54964)

Fejarang, Robert A. (1994). Impact on Property Values: A Study of the Los Angeles Metro Rail. Preprint, Transportation Research Board, 73rd Annual Meeting, Washington, D.C., January 9-13.

Garg, A. (2016). Statistical Methods for Estimating House Price Index. Journal of Business & Financial Affairs. [Online]. [Accessed at 23.8.2020]. Available at: https://www.omicsonline.org/open-access/statistical-methods-for-estimatinghouse-price-index-2167-0234-1000231.pdf

Harjunen, O. (2018) Metro investment and the housing market anticipation effect. [Online]. [Accessed at 18.7.2020]. Available at: https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/18_01_25_tyopapereita_02_Harjunen.pdf

Henneberry, J. (1998). “Transport Investment and House Prices.” Journal of Property Valuation and Investment 16(2): 144-158.

Hutcheson, G. D. (2011). Ordinary Least-Squares Regression. The SAGE Dictionary of Quantitative Management Research. Pages 224-228. [Online]. [Accessed at 15.8.2020]. Available at: https://datajobs.com/data-science-repo/OLS-Regression-[GD-Hutcheson].pdf

Hutcheson, G. D. & Sofroniou, N. (1999). The Multivariate Social Scientist. London: Sage Publications.

Keskinen, P., Karikallio, H. and Kiviholma, S. (2020). PTT-ennuste: Alueelliset asuntomarkkinat 2020. Helsinki 2020. [Online]. [Accessed at 2.7.2020]. Available at: https://www.ptt.fi/ajankohtaista/uutiset/asuntomarkkinat-2020-ennuste.html

Kodit.io. (2018). Miten uudet liikenneyhteydet vaikuttavat asuntojen hintoihin pääkaupunkiseudulla? [Online]. [Accessed at 23.8.2020]. Available at:

https://www.kodit.io/fi/uutiset/miten-uudet-liikenneyhteydet-vaikuttavat-asuntojen-hintoihin-paakaupunkiseudulla

KvantiMOTV. (2003). Regressioanalyysin rajoitteet. [Online]. [Accessed at 16.8.2020]. Available at: https://www.fsd.tuni.fi/menetelmaopetus/regressio/rajoitteet.html

KVKL. (2018). Kvkl hsp-hintaseurantapalvelun käyttöehdot 5/2018. [Online]. [Accessed 11.9.2020]. Available at: https://www.hintaseurantapalvelu.fi/kvkl-hsp-asiakassopimus-ja-palvelukuvaus-2018.pdf

Laakso, S. (1986). Metro ja kaupunkirakenne. Helsingin metron vaikutus asuntojen hintoihin ja toimistotilojen vuoriin. Helsingin kaupungin kaupunkisuunnitteluvirasto yleiskaavaosasto. Julkaisu YB:17/86. 35+27 pages.

Laakso, S. (1997). Urban Housing Prices and the Demand for Housing Characteristics. Elinkeinoelämän tutkimuslaitoksen julkaisuja A27. Tampere. Tammer-Paino Oy.

Laakso, S., Kostiainen, E. and Metsäranta, H. (2016). Liikennehankkeiden laajemmat taloudelliset vaikutukset. Kansantaloudellinen aikakauskirja – 112. vsk. – 4/2016. [Online]. [Accessed at 28.6.2020] Available at: http://www.taloustieteellinenyhdistys.fi/wp-content/uploads/2016/12/kak-4_2016-laakso-ym.pdf

Laakso, S. & Loikkanen, H. A. (2004). Kaupunkitalous. Johdatus kaupungistumiseen, kapunkien maankäyttöön sekä yritysten ja kotitalouksien sijoittumiseen. Helsinki: Gaudeamus Oy.

Laine, S. (2017). Raideliikennehankkeiden vaikutus asuntojen markkina-arvoon – Tapaus: Kehärata. Master’s Thesis. Aalto University – School of Engineering. Espoo.

Limsombunchai, V., Gan, C. & Lee, M. (2004). House Price Prediction: Hedonic Price Model vs. Artificial Neural Network. American Journal of Applied Sciences 1 (3): 193-201, 2004.

Lindblad, A., Sariola, M. & Viertola, H. (2019). What factors influence house prices and residential construction? [Online]. [Accessed at 2.7.2020]. Available at:

https://www.bofbulletin.fi/en/2019/3/what-factors-influence-house-prices-and-residential-construction/

Länsimetro. (2014). First blast on the Matinkylä–Kivenlahti metro section at the Finnoo access tunnel worksite. [Online]. [Accessed at 2.10.2020]. Available at: https://www.lansimetro.fi/en/news/first-blast-on-the-matinkyla-kivenlahti-metro-section-at-the-finnoo-access-tunnel-worksite/

Länsimetro. (2020a) Starting points. [Online]. [Accessed at 7.10.2020] Available at: https://www.lansimetro.fi/en/lahtokohdat-en-translation/

Länsimetro. (2020b). Stations. [Online]. [Accessed at 7.10.2020]. Available at: https://www.lansimetro.fi/en/stations/

Länsimetro. (2020c). Usein kysyttyjä kysymyksiä. [Online]. [Accessed at 7.10.2020]. Available at: https://www.lansimetro.fi/ota-yhteytta/usein-kysyttyja-kysymyksia/#30d0ea41

Länsimetro. (2020d). Länsimetro Oy. [Online]. [Accessed at 7.10.2020]. Available at: https://www.lansimetro.fi/en/information-on-the-project/lansimetro-oy/

Lönnqvist, H. (2015). On the effects of urban natural amenities, architectural quality and accessibility to workplaces on housing prices–An empirical study on the Helsinki Metropolitan area. [Online]. [Accessed at 27.6.2020]. Available at: https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/16_02_04_Tutkimuksia_5_2015_Lonnqvist.pdf

Mapdevelopers. (2020). Draw a circle - Create a circle on a google map using a point and a radius. [Online]. [Accessed at 27.03.2020]. Available at: https://www.mapdevelopers.com/draw-circle-tool.php

MathWorks. (2020). Stepwiselm. [Online]. [Accessed at 20.11.2020]. Available at: https://se.mathworks.com/help/stats/stepwiselm.html

McDonald, J. F. and Osuji, C. I. (1995). The effect of anticipated transportation improvement on residential land values. Regional Science & Urban Economics, 25(3), 261–278.

McKinley, S. and Levine, M. (n.d.). Cubic Spline Interpolation. Math 45: Linear Algebra. [Online]. [Accessed at 25.10.2020]. Available at: https://www.rajgunesh.com/resources/downloads/numerical/cubicsplineinterpol.pdf

Mellin, I. (2006). Tilastolliset menetelmät: Lineaarinen regressioanalyysi. TKK. [Online]. [Accessed at 6.8.2020]. Available at: https://math.aalto.fi/opetus/sovtoda/oppikirja/Regranal.pdf

Microsoft. (2018). Training and Testing Data Sets. [Online]. [Accessed at 7.11.2020]. Available at: https://docs.microsoft.com/en-us/analysis-services/data-mining/training-and-testing-data-sets?view=asallproducts-allversions

Mohammad, S., Graham, D. & Melo, P. (2015). The effect of the Dubai Metro on the value of residential and commercial properties. The Journal of transport and land use. Volume 10. 263 – 290.

Mulley, C., Liang, M., Clifton, G., Yen, B. and Burke, M. (2016). Residential property value impacts of proximity to transport infrastructure: An investigation of bus rapid transit and heavy rail networks in Brisbane, Australia. Journal of Transport Geography 54 (2016) 41–52.

Mulley, C. & Tsai, C. (2016). When and how much does new transport infrastructure add to property values? Evidence from the bus rapid transit system in Sydney, Australia. Transport Policy 51 (2016) 15–23.

Nordea. (2020). Euribor 12 kk, Prime muut viitekorot. [Online]. [Accessed att 4.11.2020]. Available at: https://www.nordea.fi/henkiloasiakkaat/korot.html

O’Brien, R. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity (2007) 41:673–690.

Oikarinen, E. (2011). Asuntohintojen kansantaloudelliset vaikutukset. Kansantaloudellinen aikakauskirja – 107. vsk. – 2/2011. [Online]. [Accessed at 2.7.2020]. Available at: http://taloustieteellinenyhdistys.fi/images/stories/kak/KAK22011/kak22011oikarinen.pdf

OP. (2020a). Euribor on yleisin viitekorko. [Online]. [Accessed at 4.11.2020]. Available at: https://www.op.fi/henkiloasiakkaat/lainat-ja-asunnot/korot-ja-hinnat/euribor

OP. (2020b). Asuntomarkkinakatsaus 2020/Q1. [Online]. [Accessed at 22.8.2020]. Available at: https://www.op.fi/documents/20556/28168687/Asuntomarkkinakatsaus_Q1_2020/68d515e8-c2ca-4d62-7b3d-a50f22e0f7cb

Ottensmann, J., Payton, S. & Man, J. (2008) Urban Location and Housing Prices within a Hedonic Model. The Journal of Regional Analysis and Policy 38(1): 19-35.

Pan, Q., Pan, H., Zhang, M. & Zhong, B. (2014). Effects of Rail Transit on Residential Property Values. Transportation Research Record Iss. 2453 (2014) 118-127.

Peltomäki, O. (2017). Länsimetron vaikutukset tulevien asemaympäristöjen kerrostaloasuntojen kauppahintoihin ja väestörakenteeseen ennen metron valmistumista. Master's Thesis. Aalto University – School of Engineering. Espoo.

Prabhu, C.S.R., Chivukula, A., Mogadala, A., Ghosh, R. & Livingston, L.M. (2019). Big Data Analytics: Systems, Algorithms, Applications. Springer Singapore Pte. Limited.

PTT. (2020). Asuntomarkkinat 2020 erikoisteema - Asuntomarkkinoiden polarisaatio jatkuu 2020-luvulla. [Online]. [Accessed at 2.7.2020]. Available at: https://www.ptt.fi/julkaisut-ja-hankkeet/uutiset/asuntomarkkinat-2020-erikoisteema-asuntomarkkinoiden-polarisaatio-jatkuu-2020-luvulla.html

Putkuri, H. (2018). Wide regional disparities in Finnish house prices and household indebtedness. [Online]. [Accessed at 11.7.2020]. Available at: https://www.bofbulletin.fi/en/2018/2/wide-regional-disparities-in-finnish-house-prices-and-household-indebtedness/

Puumalainen, K. (2019). Multivariate and econometric analysis methods. Lecture 3b Linear regression analysis. LUT University.

Re/max. (2018). Putkiremontti – uhka vai mahdollisuus? [Online]. [Accessed at 7.3.2021]. Available at: https://www.remax.fi/putkiremontti-uhka-vai-mahdollisuus/

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. Journal of Political Economy, Vol.82, pp. 34-55.

Talouselämä. (2015). Hissi, sauna, merinäköala... IS: Näin eri ominaisuudet vaikuttavat asunnon hintaan. [Online]. [Accessed at 1.12.2020]. Available at: https://www.talouselama.fi/uutiset/hissi-sauna-merinakoala-is-nain-eri-ominaisuudet-vaikuttavat-asunnon-hintaan/2a88d7be-a1e9-3a0f-96b7-275c47a9d3b3

Tapiola. (2019). Kauppakeskus AINOA rakennettiin kolmessa vaiheessa. [Online]. [Accessed at 7.10.2020]. Available at: https://www.tapiolankeskus.fi/fi/Tapiola-uudistuu/AINOA-kokonaisuus

Tilastokeskus. (2020a). Tilastokeskuksen PxWeb-tietokannat. [Online]. [Accessed at 27.03.2020]. Available at: http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin__hin__khi__kk/statfin_khi_pxt_11xf.px/

Tilastokeskus. (2020b). Tilastokeskuksen PxWeb-tietokannat. [Online]. [Accessed at 4.10.2020]. Available at: http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin__tym__tyti__nj/statfin_tyti_pxt_11c8.px/

Tilastokeskus. (2020c). Kuluttajahintaindeksi, [Online]. [Accessed at 4.10.2020]. Available at: http://tilastokeskus.fi/til/khi/index.html

Tilastokeskus. (2020d). 6.8 Kuluttajahintaindeksin käyttö. [Online]. [Accessed at 4.10.2020]. Available at:

https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_inde&lesson_id=6&page_type=sisalto&subject_id=8

Tilastokeskus. (2020e). Arvon kiinteähintaistaminen eli deflatointi. [Online]. [Accessed at 4.10.2020]. Available at: https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?page_type=esim&course_id=tkoulu_inde&lesson_id=6&subject_id=8&example_id=1

Valaja, A. (2018). Raitiotien vaikutus asuntojen hintoihin Tampereella. Master's Thesis. Master's Thesis. Aalto University – School of Engineering. Varkaus.

Yiu, C. Y. & Wong, S. K. (2005). The Effects of Expected Transport Improvements on Housing Prices. Urban Studies, Vol. 42, No. 1, 113– 125, January 2005. [Online]. [Accessed at 28.6.2020]. Available at: https://journals-sagepub-com.ezproxy.cc.lut.fi/doi/pdf/10.1080/0042098042000309720

Yle. (2013). Selvitys: Tehty putkiremontti nostaa hintaa ja lyhentää myyntiaikaa. [Online]. [Accessed at 7.3.2021]. Available at: https://yle.fi/uutiset/3-6722809

Yle. (2017). Kaikki raidehankkeet eivät nosta asunnon hintaa – tältä hintakehitys näyttää kehäradan, länsimetron ja Kruunusiltojen naapurissa. [Online]. [Accessed at 23.8.2020]. Available at: https://yle.fi/uutiset/3-9685402

Zhang, M., Meng, X., Wang, L. & Xu, T. (2014). Transit development shaping urbanization: Evidence from the housing market in Beijing. Habitat International 44 (2014) 545-554.

APPENDICES

Appendix 1 Yearly nominal and real changes of average prices per square meter by areas

All house types

Yearly changes	Lauttasaari		Koivusaari		Keilaniemi		Otaniemi	
all house types	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	14,01 %	12,64 %	6,44 %	5,16 %	18,29 %	16,87 %	22,27 %	20,80 %
2011	11,36 %	7,63 %	5,10 %	1,58 %	7,53 %	3,93 %	6,86 %	3,28 %
2012	2,83 %	0,02 %	4,82 %	1,96 %	4,18 %	1,34 %	8,00 %	5,05 %
2013	-0,09 %	-1,55 %	7,64 %	6,08 %	-4,14 %	-5,53 %	0,09 %	-1,36 %
2014	5,29 %	4,21 %	-4,52 %	-5,50 %	11,73 %	10,57 %	0,14 %	-0,90 %
2015	3,84 %	4,05 %	2,93 %	3,14 %	-2,48 %	-2,28 %	-2,61 %	-2,41 %
2016	6,19 %	5,82 %	1,67 %	1,31 %	5,21 %	4,84 %	0,37 %	0,02 %
2017	4,99 %	4,20 %	2,14 %	1,38 %	5,35 %	4,56 %	0,25 %	-0,50 %
2018	1,37 %	0,29 %	6,84 %	5,69 %	2,23 %	1,14 %	8,96 %	7,79 %
2019	5,28 %	4,21 %	7,06 %	5,97 %	5,19 %	4,12 %	1,16 %	0,14 %

Yearly changes	Tapiola		Urheilupuisto		Niittykumpu		Matinkylä	
all house types	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	13,22 %	11,86 %	9,90 %	8,58 %	5,32 %	4,06 %	19,91 %	18,47 %
2011	24,42 %	20,25 %	-2,99 %	-6,24 %	-0,57 %	-3,90 %	16,27 %	12,37 %
2012	-2,27 %	-4,94 %	12,68 %	9,60 %	17,75 %	14,54 %	11,47 %	8,42 %
2013	9,76 %	8,16 %	0,58 %	-0,88 %	16,63 %	14,93 %	-2,45 %	-3,87 %
2014	-4,48 %	-5,46 %	-6,41 %	-7,37 %	-0,79 %	-1,81 %	-9,21 %	-10,14 %
2015	6,28 %	6,50 %	6,38 %	6,60 %	-2,71 %	-2,50 %	4,33 %	4,55 %
2016	12,63 %	12,23 %	-0,45 %	-0,80 %	-12,08 %	-12,39 %	-5,21 %	-5,55 %
2017	1,43 %	0,67 %	32,94 %	31,94 %	52,45 %	51,31 %	15,13 %	14,27 %
2018	-1,62 %	-2,68 %	12,04 %	10,84 %	4,24 %	3,13 %	-1,68 %	-2,73 %
2019	9,59 %	8,48 %	3,21 %	2,16 %	7,48 %	6,39 %	-0,14 %	-1,16 %

Yearly changes	Soukka		Espoonlahti		Kivenlahti	
all house types	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	13,38 %	12,02 %	8,77 %	7,46 %	3,56 %	2,32 %
2011	-0,46 %	-3,80 %	5,45 %	1,91 %	4,30 %	0,81 %
2012	3,20 %	0,38 %	23,85 %	20,47 %	3,35 %	0,53 %
2013	1,08 %	-0,40 %	5,56 %	4,02 %	0,60 %	-0,86 %
2014	0,64 %	-0,40 %	-3,11 %	-4,11 %	-2,18 %	-3,19 %
2015	4,06 %	4,28 %	-3,55 %	-3,35 %	1,08 %	1,29 %
2016	3,85 %	3,48 %	-11,63 %	-11,94 %	1,23 %	0,87 %
2017	-0,57 %	-1,31 %	3,57 %	2,79 %	10,64 %	9,82 %
2018	2,11 %	1,02 %	-3,88 %	-4,91 %	4,36 %	3,24 %
2019	-0,12 %	-1,13 %	4,65 %	3,59 %	0,94 %	-0,09 %

Apartment buildings

Yearly changes	Lauttasaari		Koivusaari		Keilaniemi		Otaniemi	
of apt. buildings	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	13,98 %	12,61 %	6,29 %	5,02 %	18,03 %	16,62 %	19,18 %	17,75 %
2011	11,17 %	7,45 %	5,22 %	1,69 %	7,76 %	4,15 %	6,86 %	3,28 %
2012	2,89 %	0,08 %	4,47 %	1,61 %	3,36 %	0,54 %	8,00 %	5,05 %
2013	-0,16 %	-1,61 %	7,68 %	6,11 %	-3,38 %	-4,79 %	-0,65 %	-2,10 %
2014	5,44 %	4,35 %	-4,26 %	-5,25 %	9,55 %	8,42 %	0,52 %	-0,52 %
2015	3,90 %	4,12 %	2,98 %	3,20 %	-0,54 %	-0,34 %	-2,25 %	-2,05 %
2016	6,07 %	5,69 %	0,08 %	-0,27 %	5,21 %	4,84 %	0,37 %	0,02 %
2017	5,37 %	4,58 %	3,05 %	2,28 %	6,12 %	5,33 %	0,25 %	-0,50 %
2018	1,30 %	0,21 %	6,15 %	5,01 %	1,49 %	0,40 %	8,96 %	7,79 %
2019	4,97 %	3,91 %	9,15 %	8,04 %	5,19 %	4,12 %	3,39 %	2,34 %

Yearly changes of apt. buildings	Tapiola		Urheilupuisto		Niittykumpu		Matinkylä	
	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	13,94 %	12,57 %	15,75 %	14,36 %	0,99 %	-0,22 %	21,11 %	19,66 %
2011	22,73 %	18,62 %	-6,17 %	-9,32 %	2,58 %	-0,86 %	16,50 %	12,60 %
2012	-0,89 %	-3,59 %	12,21 %	9,15 %	10,09 %	7,08 %	11,81 %	8,76 %
2013	10,00 %	8,40 %	6,46 %	4,91 %	23,71 %	21,91 %	-2,21 %	-3,63 %
2014	-3,51 %	-4,51 %	-7,81 %	-8,76 %	5,16 %	4,08 %	-9,77 %	-10,70 %
2015	5,25 %	5,47 %	6,34 %	6,56 %	-5,33 %	-5,14 %	4,11 %	4,32 %
2016	12,66 %	12,26 %	-2,52 %	-2,86 %	-15,36 %	-15,66 %	-5,15 %	-5,48 %
2017	1,88 %	1,11 %	30,60 %	29,62 %	58,10 %	56,91 %	15,53 %	14,67 %
2018	-1,13 %	-2,19 %	10,08 %	8,90 %	4,57 %	3,45 %	-1,64 %	-2,70 %
2019	8,07 %	6,97 %	6,22 %	5,14 %	6,84 %	5,75 %	-0,11 %	-1,12 %

Yearly changes of apt. buildings	Soukka		Espoonlahti		Kivenlahti	
	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	8,79 %	7,48 %	8,70 %	7,39 %	3,54 %	2,30 %
2011	2,40 %	-1,03 %	6,34 %	2,78 %	4,78 %	1,27 %
2012	-2,50 %	-5,16 %	25,97 %	22,53 %	2,07 %	-0,72 %
2013	7,56 %	5,99 %	5,00 %	3,47 %	0,95 %	-0,52 %
2014	1,30 %	0,25 %	-3,26 %	-4,26 %	-1,92 %	-2,93 %
2015	0,27 %	0,47 %	-2,20 %	-1,99 %	1,83 %	2,04 %
2016	8,75 %	8,36 %	-13,15 %	-13,46 %	-1,16 %	-1,51 %
2017	-2,99 %	-3,72 %	3,61 %	2,83 %	12,85 %	12,00 %
2018	3,92 %	2,80 %	-3,19 %	-4,23 %	4,88 %	3,76 %
2019	-1,40 %	-2,40 %	5,23 %	4,16 %	0,57 %	-0,45 %

Terrace houses

Yearly changes of terrace houses	Lauttasaari		Koivusaari		Tapiola		Urheilupuisto	
	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	-	-	-	-	-9,27 %	-10,36 %	17,83 %	16,41 %
2011	67,68 %	62,07 %	-	-	50,09 %	45,06 %	-4,29 %	-7,49 %
2012	-25,91 %	-27,94 %	-	-	-23,62 %	-25,71 %	5,52 %	2,64 %
2013	12,78 %	11,14 %	-5,62 %	-6,99 %	4,56 %	3,04 %	-7,23 %	-8,59 %
2014	-14,62 %	-15,50 %	0,29 %	-0,75 %	-10,07 %	-11,00 %	-8,11 %	-9,06 %
2015	9,11 %	9,34 %	-	-	36,61 %	36,89 %	22,75 %	23,00 %
2016	14,72 %	14,31 %	-	-	-11,37 %	-11,69 %	-8,02 %	-8,34 %
2017	-32,55 %	-33,06 %	2,33 %	1,56 %	-11,55 %	-12,21 %	22,32 %	21,41 %
2018	-9,66 %	-10,63 %	1,38 %	0,29 %	16,20 %	14,95 %	8,83 %	7,66 %
2019	84,36 %	82,49 %	0,39 %	-0,63 %	9,17 %	8,07 %	-4,90 %	-5,87 %

Yearly changes of terrace houses	Niittykumpu		Matinkylä		Soukka		Espoonlahti		Kivenlahti	
	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	20,57 %	19,13 %	9,67 %	8,35 %	14,00 %	12,63 %	15,06 %	13,68 %	1,39 %	0,17 %
2011	-10,39 %	-13,40 %	9,65 %	5,98 %	-2,85 %	-6,10 %	-7,79 %	-10,87 %	-2,23 %	-5,50 %
2012	20,39 %	17,10 %	5,82 %	2,93 %	13,25 %	10,16 %	8,58 %	5,61 %	22,09 %	18,76 %
2013	11,60 %	9,97 %	-14,27 %	-15,52 %	-5,76 %	-7,13 %	1,75 %	0,27 %	2,49 %	1,00 %
2014	-21,16 %	-21,97 %	13,23 %	12,06 %	-5,67 %	-6,64 %	-3,55 %	-4,55 %	-8,74 %	-9,68 %
2015	24,60 %	24,86 %	-1,10 %	-0,90 %	6,99 %	7,21 %	2,11 %	2,32 %	11,17 %	11,40 %
2016	1,01 %	0,66 %	0,27 %	-0,09 %	-1,11 %	-1,46 %	-2,13 %	-2,48 %	-2,11 %	-2,46 %
2017	-7,43 %	-8,12 %	0,73 %	-0,02 %	0,00 %	-0,75 %	0,36 %	-0,39 %	-7,02 %	-7,72 %
2018	-6,58 %	-7,58 %	2,14 %	1,04 %	2,47 %	1,37 %	-6,79 %	-7,79 %	7,55 %	6,40 %
2019	10,04 %	8,92 %	0,22 %	-0,80 %	-1,35 %	-2,35 %	-10,04 %	-10,96 %	-4,50 %	-5,47 %

Houses

Yearly changes of houses	Koivusaari		Tapiola		Urheilupuisto		Niittykumpu	
	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	-	-	-	-	3,41 %	2,17 %	13,61 %	12,24 %
2011	-	-	64,80 %	59,28 %	-0,72 %	-4,05 %	-5,75 %	-8,90 %
2012	-15,49 %	-17,80 %	-22,74 %	-24,86 %	19,83 %	16,56 %	29,63 %	26,09 %
2013	-	-	15,81 %	14,13 %	-15,71 %	-16,94 %	-	-
2014	-	-	-	-	9,65 %	8,52 %	-	-
2015	-	-	-	-	-10,88 %	-10,70 %	1,64 %	1,85 %
2016	-	-	-	-	3,65 %	3,29 %	-	-
2017	-	-	-	-	5,82 %	5,03 %	-	-
2018	-	-	-	-	-1,72 %	-2,77 %	38,39 %	36,90 %
2019	-3,26 %	-4,24 %	49,27 %	47,76 %	13,60 %	12,45 %	-4,33 %	-5,30 %

Yearly changes of houses	Matinkylä		Soukka		Espoonlahti		Kivenlahti	
	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change	Nominal change	Real change
2010	7,60 %	6,31 %	16,52 %	15,12 %	21,14 %	19,68 %	-	-
2011	8,15 %	4,53 %	-16,34 %	-19,14 %	-5,06 %	-8,24 %	-	-
2012	5,50 %	2,62 %	14,74 %	11,61 %	-8,77 %	-11,26 %	-	-
2013	-22,15 %	-23,28 %	-13,18 %	-14,45 %	13,73 %	12,08 %	-39,87 %	-40,74 %
2014	27,61 %	26,30 %	8,46 %	7,34 %	2,95 %	1,89 %	66,13 %	64,42 %
2015	-	-	1,26 %	1,47 %	-2,54 %	-2,34 %	-17,53 %	-17,36 %
2016	-	-	-1,15 %	-1,50 %	-2,14 %	-2,49 %	19,89 %	19,47 %
2017	19,82 %	18,92 %	13,28 %	12,43 %	-1,21 %	-1,95 %	-30,10 %	-30,62 %
2018	-8,66 %	-9,64 %	-7,14 %	-8,13 %	-4,22 %	-5,25 %	5,81 %	4,68 %
2019	4,43 %	3,37 %	-4,87 %	-5,84 %	-42,55 %	-43,14 %	-	-

Appendix 2 Correlation matrix

	Log ppsm	D_apt. buildings	D_houses	D_espoonlahti	D_kivenlahti	Log dist. cc	Log dist. ms	Log squarem	Log year built
Log ppsm	1,00	-0,10	0,07	0,15	0,12	0,13	-0,11	-0,36	0,77
D_apt. buildings	-0,10	1,00	-0,56	0,10	0,12	0,06	-0,27	-0,46	-0,09
D_houses	0,07	-0,56	1,00	-0,03	-0,09	-0,14	0,16	0,36	0,11
D_espoonlahti	0,15	0,10	-0,03	1,00	-0,47	-0,09	0,18	-0,08	0,34
D_kivenlahti	0,12	0,12	-0,09	-0,47	1,00	0,77	0,01	-0,09	0,06
Log dist. cc	0,13	0,06	-0,14	-0,09	0,77	1,00	0,29	-0,07	0,09
Log dist. ms	-0,11	-0,27	0,16	0,18	0,01	0,29	1,00	0,12	-0,09
Log squarem	-0,36	-0,46	0,36	-0,08	-0,09	-0,07	0,12	1,00	-0,05
Log year built	0,77	-0,09	0,11	0,34	0,06	0,09	-0,09	-0,05	1,00
Log floor	0,10	0,33	-0,18	0,01	0,04	-0,02	-0,21	-0,13	0,06
Rooms	-0,30	-0,45	0,35	-0,05	-0,09	-0,05	0,15	0,94	0,00
Log date of sale	0,23	-0,05	0,02	0,00	0,02	0,04	0,00	0,03	0,06
Log cpi	0,22	-0,05	0,02	0,01	0,02	0,04	0,01	0,04	0,07
Employment	0,11	-0,05	0,03	-0,04	0,04	0,05	0,02	0,05	0,02
Euribor 3 months	-0,19	0,05	-0,02	0,00	-0,03	-0,05	-0,02	-0,04	-0,04
D_stage_decision	0,06	0,03	-0,02	0,07	-0,07	-0,08	-0,03	-0,02	0,07
D_stage_construction	0,12	-0,06	0,03	-0,04	0,06	0,09	0,05	0,05	-0,01
Site ownership	0,01	0,08	-0,02	0,04	0,02	0,01	-0,03	-0,04	-0,01

	Log floor	Rooms	Log date of sale	Log cpi	Employment	Euribor 3 months	D_stage_decision	D_stage_construction	Site ownership
Log ppsm	0,10	-0,30	0,23	0,22	0,11	-0,19	0,06	0,12	0,01
D_apt. buildings	0,33	-0,45	-0,05	-0,05	-0,05	0,05	0,03	-0,06	0,08
D_houses	-0,18	0,35	0,02	0,02	0,03	-0,02	-0,02	0,03	-0,02
D_espoonlahti	0,01	-0,05	0,00	0,01	-0,04	0,00	0,07	-0,04	0,04
D_kivenlahti	0,04	-0,09	0,02	0,02	0,04	-0,03	-0,07	0,06	0,02
Log dist. cc	-0,02	-0,05	0,04	0,04	0,05	-0,05	-0,08	0,09	0,01
Log dist. ms	-0,21	0,15	0,00	0,01	0,02	-0,02	-0,03	0,05	-0,03
Log squarem	-0,13	0,94	0,03	0,04	0,05	-0,04	-0,02	0,05	-0,04
Log year built	0,06	0,00	0,06	0,07	0,02	-0,04	0,07	-0,01	-0,01
Log floor	1,00	-0,13	-0,02	-0,02	-0,02	0,02	0,02	-0,04	0,02
Rooms	-0,13	1,00	0,03	0,04	0,05	-0,05	-0,02	0,06	-0,03
Log date of sale	-0,02	0,03	1,00	0,95	0,45	-0,87	0,06	0,71	0,02
Log cpi	-0,02	0,04	0,95	1,00	0,52	-0,84	0,14	0,70	0,02
Employment	-0,02	0,05	0,45	0,52	1,00	-0,42	-0,19	0,46	0,01
Euribor 3 months	0,02	-0,05	-0,87	-0,84	-0,42	1,00	-0,05	-0,78	-0,04
D_stage_decision	0,02	-0,02	0,06	0,14	-0,19	-0,05	1,00	-0,50	0,00
D_stage_construction	-0,04	0,06	0,71	0,70	0,46	-0,78	-0,50	1,00	0,03
Site ownership	0,02	-0,03	0,02	0,02	0,01	-0,04	0,00	0,03	1,00

Appendix 3 Matlab code for linear and quadratic stepwise model for all house types

```
data=thesisdata;
ppsm=data(:,11);
D_ab=data(:,1);
D_h=data(:,2);
D_espoonlahti=data(:,3);
D_kivenlahti=data(:,4);
dist_cc=data(:,5);
dist_ms=data(:,6);
squarem=data(:,7);
year=data(:,8);
floor=data(:,9);
rooms=data(:,10);
date=data(:,12);
cpi=data(:,13);
employment=data(:,14);
euribor=data(:,15);
D_stage_decision=data(:,16);
D_stage_construction=data(:,17);
site=data(:,18);

%log value
ln_ppsm=log(ppsm);
ln_cpi=log(cpi);
ln_date=log(date);
ln_dist_cc=log(dist_cc);
ln_dist_ms=log(dist_ms);
ln_squarem=log(squarem);
ln_year=log(year);
ln_date=log(date);
ln_floor=log(floor);
```

```
ln_var=[D_ab, D_h, D_espoonlahti, D_kivenlahti, ln_dist_cc, ln_dist_ms, ln_squarem, ln_year,ln_floor, rooms, ln_date, ln_cpi, employment, euribor, D_stage_decision,D_stage_construction, site];

scatter(squarem,ppsm), hold on
refline
xlabel('Square meters')
ylabel('Price per square meter')
scatter(ln_squarem, ln_ppsm), hold on
refline
xlabel('Square meters (log)')
ylabel('Price per square meter (log)')

histfit(ppsm)
ylabel('Frequency')
xlabel('Price per square meter')
histfit(ln_ppsm)
ylabel('Frequency')
xlabel('Price per square meter (log)')

%Correlation matrix and multicollinearity
all_var=[ln_ppsm,D_ab, D_h, D_espoonlahti, D_kivenlahti, ln_dist_cc, ln_dist_ms, ln_squarem, ln_year,ln_floor, rooms, ln_date, ln_cpi, employment, euribor, D_stage_decision,D_stage_construction, site];
R0=corrcoef(all_var)
format bank
all_independent=[D_ab, D_h, D_espoonlahti, D_kivenlahti, ln_dist_cc, ln_dist_ms, ln_squarem, ln_year,ln_floor, rooms, ln_date, ln_cpi, employment, euribor, D_stage_decision,D_stage_construction, site];
R1=corrcoef(all_independent);
VIF=diag(inv(R1))'

%Dividing data into training and test sets
rng('default')
```

```
Y=ln_ppsm;
X=double(ln_var(:,1:end));
cv=cvpartition(length(ln_var),'holdout',0.30);
% Training set
Xtrain=X(training(cv),:);
Ytrain=Y(training(cv),:);
% Test set
Xtest=X(test(cv),:);
Ytest=Y(test(cv),:);
format bank
format long

md1_linear=fitlm(Xtrain,Ytrain)

md1=stepwiselm(Xtrain,Ytrain,'quadratic')

plot(md1)
residuals=md1.Residuals.Raw;
mdRES=fitlm(Xtrain,residuals,'constant')
[p,F,r]=coefTest(mdRES,[1])

%checking heteroscedasticity
plotResiduals(md1,'fitted')

%testing autocorrelation
%Durbin-Watson test
plotResiduals(md1,'lagged')
plotResiduals(md1,'caseorder','linestyle','-')
[P,DW]=dwtest(md1)

%testing normality
[h,p,jbstat,critval]=jbtest(residuals,0.05)
residuals_stand=md1.Residuals.Standardized;
histfit(residuals_stand)
```

```
title('All house types')
ylabel('Frequency')
xlabel('Standardized residuals')

%testing the model
ypred2=predict(md1,Xtest);
logresults=[ypred2,Ytest];
format long
MSEtest=immse(Ytest,ypred2)
rmse=sqrt(MSEtest)

%back to functional form from log
format short
ypred2norm=exp(ypred2);
Ytestnorm=exp(Ytest);
results=[ypred2norm, Ytestnorm];
Xtest2=[Xtest(:,1:4), exp(Xtest(:,5:9)),Xtest(:,10),exp(Xtest(:,11:12)),Xtest(:,13:17)];
format bank
MSEtestnorm=immse(Ytestnorm,ypred2norm)
rmse=sqrt(MSEtestnorm)
```

Appendix 4 Output of linear model and quadratic stepwise model for all house types

	All house types		*Quadratic stepwise model All house types*	
Estimated Coefficients:	**Coeff**	**SE**	**Coeff**	**SE**
(Intercept)	-259,83 ***	4,74	-18785,58 ***	4466,46
x1 D_Apartment buildings	-0,22 ***	0,01	-195,30 ***	24,26
x2 D_Houses	0,07 ***	0,02	74,88 **	29,41
x3 D_espoonlahti	-0,10 ***	0,01	-90,88 ***	10,94
x4 D_kivenlahti	-0,08 ***	0,02	-61,03 ***	5,33
x5 Log distance to city center	1,10 ***	0,18	-18,13 ***	2,54
x6 Log distance to metro station	-0,02 ***	0,01	21,00 **	8,84
x7 Log square meters	-0,41 ***	0,03	-207,59 ***	12,30
x8 Log year built	33,88 ***	0,52	5025,07 ***	1173,61
x9 Log floor	0,04 ***	0,00	-12,49 ***	4,40
x10 Rooms	0,01	0,01	8,48 ***	2,03
x11 Log date of sale	0,04 ***	0,02	-0,41 ***	0,14
x12 Log CPI	0,29	0,39	12,26 ***	2,83
x13 Employment	0,31	0,23	342,05 **	160,19
x14 Euribor 3 months	-3,57 **	1,61	-832,10 ***	192,46
x15 D_stage_decision	-0,02	0,02	-42,02 **	18,85
x16 D_stage_construction	-0,04 *	0,03	1,27 ***	0,38
x17 Site	0,07	0,06	0,14 **	0,07
x1:x4			-0,11 ***	0,04
x1:x5			1,69 ***	0,46
x1:x6			-0,27 ***	0,04
x1:x8			23,68 ***	3,17
x1:x13			0,87 *	0,50
x2:x4			-0,31 ***	0,06
x2:x5			1,30 ***	0,49
x2:x7			-0,18 ***	0,05
x2:x8			-9,85 ***	3,77
x2:x9			-0,10 ***	0,03
x2:x12			-2,51 ***	0,80
x2:x15			0,19 ***	0,06
x2:x16			0,27 ***	0,07
x3:x6			0,07 ***	0,03
x3:x8			11,91 ***	1,43
x3:x10			-0,05 ***	0,01
x3:x15			0,05 ***	0,01

x4:x5		6,37 ***	0,53
x4:x6		0,23 ***	0,02
x4:x7		-0,19 ***	0,03
x4:x12		-0,47 *	0,26
x4:x14		-5,35 ***	2,02
x4:x15		0,05 ***	0,02
x5:x7		4,66 ***	0,68
x5:x10		-0,79 ***	0,21
x5:x14		84,75 ***	19,69
x6:x8		-2,67 **	1,16
x6:x9		-0,02 ***	0,01
x6:x16		-0,04 ***	0,01
x7:x8		22,55 ***	1,30
x7:x11		0,11 ***	0,03
x7:x12		-2,89 ***	0,67
x7:x13		3,75 ***	1,20
x7:x16		-0,07 ***	0,02
x8:x9		1,66 ***	0,58
x8:x13		-51,97 **	20,97
x10:x13		-0,98 **	0,38
x10:x14		1,47 *	0,77
x11:x15		-0,58 **	0,24
x12:x15		9,14 **	4,09
x13:x16		-1,23 **	0,53
x14:x16		16,12 *	8,47
x14:x17		-16,15 **	7,90
x6^2		-0,04 ***	0,01
x7^2		0,14 ***	0,02
x8^2		-333,34 ***	77,21
x9^2		0,02 ***	0,01
x13^2		28,63 ***	9,03
x14^2		571,39 **	276,71
Number of observations:	2243	2243	
Root Mean Squared Error:	0,147	0,116	
R-squared:	0,794	0,873	
Adjusted R-Squared	0,792	0,869	
F-statistic vs, constant model:	503	227	
Durbin-Watson	2,03	2,01	

Dependent variable: Price per square meter (log)
****** Significant on 1 % level ** Significant on 5 % level***
**** Significant on 10 % level***

Appendix 5 Output of linear model and quadratic stepwise model for apartment buildings and terrace houses

	Apt. buildings and terrace houses		*Quadratic stepwise model Apt. buildings and terrace houses*	
Estimated Coefficients:	**Coeff**	**SE**	**Coeff**	**SE**
(Intercept)	-265,34 ***	4,77	-20765,76 ***	4831,53
x1 D_Apartment buildings	-0,22 ***	0,01	-181,81 ***	23,06
x2 D_espoonlahti	-0,11 ***	0,01	-118,34 ***	13,33
x3 D_kivenlahti	-0,07 ***	0,02	-50,71 ***	4,96
x4 Log distance to city center	1,14 ***	0,18	-499,71 ***	136,07
x5 Log distance to metro station	-0,02 ***	0,01	21,65 **	9,55
x6 Log square meters	-0,40 ***	0,03	-260,57 ***	25,40
x7 Log year built	34,74 ***	0,53	6158,84 ***	1295,51
x8 Log floor	0,04 ***	0,00	0,39 **	0,16
x9 Rooms	0,01	0,01	26,01 ***	8,62
x10 Log date of sale	0,06 ***	0,01	5,75 ***	1,82
x11 Log CPI	-0,05	0,37	14,52 ***	2,82
x12 Employment	0,45 **	0,22	604,34 ***	161,47
x13Euribor 12 months	-3,41 **	1,53	-7,04 ***	2,35
x14 D_stage_decision	-0,02	0,02	0,81	0,56
x15 D_stage_construction	-0,03	0,02	1,66 **	0,72
x16 Site	0,15 ***	0,05	0,01	0,05
x1:x3			-0,11 **	0,04
x1:x4			1,85 ***	0,44
x1:x5			-0,27 ***	0,04
x1:x7			22,08 ***	2,99
x1:x11			-0,49 *	0,27
x2:x5			0,09 ***	0,03
x2:x6			-0,16 ***	0,02
x2:x7			14,88 ***	1,74
x2:x11			1,14 ***	0,22
x2:x15			-0,07 ***	0,02
x3:x4			5,08 ***	0,50
x3:x5			0,23 ***	0,02

x3:x6		-0,20 ***	0,04
x3:x10		0,05 ***	0,02
x3:x15		-0,04 **	0,02
x4:x6		3,76 ***	0,70
x4:x7		64,16 ***	17,86
x4:x9		-0,50 **	0,21
x4:x10		-0,74 ***	0,18
x5:x7		-2,52 **	1,26
x5:x8		-0,02 ***	0,01
x5:x11		-0,39 **	0,15
x5:x14		0,028 **	0,01
x6:x7		31,06 ***	3,23
x6:x10		0,13 ***	0,03
x6:x11		-2,97 ***	0,62
x7:x9		-2,78 **	1,12
x7:x12		-79,87 ***	21,27
x8:x10		0,02 **	0,01
x8:x12		-0,41 *	0,23
x9:x13		2,15 ***	0,79
x9:x15		-0,02 **	0,01
x10:x12		1,56 ***	0,59
x12:x14		-1,46 *	0,80
x12:x15		-2,70 ***	1,05
x15:x16		0,26 ***	0,09
x5^2		-0,04 ***	0,01
x6^2		0,16 ***	0,02
x7^2		-449,80 ***	88,35
x8^2		0,02 ***	0,00
Number of observations:	2152	2152	
Root Mean Squared Error:	0,14	0,111	
R-squared:	0,815	0,886	
Adjusted R-Squared	0,813	0,883	
F-statistic vs, constant model:	587	291	
Durbin-Watson	1,98	1,97	

Dependent variable: Price per square meter (log)

****** Significant on 1 % level*** ***** Significant on 5 % level***

**** Significant on 10 % level***

Appendix 6 Output of quadratic stepwise model for all house types using Euribor 12 months in the dataset

	Quadratic stepwise model Euribor 12 months	
Estimated Coefficients:	**Coeff**	**SE**
(Intercept)	-16670,09 ***	4625,24
x1 D_Apartment buildings	-190,96 ***	24,18
x2 D_House	68,01 **	29,84
x3 D_espoonlahti	-102,57 ***	12,56
x4 D_kivenlahti	-57,89 ***	5,64
x5 Log distance to city center	-304,37 **	134,81
x6 Log distance to metro station	27,82 ***	9,54
x7 Log square meters	-209,68 ***	12,43
x8 Log year built	5461,92 ***	1204,30
x9 Log floor	-12,25 ***	4,42
x10 Rooms	8,58 ***	2,04
x11 Log date of sale	55,66 ***	18,95
x12 Log CPI	-1052,06 **	413,08
x13 Employment	460,84 ***	160,37
x14 Euribor 12 months	2343,87 *	1272,14
x15 D_stage_decision	-0,54 ***	0,20
x16 D_stage_construction	58,45 ***	20,87
x17 Site	0,18 **	0,08
x1:x4	-0,10 **	0,04
x1:x5	1,71 ***	0,46
x1:x6	-0,26 ***	0,04
x1:x8	23,16 ***	3,17
x2:x4	-0,29 ***	0,06
x2:x5	1,14 **	0,50
x2:x7	-0,17 ***	0,05
x2:x8	-8,55 **	3,85
x2:x9	-0,10 ***	0,03
x2:x12	-2,83 ***	0,78
x2:x15	0,21 ***	0,06
x2:x16	0,28 ***	0,07
x3:x6	0,09 ***	0,03
x3:x8	13,45 ***	1,64
x3:x10	-0,05 ***	0,01

x3:x14	-4,25 ***	1,37
x3:x16	-0,06 ***	0,02
x4:x5	6,35 ***	0,53
x4:x6	0,23 ***	0,02
x4:x7	-0,19 ***	0,03
x4:x12	-1,08 **	0,43
x4:x14	-11,12 ***	2,60
x4:x16	-0,06 **	0,02
x5:x7	4,77 ***	0,69
x5:x8	30,04 *	17,48
x5:x10	-0,81 ***	0,21
x5:x12	11,99 **	5,60
x5:x14	132,93 ***	29,69
x6:x8	-3,57 ***	1,25
x6:x9	-0,02 ***	0,01
x6:x16	-0,04 ***	0,01
x7:x8	22,58 ***	1,31
x7:x11	0,11 ***	0,03
x7:x12	-2,62 ***	0,68
x7:x13	3,13 ***	1,19
x7:x16	-0,05 **	0,02
x8:x9	1,63 ***	0,58
x8:x13	-65,99 ***	21,08
x10:x13	-0,87 **	0,39
x10:x14	1,67 **	0,65
x11:x12	-12,28 ***	4,16
x11:x14	46,54 **	19,13
x11:x16	0,92 ***	0,32
x12:x14	-805,28 ***	275,67
x12:x16	-12,85 ***	4,57
x14:x15	42,43 **	17,26
x14:x16	69,51 ***	26,94
x14:x17	-12,80 **	6,00
x6^2	-0,04 ***	0,01
x7^2	0,14 ***	0,02
x8^2	-380,46 ***	82,40
x9^2	0,02 ***	0,01
x11^2	0,29 **	0,14
x12^2	104,22 **	44,34
x13^2	21,26 **	9,38
x14^2	1943,39 ***	652,64
Number of observations:	2243	
Root Mean Squared Error:	0,116	
R-squared:	0,874	
Adjusted R-Squared	0,869	
F-statistic vs, constant model:	206	
Durbin-Watson	2,00	

Dependent variable: Price per square meter (log)

****** Significant on 1 % level ** Significant on 5 % level***

**** Significant on 10 % level***

Appendix 7 Output of Westmetro's phase 1's quadratic stepwise model for all house types

	Quadratic stepwise model All house types	
Estimated Coefficients:	**Coeff**	**SE**
(Intercept)	24277,67 ***	872,34
x1 D_Apartment buildings	-74,405 ***	5,89
x2 D_Houses	-2,96 ***	0,44
x3 D_lauttasaari	56,88 ***	7,12
x4 D_koivusaari	34,08 ***	11,06
x5 D_keilaniemi	181,26 ***	37,72
x6 D_otaniemi	88,70 ***	17,81
x7 D_tapiola	8,74	6,89
x8 D_niittykumpu	48,43 ***	8,06
x9 D_matinkylä	-30,02 ***	10,30
x10 Log distance to city center	-13,30 ***	3,16
x11 Log distance to metro station	10,53 ***	1,86
x12 Log square meters	-49,83 ***	4,66
x13 Log year built	-6135,82 ***	203,01
x14 Log floor	-7,52 ***	1,30
x15 Rooms	8,07 ***	1,81
x16 Shopping center	13,61 *	7,51
x17 Log date of sale	42,82 ***	13,50
x18 Log CPI	-392,32 **	169,28
x19 Employment	90,83 **	44,32
x20 Euribor 3 months	-873,73 ***	73,60
x21 D_stage_decision	-153,52 ***	53,79
x22 D_stage_construction	-116,26 ***	20,52
x23 D_stage_operating	-114,95 ***	24,44
x24 Site	6,79 ***	2,03
x1:x3	0,43 ***	0,10
x1:x4	0,31 ***	0,07
x1:x8	-0,06 **	0,03
x1:x9	-0,32 ***	0,05
x1:x10	0,46 ***	0,10
x1:x11	0,04 **	0,02
x1:x12	0,15 ***	0,02
x1:x13	8,94 ***	0,77
x1:x14	0,03 ***	0,01
x1:x16	0,12 ***	0,03
x1:x18	0,36 ***	0,13
x1:x19	-0,67 *	0,38
x1:x22	-0,03 *	0,02
x2:x3	0,22 *	0,12
x2:x10	0,29 ***	0,05
x2:x17	0,069 ***	0,03
x2:x20	4,33 ***	1,55
x3:x10	1,50 ***	0,38
x3:x11	-0,22 ***	0,05
x3:x12	-0,09 ***	0,03
x3:x13	-9,45 ***	0,74
x3:x15	0,08 ***	0,01
x3:x18	0,58 ***	0,12
x3:x19	0,68 ***	0,22
x3:x20	0,66 *	0,36
x3:x21	0,07 **	0,03
x3:x22	0,03 **	0,01
x4:x10	1,43 ***	0,30
x4:x11	-0,09 ***	0,04
x4:x13	-6,22 ***	1,41
x4:x14	-0,03 ***	0,01
x4:x15	0,06 ***	0,01
x4:x18	0,23 **	0,11
x4:x21	0,06 **	0,03

x5:x10	-2,33 ***	0,36
x5:x12	-0,08 *	0,04
x5:x13	-21,58 ***	4,73
x5:x15	0,04 ***	0,02
x5:x18	0,66 ***	0,14
x5:x19	0,99 **	0,40
x6:x12	-0,06 **	0,02
x6:x13	-12,26 ***	2,36
x6:x18	0,96 ***	0,18
x6:x21	0,06 *	0,03
x6:x22	0,06 ***	0,02
x7:x10	0,67 ***	0,19
x7:x11	-0,10 ***	0,01
x7:x13	-1,85 **	0,77
x7:x22	0,03 ***	0,01
x8:x10	-0,68 ***	0,23
x8:x11	-0,04 **	0,02
x8:x12	-0,13 ***	0,03
x8:x13	-6,04 ***	0,88
x8:x14	-0,02 ***	0,01
x8:x15	0,03 **	0,01
x8:x18	0,83 ***	0,21
x8:x19	0,67 *	0,36
x8:x22	-0,07 ***	0,02
x8:x23	-0,09 ***	0,03
x8:x24	0,09 ***	0,03
x9:x10	-0,82 ***	0,26
x9:x11	-0,06 **	0,02
x9:x12	-0,18 ***	0,02
x9:x13	8,28 ***	0,96
x9:x14	-0,02 ***	0,00
x9:x18	-5,20 ***	1,23
x9:x19	1,66 **	0,73
x9:x20	-42,05 **	19,72
x9:x23	-0,03 *	0,01
x10:x11	-0,17 ***	0,05
x10:x12	-0,18 ***	0,03
x10:x15	0,07 ***	0,01
x10:x16	-0,63 ***	0,15
x10:x21	0,09 ***	0,03
x10:x22	0,04 ***	0,01
x10:x24	-0,07 ***	0,03
x11:x12	-0,027 **	0,01
x11:x13	-1,11 ***	0,23
x11:x14	-0,01 **	0,00
x11:x15	0,01 **	0,01
x11:x16	0,07 ***	0,02
x11:x19	-0,33 **	0,16
x11:x21	-0,04 ***	0,01
x11:x22	-0,03 ***	0,01
x11:x23	-0,03 ***	0,01
x11:x24	-0,1 ***	0,03
x12:x13	6,67 ***	0,63
x12:x14	0,05 ***	0,01
x12:x15	0,04 **	0,02
x12:x17	-0,05 ***	0,01
x12:x20	1,87 ***	0,40
x12:x22	0,04 ***	0,01
x12:x23	0,08 ***	0,02
x13:x14	0,97 ***	0,17
x13:x15	-1,02 ***	0,24
x13:x16	-4,39 ***	0,60
x13:x20	103,68 ***	9,59
x13:x21	2,21 ***	0,51
x13:x22	1,93 ***	0,26

x14:x15	-0,01 *	0,00
x14:x17	0,01 ***	0,00
x14:x20	0,68 ***	0,25
x14:x24	-0,03 ***	0,01
x15:x16	0,01 **	0,01
x15:x17	0,02 ***	0,00
x15:x18	-0,26 ***	0,06
x16:x18	5,47 ***	1,23
x16:x19	-1,82 **	0,73
x16:x20	41,93 ***	19,72
x16:x24	0,07 **	0,03
x17:x18	-9,56 ***	2,91
x17:x19	1,93 **	0,86
x17:x21	-2,01 ***	0,35
x17:x22	-1,55 ***	0,35
x17:x23	-1,79 ***	0,52
x17:x24	0,08 **	0,04
x18:x19	-18,38 *	9,75
x18:x21	30,94 ***	11,74
x18:x22	22,80 ***	4,58
x18:x23	25,84 ***	5,53
x18:x24	-1,20 ***	0,45
x19:x21	-9,95 ***	1,90
x19:x22	-7,11 ***	1,33
x19:x23	-7,24 ***	1,49
x20:x21	41,53 ***	5,92
x20:x22	68,42 ***	9,66
x20:x23	69,09 ***	16,27
x20:x24	2,84 ***	1,01
x10^2	0,79 ***	0,18
x11^2	-0,01 ***	0,00
x12^2	0,07 ***	0,02
x13^2	403,36 ***	13,38
x14^2	0,02 ***	0,00
x15^2	-0,01 ***	0,00
x17^2	0,46 ***	0,12
x18^2	43,65 **	18,25
x20^2	440,53 ***	97,46
Number of observations:	10890	
Root Mean Squared Error:	0,102	
R-squared:	0,893	
Adjusted R-Squared	0,891	
F-statistic vs, constant model:	227	
Dependent variable: Price per square meter (log)		
****** Significant on 1 % level ** Significant on 5 % level***		
**** Significant on 10 % level***		

Appendix 8 Matlab code for modifications of predicted values of macroeconomic variables

Code for Bank of Finland’s December 2020 predictions:

```
%predictions for year 2023 based on Bank of Finland's predictions 2020-2023
format long
employment_y=[0.716, 0.712, 0.713, 0.717]';
euribor_y=[-0.004, -0.005, -0.005, -0.005]';
cpi_y=[123.66, 124.61, 126.05, 127.96]';

%transforming yearly predictions (2020-2023) into quarterly
format bank
t_y=[45,49,53,57]';
t_q=[45:60]';
empolyment_q=spline(t_y, employment_y, t_q)
euribor_q=spline(t_y, euribor_y, t_q)
format bank
cpi_q=spline(t_y, cpi_y, t_q)
```

Code for Bank of Finland’s December 2019 predictions:

```
%predictions for year 2023 based on Bank of Finland's predictions 2020-2022
format long
t=[45,49,53]';
employment_y=[0.727, 0.73, 0.734]';
t2=[45,49,53,57]';
employment_pred=spline(t, employment_y, t2)

euribor_y=[-0.004, -0.004, -0.003]';
euribor_pred=spline(t, euribor_y, t2)

cpi_y=[124.8516941, 126.5227493, 128.4325266]';
cpi_pred=spline(t, cpi_y, t2)
```

```
%transforming yearly predictions (2020-2023) into quarterly
format long
t_y=[53,57,61,65]';
t_q=[53:68]';
empolyment_q=spline(t_y, employment_pred, t_q)
euribor_q=spline(t_y, euribor_pred, t_q)
format bank
cpi_q=spline(t_y, cpi_pred, t_q)
```