PREDICTING THE HOUSING PRICE DEVELOPMENT IN THE UPCOMING TAMPERE TRAMWAY PROJECT 2 AREAS

Lappeenranta–Lahti University of Technology LUT

Master's Programme in Strategic Finance and Analytics

Master's thesis

2024

Oskari Laamanen

Examiners: University Lecturer, Ph.D., Roman Stepanov

Post-doctoral researcher, Mariia Kozlova

ABSTRACT

Lappeenranta–Lahti University of Technology LUT
LUT School of Business and Management
Business Administration

Oskari Laamanen

Predicting the Housing Price Development in the Upcoming Tampere Tramway Project 2 Areas

Master's thesis

2024

60 pages, 6 figures, 8 tables, and 20 appendices

Examiner: University lecturer, Roman Stepanov, and Post-doctoral researcher, Mariia Kozlova

Keywords: price prediction, housing price, public transportation, Tampere tram, hedonic pricing model, regression analysis

Urbanization has been one of the megatrends during the last few decades. This has directed pressure on urban planning and efficient land use. However, innovative public transportation solutions have been found to be an effective way to develop the vitality and functionality of an urban area. Hence, a successfully executed public transportation project increases land values, consequently affecting housing prices. Tampere is one of the fastest-growing city regions in Finland. In November 2016 the city council of Tampere decided to implement its first tramway project, which first phase was completed in August 2021. The implementation of the second tramway project is planned to be voted on in October 2024.

This study aims to determine how the tram project 1 has affected the housing price development close to the tramline, and how the magnitude of the price effect has varied in different stages of the project. Based on the results of the tram project 1 areas, the housing price development of tram project 2 areas will be predicted for the end of 2026. Based on the results the housing price development in the tram project 2 areas will be predicted until the end of 2026. The research is executed by OLS regression, and the data is gathered from Hintaseurantapalvelu which is the biggest databank of realized housing transactions in Finland upheld by the Central Federation of Finnish Real Estate Agencies.

The main finding is that housing prices are expected to increase between 2024-2026 in the areas that are chosen to represent the tram project 2 areas in this research. Also, the stage of the tram project 1 affected housing prices as assumed because housing prices started to increase even before the tram started to operate due to the expectation effect.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT
LUT-kauppakorkeakoulu
Kauppatieteet

Oskari Laamanen

Asuntojen Hintakehityksen Ennustaminen Tulevan Tampereen Raitiotie Projekti 2 Alueella

Kauppatieteiden Pro Gradu -tutkielma

2024

60 sivua, 6 kuvaa, 8 taulukkoa ja 20 liitettä

Tarkastaja: Yliopisto-opettaja, Roman Stepanov ja Tutkijatohtori, Mariia Kozlova

Avainsanat: hintaennuste, asuntojen hinta, julkinen liikenne, Tampereen ratikka, hedonistinen hinnoittelu malli, regressioanalyysi

Kaupungistuminen on ollut yksi viime vuosikymmenten suurista megatrendeistä. Tämä on luonut painetta kaupunkisuunnitteluun ja tehokkaaseen maankäyttöön. Innovatiivisten julkisen liikenteen ratkaisujen on kuitenkin huomattu olevan tehokas tapa kehittää kaupunkiseudun elinvoimaisuutta ja toimivuutta. Näin ollen onnistuneesti toteutettu julkisen liikenteen hanke nostaa maan arvoa ja sitä kautta myös asuntojen hintoja. Tampere on Suomen yksi nopeimmin kasvavista kaupunkiseuduista. Marraskuussa 2016 Tampereen kaupungin valtuusto päätti toteuttaa kaupungin ensimmäisen raitiotiehankkeen, jonka ensimmäinen vaihe valmistui elokuussa 2021. Toisen raitiotiehankkeen toteutuksesta on määrä äänestää lokakuussa 2024.

Tämän tutkimuksen tavoitteena on selvittää kuinka raitiotiehanke 1 on vaikuttanut asuntojen hintakehitykseen lähellä raitiolinjaa, ja kuinka hintavaikutus on vaihdellut hankkeen eri vaiheissa. Tulosten pohjalta asuntojen hintakehitys raitiotiehanke 2 alueilla ennustetaan vuoteen 2026 asti. Tutkimus toteutetaan OLS regressiolla, ja tutkimuksessa käytettävä data on peräisin Hintaseurantapalvelusta, joka on Suomen suurin toteutuneiden asuntokauppojen tietokanta ja jota ylläpitää Kiinteistönvälitysalan Keskusliitto.

Tutkimuksen keskeisin havainto on, että asuntojen hintojen odotetaan nousevan vuosina 2024-2026 alueilla, jotka edustavat tässä tutkimuksessa raitiotiehanke 2 alueita. Myös ensimmäisen raitiotiehankeen eri vaiheet ovat vaikuttaneet asuntojen hintoihin odotetusti, sillä asuntojen hinnat alkoivat nousta jo ennen raitiovaunun liikennöinnin alkamista odotusvaikutuksen seurauksena.

Abbreviations

BRT	Bus Rapid Transit
CAR	Cumulative Absolute Error
CBD	Central Business District
CLRM	Classical Linear Regression Model
CPI	Consumer Price Index
MAE	Mean Average Error
OLS	Ordinary Least Squares
RB	Regular Bus

Table of contents

List of appendices

List of Figures

List of tables

1 Introduction

As a result of urbanization, the populations of large cities continue to grow. The increased population density causes traffic congestion, scarcity of urban land resources, and housing shortage (Wei, Zhang, Wen 2020). Transit-oriented development has been found to be an effective strategy to minimize these problems by enhancing the link between transportation and land use (Gao, Chong, Zhang, Li 2023). While public transportation improvements increase land use efficiency by improving the level of accessibility, land values tend to rise (Hennebery 1998). As a result, the increased land values are capitalized into increased values of real estate properties (Chalermpong 2007). Therefore, a successful public transportation project can be expected to affect housing prices. As the construction of new infrastructure often lasts for years, it must be considered that the expected improvements affect housing prices even before the construction is completed (Yiu, Wong 2005).

Tampere is one of the fastest growing city regions in Finland and for a long time, public transportation in the city relied entirely on bus connections. However, this was about to change when one of the largest public transportation projects in Finland of the last few years, the Tampere tramway, was gaining momentum. The implementation planning of the project took place in 2015 and 2016, and the main objectives were to streamline traveling, support the growth of the urban region, and increase the city's attractiveness (Tampereen Ratikka 2023a). On the 7th of November 2016, Tampere City Council voted in favour of the construction of the tramway (Tampereen Ratikka 2023b). In the approved plan the tramway was decided to be implemented in two phases. In Figure 1 the first phase is represented as a blue line and the second phase is represented as a grey line. The construction of the first phase took place in the years 2017-2021, and it was completed before its initial schedule. Tampere Tram started operating on the 9th of August in 2021. (Raitiotiealliansi 2023a) The construction of the second phase started in 2020 and it is planned to be completed in 2025 (Raitiotiealliansi 2023b).

Figure 1: Route presentation of tram project 1 and 2

In November 2021, a few months after the tramway started operating, the City Councils of Tampere and the neighboring city Pirkkala decided to start preliminary plans for another tram project that would continue the rails to Pirkkala and Linnainmaa (Tampereen Ratikka 2023c). The detailed implementation plans started on the 24th of April in 2023, and both City Councils are scheduled to vote for the project implementation in October 2024. If the project receives the support of both City Councils, the construction will occur in 2025-2028 (Tampereen Ratikka 2023d). In Figure 1, the tram project 2 is represented as a red line.

While increased accessibility through public transportation improves both economic and social activity, it also affects housing prices. The location of an apartment is unchangeable and therefore, the living environment and level of accessibility are not easily changeable either. This means that in addition to an apartment's structural features, people are also willing to pay for enjoyability and advanced accessibility of a neighborhood. The Tampere Tramway might influence both aspects, especially the latter.

1.1 Objectives and Research Questions

This study aims to predict the housing price development in the areas of the presumably upcoming Tampere second tramway project in 2024-2028. Assuming that the second project will be accepted by the city councils, and will be completed in time, the research window contains; the official decision of the tram project 2 implementation, construction period, and completion of the project. While most of Tampere's first tramway project is already completed, the impact of the tramway project on the housing price development in those areas is mainly known. Therefore, the housing price development in the tram project 1 areas can be used as a base for predicting the price development in the tram project 2 areas. The main research question is:

How can the housing price development be predicted most accurately in the Tampere tramway project 2 areas?

Previous research is relatively unanimous that there is a positive correlation between public transportation investments and housing prices. However, there is no clear view of what exact explanatory variables should be included in the model. In addition, most of the studies have concentrated on explanatory research while this study is about prediction. This causes a time-related issue when the period that is being predicted (prediction period) is different from the period on which basis the prices are predicted. To minimize the time-related prediction error, it must be defined whether macroeconomic conditions affect housing prices. In addition, to predict the housing price development during the tram project 2, it is also reasonable to investigate how the tram project 1 has affected housing prices in different stages of the project. In this research, the change in house prices caused by public transportation improvements is called the housing price effect. To support the main research questions two sub-questions go as follows:

What macroeconomic variables should be included in the predictive model?

How has the housing price effect of the tram line varied in different stages of the tram project 1?

1.2 Thesis Structure

The first section of the thesis has served as an introduction to the topic. In this section, the motivation, purpose, and goals of the thesis are explained, and the research questions are presented. In the second chapter, the main findings of previous research in the field of transportation infrastructure improvements, and housing prices are explained. The theoretical framework of the thesis is the hedonic pricing model, and the analysis part will be done by the OLS regression method. Those tools are also explained in chapter two. Chapter 3 introduces the data used in the analysis and goes through the data handling process. The analysis part is done in chapter 4. It starts by testing the assumptions of the OLS regression, after which the final functional form is chosen, and at the end, the regression is run. In this chapter also the predictive model, which is created based on the regression, is explained. The results of the prediction model are presented and analyzed in Chapter 5. Chapter 6 combines the main findings of the thesis and discusses the limitations of the research.

2 Literature Review

Public transportation infrastructure is known to have strong historical evidence in shaping the pattern of urban development (Hennebery 1998; Dorantes, Paez, Vassallo 2011). Better accessibility increases economic activity by bringing consumption possibilities closer and enabling employers and employees to find each other. Rode, Floater, Thomopoulos, Docherty, Schwinger, Mahendra, and Fang (2014) estimated that 1 billion US dollars of spending on public transportation indirectly creates $3,6 billion of output and generates over 36,000 new jobs. In the long run, the increased social, technological, and economic activity caused by new public transportation improvements impact house prices as fundamentally as the improved accessibility does in the short run (Henneberry 1998). In line with the above findings, it is the argument of Zhou and Yang (2021) that the urban development of a city can be predicted by transit-based accessibility.

2.1 Housing Price Formation

Structural characteristics like square meters, floor plan, condition, and house type play a key role in determining house prices. Still, a dwelling is a heterogeneous good whose price is not determined only by its structural characteristics but also by the surrounding environment and the level of accessibility (Coppola, Cordera, dell'Olio & Ibeas 2019). This is because the location is immobile determining the neighborhood and the proximity to everyday activities. Because the living space is limited in cities, when a certain location becomes desirable, the demand increases which pushes the house prices up (Debrezion, Pels, Rietveld 2007). According to land rent theory, one of the key factors that drive the demand for houses in a neighborhood is accessibility to goods and services. When accessibility increases, it affects positively the land rents and the land values. (Mulley 2014) Thus, higher land values raise house prices.

In cities, services are typically concentrated in central business districts (CBD). Therefore, the proximity of CBD increases the accessibility to services which is reflected in land values.

Because there is limited space close to the CBD, the accessibility of areas further away from the CBD can be increased by successful urban planning and well-functioning public transportation. The accessibility in the areas along transport connections makes it easier to reach the services of the CBD, but the improved transportation also attracts retailers to bring the services into those areas (Bowes & Ihlanfeldt 2001).

Public transportation investments are typically financed by public money. The desire of public authorities to finance public transportation projects stems from the desire to promote the functionality and vitality of the city. There are three main ways to cover the cost of a public transportation project, user fees, tax increases, and cuts in other public spending. The latter two are typically more significant in monetary terms. (Gihring & Smith 2006) While the accessibility benefits from this kind of project are distributed asymmetrically within the city, it is reasonable to seek to allocate the costs to those who benefit from the project. To do so, public authority can use value capturing in which the increase of land values will be levied by increasing property taxes (Gihring & Smith 2006).

2.2 The Impact of Public Transportation on Housing Prices

The relationship between public transportation infrastructure and housing prices is widely researched in academic literature. A city is always a combination of different factors, such as population density, geographical location, income level, and cultural elements. While all those elements affect the magnitude of the housing price effect caused by public transportation, the research outcomes of the study field vary. Most of the literature has found a positive correlation between public transportation and housing prices. Dorantes et al. (2011) researched the price effect of the Metrosur metro line which connects five Madrid urban region municipalities, and the Cernicas metro line which connects the municipalities to the city of Madrid. They estimated that a house right next to a Metrosur station is from 2.18 % to 3.18 % higher valued than a house located 1000 meters from the station. The impact of Cernicas is even greater since the value of a house right next to a Cernicas station is from 3.38 % to 5.17 % higher than a house located 1000 meters from the station. The difference is presumably due to the connection of Cernicas to the city of Madrid. The

connection to the CBD of Madrid is the most central aspect affecting housing prices. Via Metrosur the travel cost and time to Madrid is higher than via Cernicas.

The intensity of the housing price impact caused by public transportation is dependent on the quality of the transportation infrastructure. In literature, there are several different categorization methods to define the quality of transportation infrastructure. So, Tse and Ganesan (1997) divided the impacts of transportation on housing prices into four categories, which are; the availability of transportation, travel time, transportation costs, and convenience. The main idea of the classification has been supported by several different studies. Debrezion et al. (2011), researched that a large number of destinations and transportation frequency, which both are elements of availability, affect positively housing prices. Also, So et al. (1997) emphasized that transportation frequency is an important attribute of accessibility, which they found to have a positive effect on housing prices. Ryan (1999) pointed out that when transportation facilities provide time savings, property values tend to show the theoretically expected relationship with transportation access. Coppola et al. (2019) estimated that every additional travel minute to the City of Rome decreases housing prices by 0,6 % and the corresponding effect in the City of Santander is 1,0 %. Generally, it can be said that buying a house in a city region is a trade-off between lower housing prices and transportation costs, which include travel time and transport fares. People have a higher willingness to pay for accessible locations. (Dorantes et al. 2011) However, if the improvements in public transportation are not large enough compared to the previous system the housing price effect may not occur at all (Ryan 1999).

2.2.1 Comparison Between Rail and Bus Transportation

The two most common categories of how public transportation can be implemented are bus and rail connections, and several studies have investigated the differences between them. To compare bus and rail transportation the types of bus connections must be defined because the performance of bus connections is highly dependent on the structure of public transportation networks within the city. In the context of intra-city public transportation, bus connections can be separated into regular bus connections (RB) and bus rapid transit connections (BRT). In RB systems the stop spacing is typically short, the frequency is low,

but the number of lines is high so that most of the city is covered by some kind of public transport. (Vuchic 2002) Implementing the whole public transport network on regular buses makes it inefficient. Therefore, BRT systems have gained popularity as a cost-efficient way to improve public transportation (Cervero & Kang 2011). The idea of BRT is to reorganize the bus network and create a few trunk lines that can operate with high speed, frequency, and capacity. (Wirasinghe, Kattan, Rahman, Hubbell, Thilakaratne, & Anowar 2013) In the BRT system, there are feeder lines that connect passengers to the trunk line (Vuchic 2002). The operating logic of the BRT systems is similar to the light rail transit systems (LRT) and therefore they are often seen as alternative options to each other (Beaudoin & Tyndall, 2023).

While the BRT system is typically an alternative to the LRT system it is reasonable to compare them with each other and exclude the regular bus service from the discussion. Probably the most apparent difference is the implementation cost because the LRT system requires massive investments in rail infrastructure while the BRT system can utilize the existing road infrastructure. However, the implementation of a BRT system is not completely free, because the bus stops serve as transport hubs, and therefore the logistics might require reorganizing the bus stop median facilities (Rathwell & King 2011). Also, the implementation time is lower in the BRT system compared to the LRT system (Deng & Nelson 2011). However, when completed the LRT system can operate at a lower cost, higher speed, and higher capacity (Vuchic 2002). According to Wei et al. (2020), people prefer public transportation with large capacity and high speed. Another strength of the LRT system is its right-of-way which makes it possible to avoid traffic congestion enabling higher speed and reliability (Wirasinghe et al. 2013). One of the main advantages of the BRT system is its flexibility. At the implementation stage, the route is easier to implement because there are hardly any of the land use issues related to rail infrastructure. (Deng & Nelson 2011)

As stated earlier, LRT systems have been found to have a significant effect on both land and housing prices but also BRT systems have increased land and house values (Beaudoin & Tyndall, 2023; Cervero & Kang 2011; Munoz-Raskin 2010). Zhang and Yen (2020) combined the information from 23 studies, and they found that both systems affect land and housing prices significantly, but the price effect of BRT systems might arise with lag.

Conventional bus services have not been found to have a significant impact on housing prices (Wei et al. 2020; Vuchic 2002). However, increases in land prices and consequently rising housing prices happen in the short term. In the long-term transportation infrastructure tends to affect land use and urban development. The more fixed the transportation system is the higher the impact it has on the urban development (Vuchic 2002). Therefore, one of the main advantages of the BRT system, flexibility, is also a weakness from a land development point of view (Deng & Nelson 2011).

In conclusion, the BRT system compared to the LRT system is a cost-effective way to increase accessibility which affects housing prices (Beaudoin & Tyndall, 2023). But as Henneberry (1998) states the long-term effects of improved transportation to urban development affect the housing prices as fundamentally as the increased accessibility in the short run. While the impact of the BRT system on urban development is comparably weak it gives to LRT system an advantage in the long run.

2.2.2 Other Factors Affecting the Strength of Housing Price Effect

The magnitude of the housing price effect caused by improvements in public transportation varies depending on several factors. One of them is the income level of the area which affects the housing prices in multiple ways. In low-income areas, people rely more on public transit which is why the residents have more willingness to pay for houses close to improved public transportation. While in high-income areas where people do not use public transportation as much the nuisance effect dominates. (Bowes & Ihlanfeldt 2001) Also, Wei et al. (2020) found that a high car ownership rate weakens the impact of public transportation on housing prices. On the other hand, Hess and Almeda (2007) found that the price effect of improved public transportation is greater in high-income areas compared to low-income areas. This may be because in high-income areas people who use public transportation appreciate more the timesaving caused by improved transportation.

Munoz-Raskin (2010) researched the housing price effect in Bogotá, Colombia. He found that the price effect was positive only in middle-income areas, while in the low-income and

high-income areas, the effect was negative. The negative effect in low-income areas is that even after the new BRT system was implemented people still use the old public transport because they prefer the slightly lower price to the time savings. In high-income areas, people use more private vehicles, and housing prices in those areas decreased as much as 14,9 % because of the new BRT system.

Increased accessibility might also increase the crime rate in the area. Bowes and Ihlanfeldt (2001) analyzed that in high-income areas the risk of being robbed and the potential loot are higher. As a consequence, they found that in high-income areas the housing price effect of public transportation is negative within a quarter mile of a station. But between one-quarter and three miles of a station, the crime effect no longer dominates, and a station affects positively housing prices. (Bowes & Ihlanfeldt 2001) Overall, the housing price effect of public transportation is affected by income level, but the direction and the intensity of the effect are highly dependent on the characteristics of the area.

2.2.3 Price Effect in Different Stages of Traffic Infrastructure Project

The price effect of a new public transportation project does not arise only after the project is completed. Because public transportation projects are typically long-lasting and include several stages, the magnitude and even the direction of the price effect might vary during the project. Yiu and Wong (2005) divide the effects into three categories, expected benefits of transportation improvements, costs of the construction nuisance, and the information costs of the improvements. If the information costs are zero and the expectation effect is stronger than the nuisance effect the housing prices will rise before a project is completed (Yiu & Wong 2005).

Hennebery (1998) investigated the price effect of the Supertram light rail project in New Yorkshire, UK. In the research, there were 3 datasets from different time points of the project, the first before the official decision of the project, the second right before the construction started, and the third four months after the construction was completed. The result showed that in the second timepoint, the anticipated nuisance of the construction had

decreased the house prices, but when the project was completed the nuisance effect disappeared. (Hennebery 1998) Yiu and Wong (2005) divided their data into three time periods attempting to analyze the housing price effect at different stages of the Western Harbour Tunnel project in Hong Kong. They found that the price effect took place well before the completion, even though the overall price increase was minimal during the middle period because of the expected construction nuisance (Yiu & Wong 2005).

2.3 Hedonic Pricing Model

To distinguish which part of the change in housing prices is due to improved transport connections it is necessary to investigate also the other factors affecting housing prices. In literature, there are two main ways to evaluate housing prices which are; the repeated sales approach and the hedonic pricing model. The strength of the hedonic model is concreteness because it prices each physical attribute that is included in the model. The weakness is that the method does not provide the right functional form for the model. (Sommervoll 2006) Even if the repeated sales approach does not require function form selection there is a crucial disadvantage related to sample size. The repeated sales approach requires a large number of sale transactions for each house which might cause problems, especially in new houses. (Dorantes et al. 2011) In academic literature, the hedonic model is much more common compared to the repeated sales approach (Sommerwoll 2006). Moreover, unlike the repeated sales approach the hedonic model can estimate the effect of a single variable on housing prices (Chin & Chau 2003). This feature is essential because the purpose of this research is to define the house price effect of the new tram line.

Hedonic methods were utilized long before the conceptual framework was determined (Chin & Chau 2003). While the method has evolved step by step, several researchers have played a central role in the development of the method. Several academic papers have mentioned the pioneering research of Andrew Court in 1939 when discussing the genesis of the hedonic pricing model (Dorantes et al. 2011; Goodman 1997; Malpezzi 2003) However, Colwell and Dilmore (1999) argue that Haas has created the guidelines for the hedonic models in the early 1920s. The hedonic model achieved a more generalized form after two groundbreaking approaches by Lancaster (1966) and Rosen (1974). Both approaches aimed to determine the

price of a heterogeneous product by specifying the different attributes of the product and pricing them individually. Despite the similarity of the models, the consumer price theory derived by Lancaster divides goods into groups according to their characteristics while in the model of Rosen, there is rather a range of goods than groups. Thus, the first-mentioned model fits better to the consumer goods and the last-mentioned to the durable goods. (Chin & Chau 2003)

The hedonic pricing model is widely used in academic research because it has found to be a powerful tool to determine the price of multidimensional goods such as real estate (Malpezzi 2003). While the value of a dwelling consists of various components that people value in different proportions the hedonic method determines the price for different components separately. Consequently, the price of a dwelling is a sum of implicit prices that the market ascribes to the various attributes of the dwelling. (Rosen 1974) Those implicit prices can be estimated through a regression analysis (Chin & Chau 2003).

In the hedonic model, the characteristics of dwellings are typically divided into three categories which are; structural, neighborhood, and locational characteristics (Hennebery 1998; Dorantes et al. 2011) Structural characteristics consist of physical elements of a dwelling like living space, number of bedrooms, and age. The neighborhood characteristics represent the pleasantness of the area that can be estimated for example by proxies such as crime rate, income rate, and the share of owner-occupiers. Locational-related features characterize the accessibility that can be measured by distance to certain places like the central business district or distance to the closest bus or rail stop. (Bowes & Ihlanfeldt 2001)

Rosen (1974) suggests that a good can be described as a vector of its measurable attributes and the price of each attribute can be described via the hedonistic model. When combining the set of specific attributes associated with a good and the price of each attribute, the price of a good is composed of the implicit prices of each attribute. Therefore, the hedonistic equation can be written as follows:

$$P(Z) = P(z_1, z_2, \dots, z_n)$$

Where P is price, Z is the combined vector of all attributes, and z_i represents the *i*th attribute of a good. While the attributes can be divided into three categories the hedonic equation can also be described as a function of those categories:

$$P = f(S, N, L)$$

Where S describes the structural attributes, N describes the neighborhood attributes, and L describes the locational attributes. (Rosen 1974)

2.3.1 Weaknesses of Hedonic Model

Even though the hedonistic model is a powerful tool to determine prices for differentiated products a model is always a simplification of reality, meaning that every model has its weaknesses. The hedonic pricing model offers a framework to determine the price for each attribute of a good, but unfortunately, the method does not take a position on what attributes should be included in the model and the right combination of components varies across different markets (Chin & Chau 2003). Malpezzi (2003) notes that there are hundreds of potential variables that could be included in the model. In the case of over-specification, the independent variables are still unbiased and consistent, but the model is inefficient due to irrelevant variables. Even bigger problem is under-specification which offers both biased and inconsistent coefficients. (Chin & Chau 2003) However, the method offers the categorization of the attributes which makes it easier to find the right variables. In the selection of variables, the only problem is not finding the right variables but also the availability of data. Even if the right variables are known the data might be unavailable for some of the variables. To avoid omitted variables the usage of proxy variables is possible. Though, if the proxy variable is incomplete it leads to measurement error (Chin & Chau 2003).

Another major issue with the model that is not defined is the right functional form (Rosen 1974). Chin and Chau (2003) note several functional forms that are typically used in hedonic models which are linear, semi-log, and log-log forms. A linear form has been criticized that it does not allow interactions between independent variables. For example, the price effect on an additional bathroom may vary depending on the size of a house. Logarithmic model makes it possible to detect such interactions. (Dubin 1998) Logarithmic forms also reduce the heteroscedasticity in the model (Dubin 1998; Ottensmann, Payton & Man 2008). On the other hand, Dubin refers to Goldberg (1986) who mentioned that in predictive research the logarithmic form causes problems because the unbiased estimators of the logarithmic model turn to biased ones when the estimators are turned back to price.

2.4 Ordinary Least Squares

Ordinary least squares (OLS) is the most common way to estimate the classical linear regression model (CLRM) (Brooks 2002, 44). OLS is a generalized and widely used linear modeling technique that allows one to explain and predict the dependent variable by one or more independent variables (Hutcheson & Sofroniou 1999, 55). The linearity requirement in OLS means that the model does not have to be linear in variables, but it has to be linear in parameters. In other words, the variables may contain for example logarithmic changes but the relationship between x and y must be capable of being presented linearly. (Brooks 2002, 54) OLS regression method enables to use of dichotomous or categorical independent variables if those are coded properly into dummy categories (Hutcheson & Sofroniou 1999, 55).

In the OLS method, the regression line is drawn to pierce the observation mass in a way that minimizes the sum of the squared vertical deviations of the observed values from the regression line. The errors are squared so that positive and negative errors do not cancel out each other. (Wilson, Keating & Beal-Hodges 2016, 28) The regression line describes the estimated values of the dependent variable, and the observed values are the actual values. Therefore, the vertical difference between an observation and the regression line reflects the estimation error of the particular observation in the model.

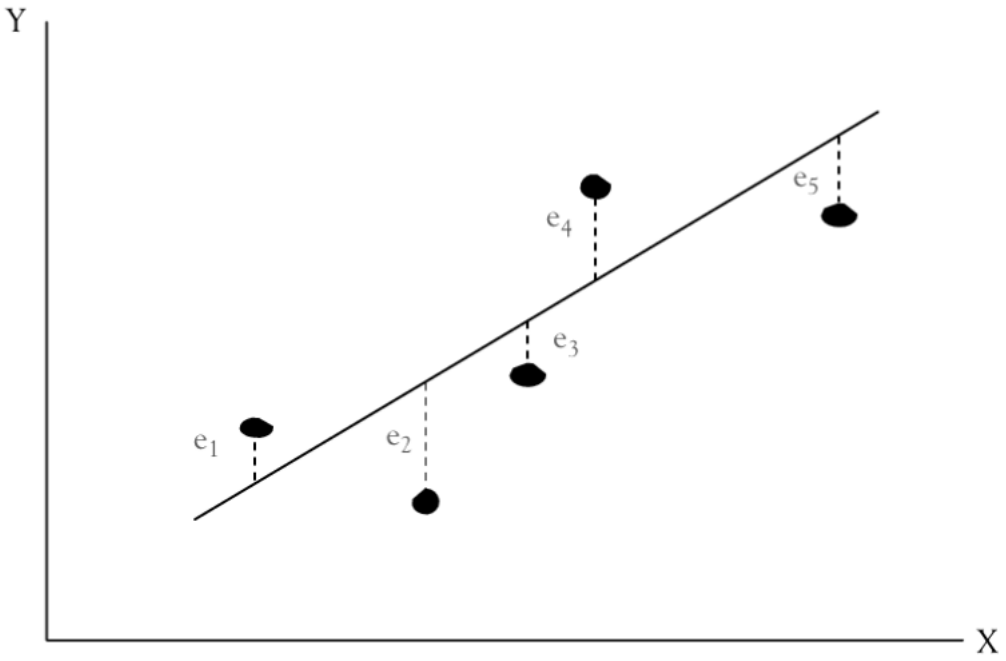


Figure 2 Residual illustration in OLS regression (Wilson et al. 2016, 27)

The population regression function (PRF) is a mathematical representation of a model that is generated by using all the actual data, and therefore describes the actual relationship between variables (Brooks 2002, 53):

$$y_t = \alpha + \beta x_t + u_t$$

Where α and β are the true parameters of the whole population. While the model typically contains only a small sample of the actual data the estimated sample regression function (SRF) is usually formed as:

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the estimated parameters. The real interest is on the real relationship between x and y, but because the PRF is not available the estimated SRF is used as a best guess of the parameters. Although the certainty about the relationship between the actual and estimated parameters cannot be achieved, based on the parameters estimated in SFR it is possible to estimate the probability that the corresponding population parameters take on certain values. (Brooks 2002, 53-54)

After the estimation of the model, the estimators $\hat{\alpha}$ and $\hat{\beta}$ are known in addition to the observations of x_t. Because y_t is also dependent on the error term $\hat{u}_t$ the goodness of the model is dependent on how the error term is generated. In OLS regression there are 5 assumptions for the error term which are:

1. $E(u_t) = 0$	The mean value of errors is zero
2. $var(u_t) = \sigma^2 < \infty$	Homoscedasticity: The variance of errors is constant
3. $cov(u_i, u_j) = 0$	No autocorrelation: The errors are linearly independent of one another
4. $cov(u_t, x_t) = 0$	No correlation between the error and the corresponding independent variable
5. $u_t \sim N(0, \sigma^2)$	The errors are normally distributed

When assumptions 1-4 hold the estimators of OLS are known to be the best linear unbiased estimators (BLUE). This means that the model has the lowest variance among the class of linear unbiased estimators, the model is linear to its estimators $\hat{\alpha}$ and $\hat{\beta}$, and on average the estimators $\hat{\alpha}$ and $\hat{\beta}$ are equal to the true values of α and β. To make valid interfaces about the population parameters from the sample parameters the fifth assumption is required. (Brooks 2002, 55-57)

3 Data

In this chapter, the data used in the thesis is presented, and the detailed data handling process is explained. The data of the thesis is obtained from Hintaseurantapalvelu (HSP) which is the most comprehensive statistic of Finnish housing market transactions, and it is maintained by the Central Federation of Finnish Real Estate Agencies. HSP is a private information pool that is intended for companies operating in the real estate and construction sector. The database includes all the housing transactions made through real estate agents in Finland since 1999. (KVKL 2023) The usage of realized transactions gives a truthful overview of the housing market.

3.1 Data Processing

In the beginning, the whole dataset contained 915 989 house transactions in Finland from the start of 1999 to October 2023. Because the tram projects are located in Tampere and Pirkkala, the data-handling process starts by removing houses that are located in other municipalities. Second, transactions outside the research window, from the beginning of 2015 to the end of 3rd quartile of 2023, are removed. This study aims to investigate the price development of residential buildings, and therefore real estate types such as business premises, commercial buildings, farm and forest estates, garages, holiday homes, morning places, parking spaces, plots, and storage facilities are removed. Houses with no living area information or houses with less than 10 m^2 living area are removed. Also, houses with higher than 500m^2 living area are removed. Based on theory, features like built year, lot ownership, location, and number of rooms are relevant when determining housing prices, and therefore observations that have empty values in any of those categories are removed. Lastly, houses further than 1 km from the closest tram station are removed. Number of observations after these steps is 15 844.

Table 1 Steps of the data handling process

Description of the operation	Removed observations	Remaining observations
The initial dataset		915 989
Removing houses that are located outside Tampere or Pirkkala	850 139	65 850
Removing observations outside the time window	13 897	51 953
Removing irrelevant real estate types	4 797	47 156
Removing houses with less than (<=) 10 m2 living area or with no living area information	128	47 028
Removing houses with more than (>=) 500 m2 living area	1 693	45 335
Removing houses with no built year information	2 057	43 278
Removing houses with no lot ownership information	85	43 193
Removing houses with no location information	1 927	41 266
Removing houses with no room information	134	41 132
Remove houses over 1km from the closest tram station	25 288	15 844

The distances in the model are calculated based on the coordinates that were already in the initial dataset. Coordinates of all the tram stations, universities, and the city center are defined by using Google Maps. The coordinates of the railway station are used as the coordinates of the city center. When all the necessary coordinates were determined, the distances were calculated by Euclidean distance.

The house type categories were rearranged. At this stage, the data consisted of the following house types: apartment houses, small apartment houses, balcony access buildings, detached houses, semi-detached houses, single-family houses, wooden houses, and row houses. The names of the new house categories are apartment, detached, and terraced. The apartment category consists of apartment houses, small apartment houses, and balcony access buildings. The detached category consists of detached houses, semi-detached houses, single-family houses, and wooden houses. The terraced category consists of row houses. Changes are also made in the category of lot ownership. Initially, there were categories owned, rented, and flexible rented. The owned category remained unchangeable, but the flexible rented category is joined into the rented category.

Table 2 Reorganization of categories

New house type categories	Initial categories
apartment	apartment_houses
	small_apartment_houses
	balcony_access_building
detached	detached_house
	semi_detached house
	single_family_house
	wooden_house
terraced	row_house

New lot ownership categories	Initial categories
owned	owned
rented	rented
	flexible_rented

3.2 Variable Selection

In addition to debt-free prices, HSP provides house-specific information related to the sale date, amount of liabilities, postcode, coordinates, municipality, building type, condition, living area, rooms, built year, floor number, total floors, lift, sauna, balcony, shore, lot ownership, lot area, energy class, and heat source. The hedonic pricing model provides a framework for selecting variables. Based on theory the chosen variables should cover structural, neighborhood, and locational aspects of a house. Unfortunately, the model does not tell which individual variables should be included in the model. In the first stage, the variables are selected based on previous academic research considering the limitations of available data. In the second stage, the final variable combination is chosen based on econometrical requirements. Therefore, some of the variables from the first stage are removed from the model if the variables are insignificant, there are multicollinearity issues, or the variable does not affect the coefficient of determination.

3.2.1 The First Stage of the Variable Selection

Considering the structural variables used in previous studies the variables living area, rooms, floor number, built year, and building type are included in the model at the first stage. Also, the lot ownership is included in the model. When the lot is owned the price of the lot is typically included in the house price, while the price of a house with a rented lot is typically lower because the lot rent increases the living costs, not the selling price. Also, a lot with a great location has an excellent potential for future value increase. (YIT, 2024) Therefore, the lot ownership is reasonable to be taken into account. Sauna and balcony could have been included in the model otherwise but most of the houses have empty value in those categories. Also, the condition variable is excluded from the model due to the huge number of empty values.

There are substantial issues related to the availability of the neighborhood variables. Based on previous research, there are several ways to determine the goodness or desirability of a neighborhood. As a proxy variable, there have been used the crime rate in the area and the ratio of owner-occupied houses to rented houses. According to Henneberry (1998), multiple studies have also used a selection of socio-economic variables collected from the national census. Unfortunately, any of this kind of neighborhood-related data has not been collected by residential areas in Tampere. Therefore, the model does not include neighborhood-related variables, which has to be considered when analyzing the results.

The last category in the hedonic pricing model is locational attributes. People are typically willing to pay for proximity to services and everyday activities. Because those services are concentrated in the CBD the distance to the CBD is one of the locational variables in the model. However, even if the services are located far away, well-working transportation connections can bring the services timewise rather close. Therefore, the distance to the closest tram station can be expected to have a significant effect on housing prices, and because one of the main interests of this study is to investigate the housing price effect of the tram line, the proximity of the closest tram station is included in the model. The third locational variable that is chosen is the proximity of a university. In Tampere, there are three

universities which are the University of Tampere, the University of Technology, and the University of Applied Sciences. The variable's value is the distance of a house to the closest university. Several studies also used the proximity to the closest highway but in this study, it is excluded from the model. The reason is that multiple highways get started or end in Tampere, and while a highway turns into a smaller road gradually, it is challenging to determine the clear distance points.

This study also aims to investigate the intensity of the housing price effect in different stages of the tram project. Therefore, a tram-related categorical variable is included in the model, indicating at which stage of the tram project the housing transaction has been completed. The variable divides the observations into three categories according to the time of occurrence. The first category is the period before the official decision of the implementation of the tram line, the second category indicates the construction period of the project, and the third category is the period after the tram project has been completed.

In addition to the structural, neighborhood, and locational variables some macroeconomic variables are included in the model. This is because there have been changes in macroeconomic conditions during the estimation window of 2015-2023, and the prediction window takes place just after that. According to Hypo (2023) inflation, interest rates, and the overall economic conditions affect the housing market development. The chosen macroeconomic variables are Euribor 3 months, consumer price index (CPI), employment rate, and the number of housing projects started 3 years ago. The variables are selected based on the relevance related to the housing markets, and the availability of the future forecasts of that variable, because the estimated future values of the macroeconomic variables are required to predict the housing price development.

Interest rates affect the willingness of an individual to buy a house because typically housing transactions are financed by loan. When interest rates are high it is less attractive to obtain a mortgage, and therefore the housing prices must be elastic downwards to find an equilibrium price. According to OP (2024), the most typical reference rate for mortgages in Finland is

the Euribor 12 months, but because the Bank of Finland offers a forecast only for the Euribor 3 months, the latter mentioned is selected in the model. The Euribor 3 months rates are collected from the databank of the Bank of Finland (2024a). The nominal housing prices tend to increase over time due to inflation, and to eliminate this effect the CPI is also included in the model. The employment rate is the third variable that is typically used to determine the general macroeconomic conditions. It also indicates the amount of people who have an opportunity to acquire a house. Therefore, the employment rate is included in the model. The data on CPI and employment variables are gathered from Statistics Finland (Tilastokeskus 2024a; Tilastokeskus 2024b). In addition to Euribor 3 months, the Bank of Finland offers forecasts for the CPI and employment rate. The forecasts extend until the end of 2026.

In addition to those more general macroeconomic variables, the model also includes the number of housing projects started 3 years ago to get information on how the supply side affects housing prices. The variable is calculated based on the started housing projects in Pirkanmaa, which is the province of which biggest city Tampere is. The data is offered by Statistics Finland (Tilastokeskus 2024c). The variable development process started by creating 4 lagged variables, from 1 year to 4 years. Only the variables lagged by 3 and 4 years were statistically significant, and while the variable lagged by 3 years offered a higher coefficient of determination it was selected for the model. An even more accurate variable would have been the number of completed house projects, but there is no forecast available for it. The strength of the lagged variable is that its values are possible to determine as far to the future as much the variable is lagged. Therefore, the values of the variable are possible to determine as far as the Bank of Finland has forecasted the other macroeconomic variables, to the end of 2026.

The analysis uses quarterly data of all the macroeconomic variables when determining the regression with the previous data. Because the Bank of Finland offers only a yearly forecast for the Euribor 3 months, CPI, and employment rate, the housing price prediction will be done by using yearly data for these variables (Bank of Finland 2024b). The 3-year lagged variable of the started housing projects is possible to determine quarterly based for the future, and therefore for this variable the quarterly data is used also for the prediction part.

3.2.2 The Second Stage of the Variable Selection Process

At the second stage of the variable selection process, the optimal model is determined based on the econometrical performance of each variable. In the beginning, the dependent variable of the model is chosen. Options for the dependent variable are price per square meter and total price. The aim of this study is to predict the future housing price development, and therefore the price per square meter would be more reasonable because the total price is much more dependent on the size of the houses that are sold in the future. However, in all the tested models where the price per square meter is the dependent variable autocorrelation problems arise. Also, the coefficient of determination is depending on the variable combination around 10 % higher in the models where the total price is the dependent variable. For these reasons, the total price is chosen to be the dependent variable.

After the first stage of the variable selection process, the possible independent variables for the final model are living area, number of rooms, built year, floor, house type, lot ownership, stage of the tram project, distance to tram station, distance to university, distance to CBD, Euribor 3 months, CPI, employment rate, and the 3-year lagged variable of the started housing projects. The final model is formed by using a stepwise method, at the beginning of which all variables are included in the model. After every regression, one variable is dropped from the model, and then the regression is run again. This procedure is done as many times as it is reasonable to drop any variable.

In the first three regressions, there have been variables whose P-value is over 0,05 which is a border level of insignificance. The insignificant variables are lot ownership, employment rate, and house type, whose P-values were 0,306, 0,072, and 0,063 respectively. It must be noted that the insignificance of the employment rate might be a consequence of multicollinearity with other macroeconomic variables, but the final selection of the macroeconomic variables is done later. After removing those three variables there are no insignificant variables in the model, but there are still variables whose effect on the coefficient of determination is negligible, and variables that increase the multicollinearity among the independent variables.

At the beginning of the process the adjusted R-squared, which describes the coefficient of determination of the model, was 0,6766. After removing the three insignificant variables the adjusted R-squared is 0,6725. Now the model does not include any insignificant variables, but the process continues by testing the effect of the floor variable on the adjusted R-squared because even floor variable is significant, its P-value is still the highest among the remaining variables being 0,008. Removing the floor variable the adjusted R-squared decreased from 0,6725 to 0,6724, meaning that its effect on the coefficient of determination of the model is almost zero, and therefore it excluded from the model. Next, the omission of the number of rooms variable is tested, because it is reasonable to assume that its independent explanatory power is weak because the living area variable is already included in the model. The omission of the variable decreases the adjusted R-squared from 0,6724 to 0,6707, which is still quite little. Also, the removal of the room variable decreases the VIF value of the living area variable from 4,77 to 1,08, which describes the multicollinearity of the variable among the independent variables. For these justifications, the room variable is removed from the model. Multicollinearity and VIF values are discussed in more detail later.

At this stage, there are only two structural variables, living area and built year, which both have a strong effect on the coefficient of determination. Removal of the built year would decrease the adjusted R-squared from 0,6707 to 0,5809, and removal of the living area would decrease the adjusted R-squared from 0,6707 to 0,1842. Therefore, both of the variables will be included in the final model. As the final structural variables are selected, and due to the data availability issues there are no neighborhood variables the attention focuses on the locational variables, which are distances to the tram station, university, and CBD. Among those three variables, the distance to the CBD is the most important variable in determining the housing prices, because its omission decreases the adjusted R-squared from 0,6707 to 0,4704. Removal of the tram station distance would decrease the adjusted R-squared from 0,6707 to 0,6609, which is a moderate drop but not huge. Also, the removal of the variable would increase P-values among the remaining variables, and the variable is fundamental considering the research questions. Hence, it is included in the final model. On the other hand, the removal of the university distance variable decreases the adjusted R-squared no

more than from 0,6707 to 0,686, which is so little that it is sufficient reason to exclude the variable from the final model.

At last, the combination of macroeconomic variables is selected. The only macroeconomic variable that is removed so far is the employment rate. But because its insignificance might be a consequence of multicollinearity with other macroeconomic variables, depending on the variable combination it could still be included in the model. Removal of the 3-year lagged variable of the started housing projects would make the stage of the tram project variable insignificant, and hence it is reasonable to keep it in the model. The exclusion of the Euribor variable would decrease the adjusted R-squared from 0,6686 to 0,6658 which is quite a small drop. A possible explanation for this is that during the estimation window, from the beginning of 2015 to the end of September 2023, the Euribor 3 months have been almost flat most of the time. And if there is no variance in a variable, it is impossible to capture its effect on the dependent variable. In the first 30 quartiles the average of the Euribor 3 month have been negative, fluctuating between -0,002 and -0,575. Only the last 5 quarters of the data have offered notable variability, and during that time the 3 months Euribor have risen from -0,575 to 3,618. And because the Bank of Finland has forecasted Euribor 3 months to be 2,6 on average already in 2024, which can be considered a significant decline compared to the current interest level, it is reasonable to keep the variable in the model from the point of view of the prediction. (Bank of Finland 2024b)

The last two macroeconomic variables to cover are CPI and employment rate. Even though there are good reasons for both of the variables to be included in the model, those variables do not fit in the same model. This is because multicollinearity would be extremely high, and the P-value of employment would be 0,034 which is rather close to the insignificance borderline of 0,05. The CPI is reasonable to keep in the model, because otherwise the stage of the tram project variable, which categories are determined based on the dates of housing transactions, would include not only the effect of the tram project but also the inflation. The disadvantage of the model where CPI is included is that there is still rather high multicollinearity, the VIF-values being on average 5,45. By replacing the CPI with employment rate the adjusted R-squared is almost the same, but the mean VIF declines to

2,55. Also, the P-values of all of the variables are 0,000, while the P-value of the during construction category of the stage variable is 0,006 in the model where CPI is used instead. Therefore, the model where the employment rate is used is econometrically better, but its weakness is that the inflation is partially contained by the stage of the tram project variable.

Both models, the one with CPI and the other with employment rate, have their advantages and disadvantages. Fortunately, choosing between the CPI and employment model is not mandatory, because the analysis and prediction can be done by both models and compare the results with each other. From now on, the models will be referred to as the CPI model and the employment model. The last variable whose inclusion in the model has not been discussed yet is the stage of the tram project variable. Because it is a crucial variable from the point of view of the research questions, and it is statistically significant in both models, it is included in the models.

Table 3 Table of the final variables

Dependent variable			
	Variable	Source	Type
	Price	HSP	Continuous
Independent variables			
Category	Variable	Source	Type
Structural	Living area	HSP	Continuous
	Built year	HSP	Continuous
Locational	Distance to CBD	Created (based on coordinates from HSP)	Continuous
	Distance to tram station	Created (based on coordinates from HSP)	Continuous
Tram related	Stage of the project	Created (based on sale dates from HSP) *Dummy variables: Before the decision, During the construction, After the completion* *Reference variable: After the completion*	Dummy
Macroeconomic	3-year lagged variable of started housing projects	Created (based on data from Statistics Finland)	Continuous
	Euribor	Bank of Finland	Continuous
	CPI (only in one of the models)	Statistic Finland	Continuous
	Employment Rate (only in one of the models)	Statistic Finland	Continuous

4 Methodology

In this research, the housing price predictions are done by OLS regression. To do so, the first regressions are run by the housing transaction data close to the tram project 1 stations during the estimation window. Based on the results the housing price predictions are done for the tram project 2 areas. To evaluate the prediction accuracy, the original data set is split into a training set and a test set in a ratio of 70-30. The regressions and the assumption tests are done with the training set, and before the actual prediction, the housing prices are estimated for the test set. Because the actual prices of the test set are already known, the estimated prices of the model can be compared with the actual prices.

4.1 Testing the OLS Assumptions

Unfortunately, the hedonic model does not offer the right functional form for the regression, and therefore a suitable model must be determined. The functional form of the regression is chosen between linear and logarithmic models. The advantage of the linear model is that the results are easier to read when there are no variable transformations. On the other hand, the logarithmic functional form often mitigates heteroscedasticity in the model (Malpezzi 2003). However, both functional forms have been used in previous research, and the better-fitting functional form depends on the data. To select the better fitting functional form the linear and logarithmic models are compared side by side. To do so the logarithmic transformations must be done for the continuous variables. Because the logarithmic model under comparison will be in the form of log-log also the dependent variable must be changed into logarithmic form.

Before getting into the results of the regression models, let’s compare the OLS assumptions of the linear models and the logarithmic models. Because as the result of the variable selection process, the research has decided to carry with two models, the CPI and the employment model, there are in total of four models to compare when the logarithmic modifications of those models are taken into account. While the linear models have reached

their final form, as a consequence of the modifications some variables in the logarithmic models have turned into insignificant. In the logarithmic CPI model, all three macroeconomic variables are insignificant with P-values of 0,623, 0,456, and 0,965. By dropping the Euribor 3 months variable the other two turned to significant. In the other logarithmic model, the only option to get all the macroeconomic variables significant is to remove both the 3-year lagged variable of the started housing projects and the employment rate variable. Therefore, the only macroeconomic variable that is included in the model is Euribor 3 months. Also, the stage of the tram project turned out to be insignificant in both logarithmic models, hence it is removed from the logarithmic models. Now the four models to compare are the linear CPI, the linear employment, the logarithmic CPI, and the logarithmic Euribor.

While the regression is automatically run in a way that the mean value of errors is zero, the first OLS assumption to be tested is homoscedasticity. If a model is not homoscedastic, meaning it is heteroscedastic, the estimators are still unbiased, but because the variance among the class of unbiased estimators is not the minimum, the estimators are not BLUE anymore (Brooks 2002, 150). Two widely used tests for homoscedasticity are the White test and the Breusch-Pagan test. Both linear models and the logarithmic CPI model get the P-value of 0,000 in both tests, meaning that the null hypotheses of homoscedasticity are rejected, and those three models are heteroscedastic. The result of the logarithmic Euribor model seems to be better. Even though the White test claims that also this model is heteroscedastic with P-values of 0,000, the Breusch-Pagan test confirms the null hypothesis of homoscedasticity with P-value of 0,7078. The same deduction can be also done based on Figure 4, where the residuals of each model are plotted against the fitted values. From the graphs of the linear models, it is easy to see that the variance of the residuals increases when the fitted values increase. The heteroscedasticity is much weaker in the logarithmic CPI model, but also in that graph the variance residuals tend to be a bit lower when the fitted values are lower. The same pattern cannot be seen from the graph of the logarithmic Euribor model.

Table 4 Test statistics of the White and Breusch-Pagan tests

	Model			
Statistical test	**Linear CPI**	**Linear Employment**	**Logarithmic CPI**	**Logarithmic Euribor**
White test	0,0000	0,0000	0,0000	0,0000
Breusch-Pagan	0,0000	0,0000	0,0000	0,7078

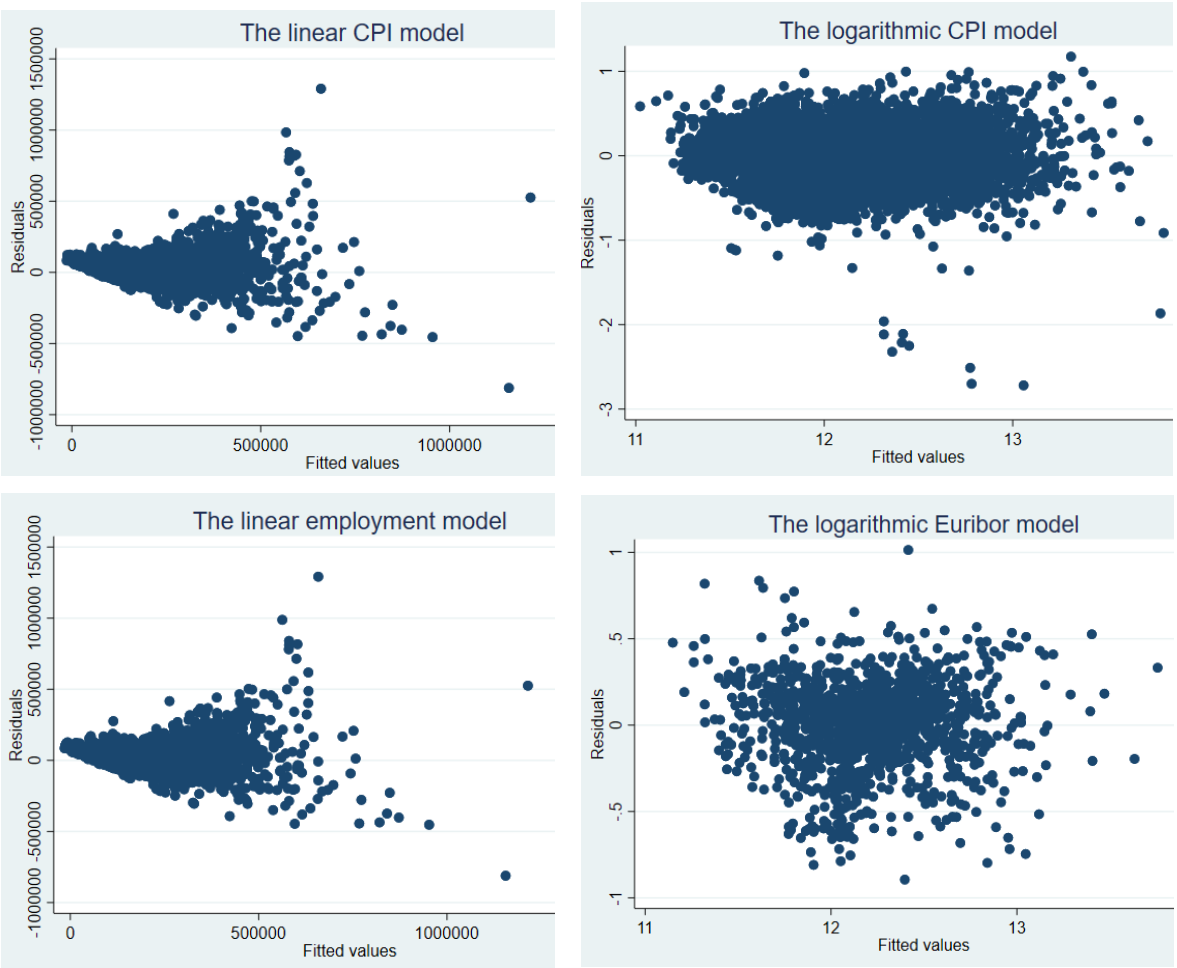


Figure 3 Variation of the residuals in the linear and logarithmic models

The next assumption is that there is no autocorrelation in the residuals. The consequences of autocorrelation are similar compared to heteroscedasticity. Autocorrelation makes the estimators inefficient, and the standard errors could be wrong. Anyhow, the estimators are

still unbiased. (Brooks 2002, 166) Autocorrelation can be tested by the Breusch-Godfrey test and the Durbin-Watson test. In the Breusch-Godfrey test, the logarithmic CPI model is the only one that does not exceed the critical value of 0,05, hence it is the only model where the autocorrelation occurs. However, in the logarithmic Euribor model, the autocorrelation is rejected with a much higher margin compared to the linear models. The P-values of the linear models are 0,1893 and 0,1767, while the corresponding value of the logarithmic Euribor model is 0,7078. However, in none of the three models, the P-value is not even close to 0,05, and hence it is safe to state that there are no autocorrelation problems in those models. In the Durbin-Watson test the test statistic gets a value between 0 and 4. Number 0 indicates perfect negative autocorrelation, 2 indicates no autocorrelation, and 4 indicates perfect positive autocorrelation. The test statistics of each model can be seen in the table 5. All values are rather close to the value of 1,4 meaning that there could be a weak negative autocorrelation. However, because the results of the Breusch-Godfrey test rejected the autocorrelation with a great margin in the three models, the only model where the autocorrelation is a problem is the logarithmic CPI model.

Table 5 Test statistics of the Breusch-Godfrey and Durbin-Watson tests

	Model			
Statistical test	**Linear CPI**	**Linear Employment**	**Logarithmic CPI**	**Logarithmic Euribor**
Breusch-Godfrey	0,1893	0,1767	0,0000	0,7587
Durbin-Watson	1,3373	1,3355	1,2908	1,4660

The last mandatory assumption requires that the residuals are normally distributed. It can be tested by the Shapiro-Wilk test. In all four models, the test statistic gets a value of 0,000, and therefore the models are not normally distributed. By looking at the graphs in Figure 5 it can be seen that in the linear models, the kurtosis is rather high, while the logarithmic models are not far from the normal distribution but there is a bit of kurtosis and negative skewness in both of the models. According to Brooks (2002, 182) in large sample sizes, which is the case in this study, the violation of normally distributed error terms is inconsequential. Therefore, the violation of that assumption can be ignored.

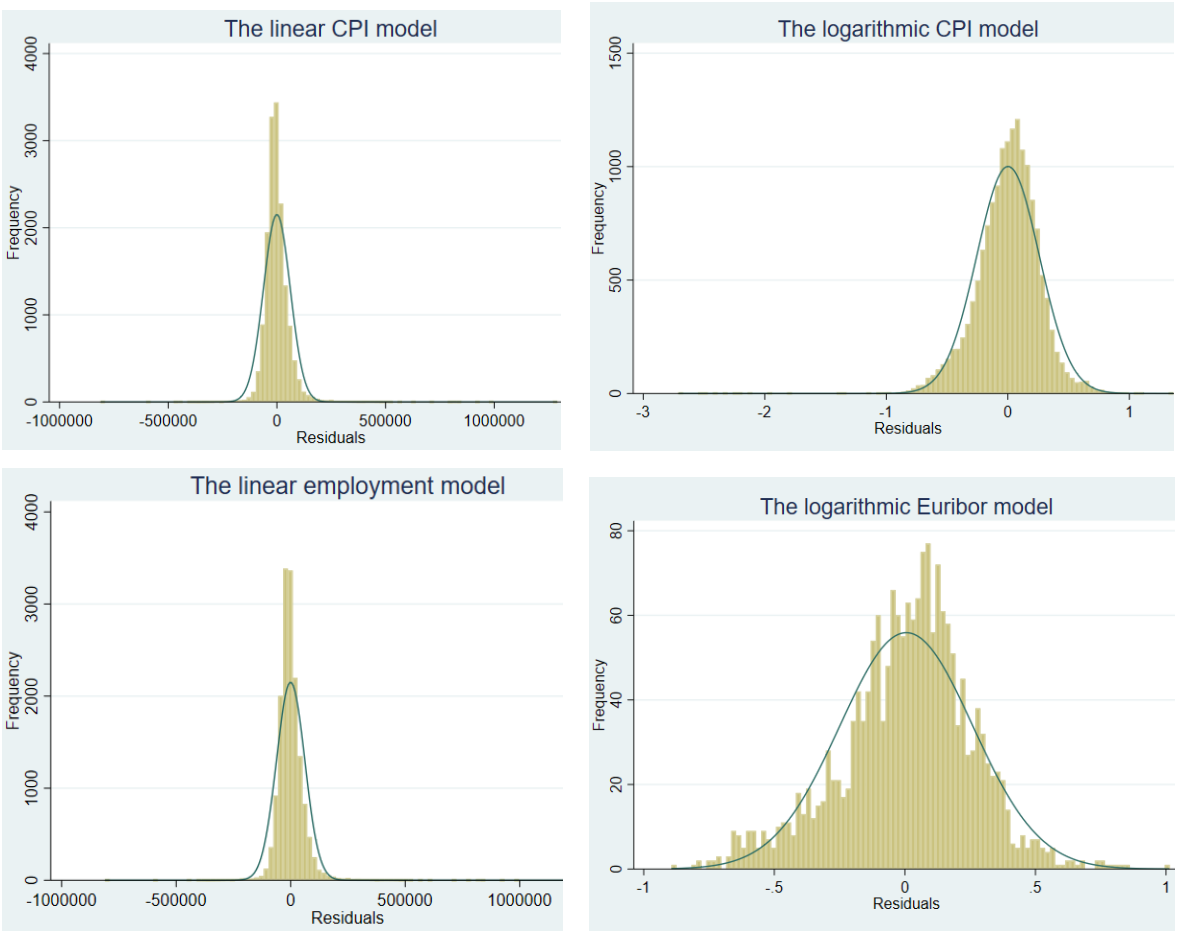


Figure 4 Distribution of residuals in the models

An additional assumption is no multicollinearity. It means that the correlation between independent variables should not be too strong. It can be tested by the VIF (Variance Inflation Factor) test. The test offers VIF value and 1/VIF value. Typically, when the VIF value is higher than 10, multicollinearity problems arise. The 1/VIF value tells how much the variable has independent explanatory power in the model.

Table 6 VIF tests for linear and logarithmic models

VIF tests for **linear** models		
	CPI	Employment
Living area	1,05	1,05
Built year	1,12	1,12
Tram distance	1,15	1,15
Centrum distance	1,05	1,05
Stage Const.	10,38	5,96
Stage Completed	9,44	3,95
Started houses lag 3	1,55	1,78
Euribor 3	5,04	1,38
CPI	18,25	-
Employment	-	5,47
Mean VIF	**5,45**	**2,55**

VIF tests for **logarithmic** models		
	CPI	Employment
Living area	1,05	1,04
Built year	1,09	1,14
Tram distance	1,08	1,05
Centrum distance	1,16	1,19
Stage Const.	-	-
Stage Completed	-	-
Started houses lag 3	1,18	-
Euribor 3	-	1,00
CPI	1,18	-
Employment	-	-
Mean VIF	**1,12**	**1,08**

The VIF values of each independent variable can be seen in Table 6, also the average VIF values of each model are calculated. As can be seen, the CPI causes problems in the linear CPI model, which is expected because it correlates with the categories of stage variable. That is the weakness of the model which is discussed above. Also, the Euribor 3 months gets a rather high VIF value in the model. In the linear employment model, the multicollinearity is much lower due to the CPI variable being omitted. Though, the employment seems to correlate a bit with the stage variable, but not as much as the CPI does in the other linear model. In the logarithmic models, the VIF values are much lower compared to the linear models. This is pretty obvious because there are fewer variables in the models. In the presence of high multicollinearity, the R-squared might be high even if the individual variables are not significant, which was the case before decreasing the amount of the macroeconomic variables in the logarithmic models. The standard errors of the significant variables might also increase. Second, the model becomes very sensitive to specifications, hence removing or adding variables might strongly affect the significance of other variables in the model. (Brooks 2002, 191) An example of this was the removal of the employment rate in the variable selection process. The variable was not insignificant due to its lack of explanation power but due to the multicollinearity with other variables. When the

combination of macroeconomic variables was changed the variable fitted into the other model.

In addition to the OLS assumptions, the model specification is tested. It will be done by the Ramsey reset test. The method uses higher orders of the dependent variable to explain the error terms, hence revealing the possibly wrong functional form and omitted variables (Brooks 2002, 194). The test shows that in all of the models, there are model specification issues. Based on the theory, the probable reason is that there are no neighborhood variables in the model due to the availability of data, meaning that there are omitted variables. This is the primary weakness of the model because it makes the estimated coefficients of the variables and the standard errors biased (Brooks 2002, 198).

4.2 Choosing the Functional Form

Based on the OLS assumption tests the weakest model is the logarithmic CPI model, which is the only one where the autocorrelation is present, and it did not overperform other models in any other assumption tests. Instead, the logarithmic Euribor model offered the best econometrical performance, because it is the only homoscedastic model, the autocorrelation was rejected by the largest margin and the multicollinearity was the lowest among the models. Also, the linear models performed rather well, though both of the models are heteroscedastic, and the multicollinearity is a bit higher, especially in the linear CPI model.

According to the assumption tests, it is clear that the logarithmic CPI model is rejected at this stage, but also the logarithmic Euribor model is rejected despite its better econometrical performance compared to the linear models. There are several reasons for that. At first, before the assumption tests, the aggressive removal of independent variables was undesirable from the point of view of the research questions and prediction vice. The dropout of the stage variable makes it impossible to determine the housing price effect of the stage of the tram project, and the removal of two of the three remaining macroeconomic variables also makes it much harder to execute the prediction for the tram project 2 areas. While the adjusted R-squared of the logarithmic Euribor model is effectively the same compared to the

linear models, it is reasonable to continue with the linear models, which are also easier to use. Therefore, the research will be done by the linear CPI model and the linear employment model.

4.3 Regression Models

Once the data is processed, OLS assumptions are tested, the variable selection is done, and the models with the right functional form are selected, it is time to implement the regressions. There is one categorical variable in both of the regressions, which is the stage variable of the tram project. It is encoded to a dummy variable in a way that the models assume that an observation belongs to the after completion category, and if that is not the case, either of the dummy variables before decision or during construction gets value 1, depending on which category the observation belongs to. Therefore, the final regression equation for the linear CPI model is:

$$Price = \beta 0 + \beta 1 * Living\ area + \beta 2 * Built\ year + \beta 3 * Distance\ to\ centrum + \beta 4 * Distance\ to\ tram\ station + \beta 5 * Stage\ (before\ decision) + \beta 6 * Stage\ (During\ construction) + \beta 7 * Started\ housing\ projects\ with\ 3\text{-}year\ lag + \beta 8 * Euribor\ 3\ months + \beta 9 * CPI$$

The final regression equation for the linear employment model is:

$$Price = \beta 0 + \beta 1 * Living\ area + \beta 2 * Built\ year + \beta 3 * Distance\ to\ centrum + \beta 4 * Distance\ to\ tram\ station + \beta 5 * Stage\ (before\ decision) + \beta 6 * Stage\ (During\ construction) + \beta 7 * Started\ housing\ projects\ with\ 3\text{-}year\ lag + \beta 8 * Euribor\ 3\ months + \beta 9 * employment$$

The regression results are shown in Table 7. In both regressions, all the variables, get an expected sign. However, there is no previous research where the lagged variable of the started housing projects would have been used to determine housing prices, and hence the expected sign of the variable is not clear. First, the variable explains the development on the

supply side, and typically increased supply tends to decrease the market equilibrium price. On the other hand, the constructors execute housing projects more actively during favorable housing market conditions and are expected to try to schedule the completion of the housing projects during periods when the prices are high. In addition, the 3-year lagged variable of the started housing projects assumes that housing construction projects would last three years, which is not always the case because the construction period can vary depending on several reasons, for example, the size of the project.

Table 7 Regression results of the linear CPI and employment models

The linear CPI model

Variable	Coefficient	P-value
Constant	-2702507	0,000
Living area	2903,8	0,000
Built year	1266,4	0,000
Centrum distance	-22534,3	0,000
Tram distance	-38749,2	0,000
Before decision	-20065,8	0,000
During constr.	-10263,7	0,006
After completion	0	0,000
Started houses 3lag	9,0804	0,000
Euribor 3 months	-14222,3	0,000
CPI	2828,4	0,000

P-value	0,0000
R-squared	0,6688
Adj. R-squared	0,6686

The linear employment model

Variable	Coefficient	P-value
Constant	-2620051	0,000
Living area	2903,8	0,000
Built year	1266,7	0,000
Centrum distance	-22548,9	0,000
Tram distance	-38820,1	0,000
Before decision	-32778,8	0,000
During const.	-22531,0	0,000
After completion	0	0,000
Started houses 3lag	9,2182	0,000
Euribor 3 months	-6458,8	0,000
Employment	2981,7	0,000

P-value	0,0000
R-squared	0,6684
Adj. R-squared	0,6681

While the estimates of structural and locational variables are quite similar, the biggest differences are in the stage variable and macroeconomic variables. In the employment model, the impact of the stage variable is much stronger, but as discussed before, it is an obvious consequence of the fact that the CPI variable is not in the model, and consequently the impact of inflation leaks into the stage variable. Therefore, the CPI model offers more reliable information on the impact of the stage of the tram project. Another clear difference between the models is that the impact of the Euribor 3 months is more than twice as much in the CPI model compared to the employment model. This is because there was multicollinearity between the macroeconomic variables, which makes it harder to observe

the independent explanatory power of variables where the problem is present. While the problem was solved by implementing two models with different combinations of macroeconomic variables, the models divide the impact between each macroeconomic variable differently. Unfortunately, it is rather hard to say which model offers more reliable information on the impact of interest rates on housing prices.

The coefficient of determination in the models is practically the same, while the adjusted R-squares of the models are 0,6686 and 0,6681, which are rather high values. The P-value of each variable in the employment model is 0,000, but in the CPI model, the P-value of the during construction category in the stage variable is 0,006. Anyway, the value is not even close to 0,05, and hence it does not cause problems.

The price impact of the proximity to the CBD seems to be rather high, and according to both models, every additional kilometer to the CBD decreases the house prices approximately by 22 500 euros. The proximity of a tram station is also crucial. While the models include only houses that are no more than 1 kilometer from the closest tram station, it is more descriptive to say that every additional 100 meters to the tram station decreases the housing prices by 3 800 euros. Also, the stage variable affects the house prices as expected. According to the CPI model, which is more accurate when determining the impact of the stage variable, housing prices have increased as a consequence of the completion of the tram project. In addition, the expectation effect of the upcoming tram line has had an impact on housing prices, because also the official decision of the tram project has increased the housing prices. Compared to the time the tram operated, the price of a house was approximately 20 000 euros lower before the official decision of the project, and around 10 000 euros lower during the construction period, ceteris paribus.

4.4 Prediction Model

At this stage, the housing prices of the tram project 1 areas are determined by two different models. Based on the information gathered from the regressions, the future price development is predicted for the tram project 2 areas. This will be done by first defining the

actual area of which price development will be predicted. After that, a time window is defined within which the completed housing transactions are used to implement the prediction. The prediction will be done by estimating the housing prices for the determined prediction dataset if those houses would be sold in the future. The same prediction dataset is used for every time point in the future in which the prediction is executed. Otherwise, the predicted prices would vary depending not only on the market conditions but also on the differences in the houses of the dataset. The predictions will be done from the 4th quartile of 2023 to the end of 2026.

4.4.1 Accuracy of the Prediction

Before getting into the actual prediction, the accuracy of the prediction model is estimated. This is possible because the data was divided into training and test sets before the regressions were executed. While the regressions were executed by using the training data, the prediction accuracy can be estimated by predicting the housing prices for the test set and comparing the estimated prices with the actual prices. The methods that are used for estimating the accuracy of the prediction model are mean absolute error (MAE) and cumulative abnormal return (CAR). The MAE indicator tells how much an estimated house price differs from the actual price on average. Instead, the CAR calculates the cumulative prediction error of the prediction model. The main difference between the indicators is that in the MAE indicator, the positive and negative values do not compensate for each other, which is the case in the CAR.

The MAE and CAR values of the models can be seen in Table 8. Both of the indicators are extremely close to each other's, hence it is impossible to confirm the superiority of either model based on these values. According to the MAE values, the estimation error of both models is approximately 38 000 euros, which is rather high. However, the indicator is highly affected by several extremely high values. In both prediction models, there is one observation where the estimation error is more than 900 000 euros, and in both models, there are around 250 observations where the estimation error is more than 100 000 euros. Most of the high estimation errors have occurred in comparably expensive houses, which is in line with the fact that there is heteroscedasticity in both models. Anyway, the values of the CAR

indicators are pretty close to zero, meaning that the positive and negative errors compensate each other almost entirely.

Table 8 Prediction accuracy indicators of both models

	MAE	CAR	Max error	Count of errors over 100 000 euros	Observations
CPI model	38 155	-425	927 100 €	246	4743
Employment model	38 217	-304	927 522 €	253	4743

To get a clearer view of the relationship between the actual prices and the predicted prices, the price development of the actual prices and the corresponding predicted prices are illustrated with a graph. This is done quarterly, due to which the quarterly average prices are calculated for the actual prices and both of the prediction models.

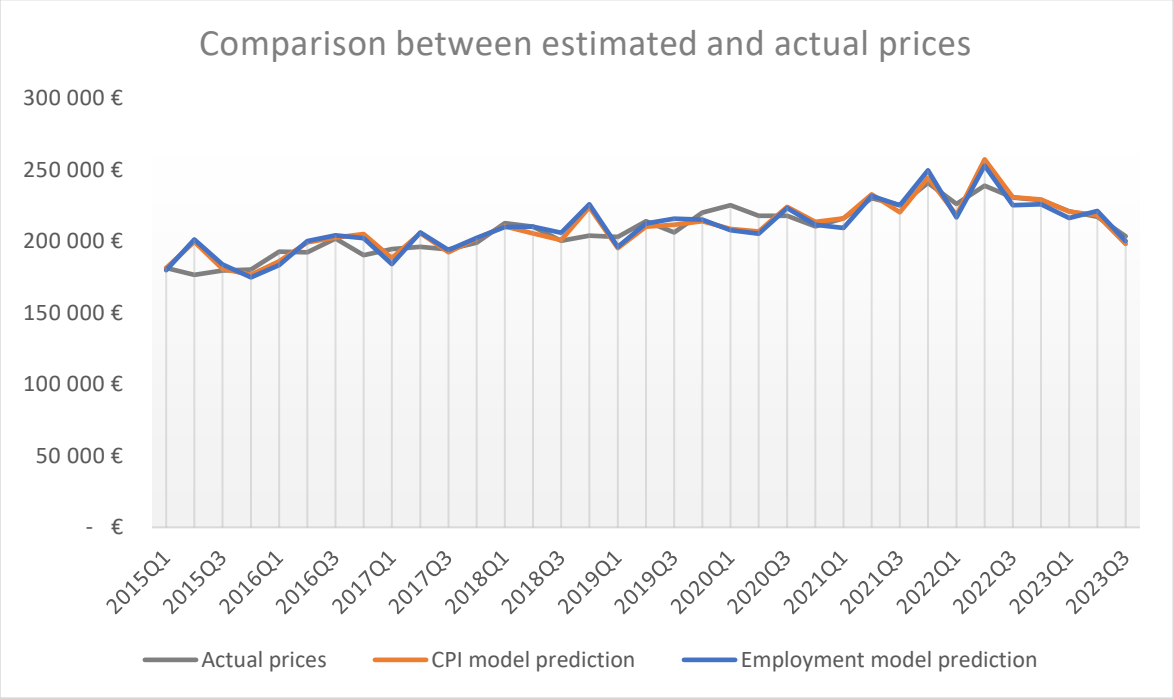


Figure 5 Comparison between estimated prices and actual prices

Figure 6 shows that the estimated prices have been quite close to the real prices. There are a few quartiles where the estimated prices have differed moderately from the real prices. There are a pair of possible reasons for that. At first, when there are large scale changes in macroeconomic variables or the stage variable, the model assumes that it affects housing markets overnight. In real life these kinds of changes could affect housing prices both beforehand, due to the expectation effect, and afterward, because people might react with time delay. Therefore, the prediction models might react to the changes more aggressively compared to reality. Second, while the price development of the real prices is done by using total prices instead of price per square meters, the quarterly price average is highly dependent on what kind of houses are sold in a certain quartile. For example, a few very expensive house transactions might have a comparably strong effect on the average price in the quartile.

As a conclusion can be said, that due to the high MAE values, the prediction models are not effective in predicting the prices for individual houses, at least when it comes to expensive houses. However, the low CAR values indicate that the models are appropriate for predicting the overall price development. Also, Figure 6 shows that despite a few exceptions, the predicted prices go hand in hand with the actual price development.

4.4.2 Defining the Prediction Model

At the beginning of defining the predictive model, the specific area of which price development will be predicted is determined. Because the implementation plan is still in progress, there is imperfect information of where the tram stations will be located, which is required to determine the distance to the closest tram station. Fortunately, the draft town plan from the areas of Härmälä and Rantaperkiö is already public information, and hence the probable locations of the two tram stations, Härmälä and Rantaperkiö, are known. Therefore, those areas are used in the prediction. Second, the prediction dataset is determined based on the HSP data. The same data cleaning process is done for that data as in chapter 3.1, but this time the houses located further than 1 kilometer from the stations of Härmälä and Rantaperkiö are removed. Then the dataset must be limited in time. The prediction dataset is decided to cover the housing transactions from the last known 8 quartiles, which are the

quartiles from 2021 Q4 to 2023 Q3. After these steps, the final prediction dataset contains 459 observations.

The predictions are made by estimating the prices for the prediction dataset based on the CPI and employment models if those houses would be sold in the future. Because the structural and locational variables do not change, the future price development is dependent on the macroeconomic variables and the stage variable. The Bank of Finland has created forecasts for the macroeconomic variables until the end of 2026, which is the limiting factor for the prediction window, and therefore the predictions are done from the last quartile of 2023 to the end of 2026. While the official decision of the tram project 2 will be made in October 2024 and the project is planned to be completed in 2028, the prediction period contains only the before decision and during construction categories of the stage variable. Of course, assuming that the official decision of the tram project will be confirmed. Because the specific date of the decision is not known, the whole 4^{th} quartile of 2024 is predicted by using the during construction category. The forecasts of the Bank of Finland are yearly based, but the 3-year lagged variable of the started housing projects have been calculated quarterly.

5 Results

The primary aim of this study has been to predict the housing price development in the tram project 2 areas, represented in this study by Härmälä and Rantaperkiö. In this chapter, the housing price predictions for those areas are presented from the 4th quartile of 2023 until the end of 2026. The predictions based on the CPI and the employment models are presented side by side, which makes it easier to compare predictions to each other. After the predicted housing price developments are established, the results are analysed and the factors affecting the reliability of the predictions are discussed.

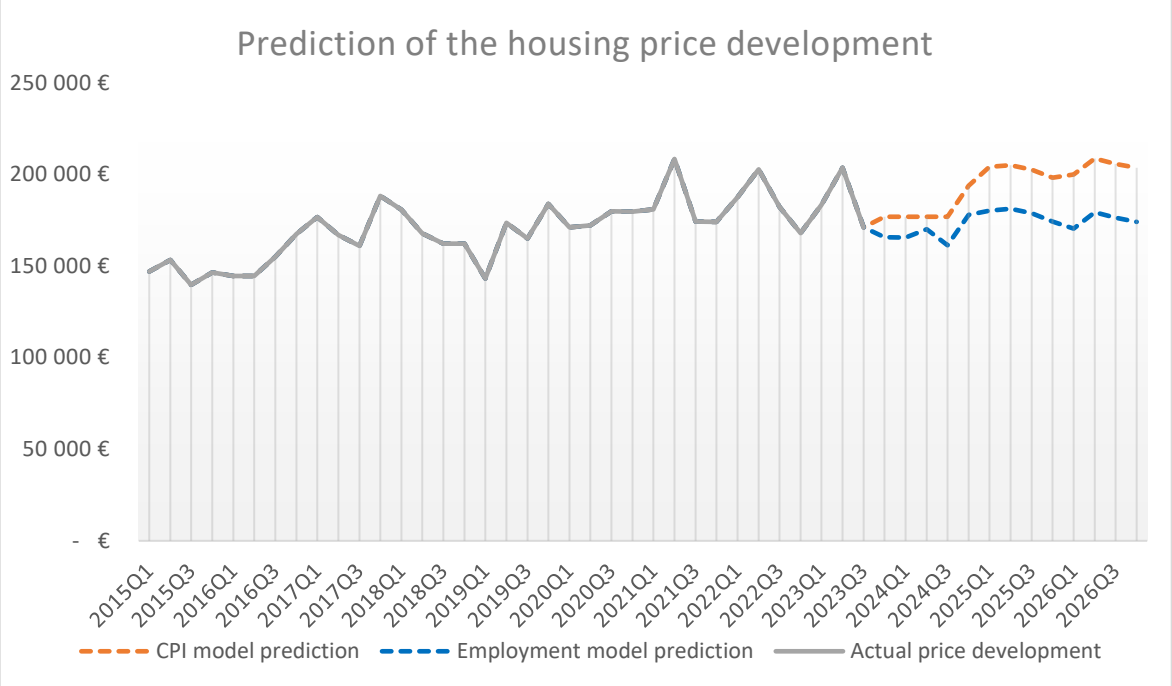


Figure 6 Predictions of the housing price development in Härmälä and Rantaperkiö

The predicted housing price development in the areas of Härmälä and Rantaperkiö is presented in Figure 7 as a continuation of the historical price development. The prediction of the CPI model is more optimistic. During the estimation window, the housing prices have increased by approximately 1,4 % per year in the areas of Härmälä and Rantaperkiö. The CPI model predicts that the yearly price increase in the areas will be around 2,4 % until the

end of 2026, whereas the corresponding value of the employment model is 1,2 %. Even though both of the models predict the price development to increase, the CPI model assumes an acceleration of the price development and the employment model assumes the growth pace is going to slow down.

A possible explanation for the prediction difference between the models is related to the consideration of inflation. While the employment model does not include the CPI variable, the effect of inflation is partially leaked into the stage variable. In the estimation window, the inflation started its intense increase only during the year 2021, which was the year the tram project 1 was completed. Therefore, most of the inflation effect that is leaked into the stage variable is in the after completion category, and while that category is not included in the prediction window, the effect of inflation is broadly ignored. On the other hand, the employment rate typically varies back and forth within a certain range, whereas inflation tends to increase in time. By a change, during the estimation window, the employment rate has steadily increased, and consequently, due to the correlation of those two variables, inflation might have leaked into the employment variable also. However, the employment variable has captured the inflation effect only partially because the variables are not perfectly correlated. As a consequence, the employment model offers a more pessimistic housing price prediction.

Another observation is that the predicted prices change more steadily, whereas in history the price changes between the quartiles have been more radical. One obvious reason for that is the fact that the historical price development of the quarterly average prices is highly dependent on the houses that are sold in a certain quartile, while in the predictions the same dataset is used in every quartile, and therefore the predictions are not affected by the differences in the sold houses in each quartile. However, for that reason, the historical price development does not give an exact picture of the price equilibrium in an individual quartile, but it is accurate enough to offer the big picture of the overall housing market development.

In the prediction curves, there are a few worth mentioning observations. At first, the price jump caused by the official decision is easy to point out between the last two quartiles of 2024. Of course, the price increase requires that the result of the official decision is to accept the implementation plan. Right after that, the prediction difference between the models becomes wider from 2025 onwards. That is a consequence of the forecasted decrease in interest rates. While the weight of the Euribor 3 months is more than twice as much in the CPI model compared to the employment model, it is clear that the drop from 3,6 % to 2,8 % in the variable affects more to the prediction of the CPI model.

In both models, the issues recognized in the OLS assumption tests affect the reliability of the predictions. At first, the heteroscedasticity exposes the models to possible distortion caused by expensive houses in the prediction dataset. As presented in chapter 4.4.1 the MAE of the models is rather high, but because the models do approximately as many positive and negative estimation errors, the CAR values are low. To minimize the estimation error caused by expensive houses, the prediction dataset was gathered from a comparably long time period, two years, which makes it more probable that there are roughly as many large-scale positive and negative estimation errors that would compensate for each other. The more crucial issue is the model specification problem, which is probably a consequence of omitting all the neighborhood variables. Therefore, the predictions do not take into account the attractiveness of the neighborhood, which would probably increase the predicted prices, because both Härmälä and Rantaperkiö are located by the lake. Also, the omission of the neighborhood variables might have affected other estimators in the models, which consequently makes the prediction models biased.

6 Conclusions

The thesis has predicted the housing price development in the Tampere tramway project 2 areas by answering the main research question which was formulated as follows:

> *How can the housing price development be predicted most accurately in the Tampere tramway project 2 areas?*

The research process started by estimating the best model to describe the housing prices in the tram project 1 areas. However, determining unequivocally best model turned out to be hard, and hence the research decided to continue with two different models, which differed from each other by the combination of macroeconomic variables. The regression results of both models were consistent with the previous research for all variables included in the models. However, there was no previous research where the lagged variable of the started housing projects would have been used. In the context of this research, the most relevant finding was that the availability of efficient public transportation has a great impact on housing prices, which can be noticed from the importance of the tram station's proximity and the stage of the tram project.

The predictions for the tram project 2 areas were executed based on the regression results from the tram project 1 areas. Both models forecasted increasing housing prices in the following years, however, the pace of the price development differs clearly between the models. The average annual price increase of the more optimistic prediction is 2,4 % in the following years, whereas the more pessimistic prediction forecasted the annual price increase to be just 1,2 % on average.

In this study, two sub-questions were created to support the main research question, and the first of the sub-questions was:

What macroeconomic variables should be included in the predictive model?

The theoretical framework of this study was the hedonic pricing model, which offered great guidelines for choosing the most relevant variables for determining housing prices. However, the hedonic pricing model is quite a static framework that pays attention only to rather permanent characteristics of a house, ignoring the dynamic components that affect housing prices. Because this study focuses on housing price prediction also the macroeconomic conditions must be taken into account. A limiting factor in the context of a predictive model is that from the chosen macroeconomic variables there must be some kind of forecasts available so that the prediction is possible to execute by utilizing the variables.

Given the requirements, the chosen macroeconomic variables were the 3-year lagged variable of the started housing projects, Euribor 3 months, CPI, and employment rate. Due to the multicollinearity issues the CPI and employment rate did not fit in the same model, which was the reason to execute two separate prediction models. Based on previous research it is pretty clear that both variables have independent explanatory power on housing prices, but because during the estimation window, the development of those variables was so similar that the statistical methods used in this study were not able to separate the independent explanatory power of those variables. To avoid that problem the estimation window should be so long that it would contain a more diverse combination of different types of market conditions.

The second sub-question of the study was:

How has the housing price effect of the tram line varied in different stages of the tram project 1?

In this study, the interest has been focused on housing price prediction in the areas where a significant public transportation improvement is under process. Because it is assumed that the improvement affects housing prices and the price effect does not happen overnight, the

intensity of the price effect in different stages of the project must be determined. This was executed by creating a categorical stage variable that separated housing transactions into categories before decision, during construction, and after completion. The price effect of the tram project 1 has been quite expected. Referring to the result of the CPI model, which is more relevant when estimating the price effect of the stage variable due to the consideration of inflation, the housing prices were approximately 20 000 euros lower before the official decision of the tram project 1 compared to the after completion period, ceteris paribus. During the construction period, the housing prices were just around 10 000 euros lower compared to the after completion period. Therefore, the expectation effect of the project has increased housing prices even before the start of the actual tram traffic.

In this study the nuisance effects of the tram construction, considering the air pollution, construction noise, and temporary traffic solutions, have not been determined and separated from the expectation effect. Hence, the during construction category includes not only the expectation effect of the tram project, which is presumably positive but also the nuisance effect of the construction, which is presumably negative. Therefore, the expectation effect of the Tram Project 1 is assumed to have been even stronger than the stage variable implies.

6.1 Limitations and Reliability of the Results

The biggest limitations of the research are related to the omission of neighborhood variables, which was due to the poor availability of residential area specific data. It does not only exclude the price effect of the attractiveness of a neighborhood, but it is also the probable cause for the model specification problems. Because a sufficient solution for the issue was not found, the pending insufficiency in the model specification affects the estimators of the other variables in the model making them biased. Therefore, the results of this research must be considered with caution.

Another factor to be considered is that the research predicted the development of total house prices. For the prediction the price per square meter could have been more reasonable because it is not that dependent on the size of the houses that are sold in the prediction window. However, the total house price was selected as the dependent variable because it removed the autocorrelation problem that was present in both of the models. The usage of the same prediction dataset over the whole prediction period minimises the problem making the predicted quarterly average prices comparable with each other.

The last issue to cover is the selection of the right functional form. The thesis was decided to execute by linear models because the selection of logarithmic models would have required the removal of variables that are essential for answering the research questions. However, the econometrical performance of the logarithmic Euribor model was slightly better compared to the linear models which imply that the relationship between each independent variable and the dependent variable is not necessarily linear. Therefore, the linear models used in the analysis may contain variables that affect the dependent variable non-linearly in reality.

References

Bank of Finland (2024a). Euriborkorot ja Eoniakorko. [Online]. [Accessed 23.1.2024]. Available at: https://www.suomenpankki.fi/fi/Tilastot/korot/taulukot2/korot_taulukot/euribor_korot_xml_long_fi/

Bank of Finland (2024b). Forecast tables for 2023–2026. [Online]. [Accessed 23.1.2024]. Available at: https://www.bofbulletin.fi/en/2023/6/forecast-tables-for-2023-2026-december-2023/

Beaudoin, J. & Tyndall, J. (2023). The effect of bus rapid transit on local home prices. Research in transportation economics, Vol. 102, 101370.

Bowes, D. R. & Ihlanfeldt, K. R. (2001). Identifying the Impacts of Rail Transit Stations on Residential Property Values. Journal of Urban Economics, Vol. 50 (1), 1-25.

Brooks, C. (2002). Introductory Econometrics for Finance. 1st edition. Cambridge University Press, Cambridge.

Cervero, R. & Kang, C. D. (2011). Bus rapid transit impacts on land uses and land values in Seoul, Korea. Transport policy, Vol.18 (1), p.102-116.

Chalermpong, S. (2007). Rail Transit and Residential Land Use in Developing Countries: Hedonic Study of Residential Property Prices in Bangkok, Thailand. Transportation research record, Vol. 2038 (1), 111-119.

Chin, T. L. and 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.

Colwell, P.F & Dilmore, G. (1999). Who was first? An examination of an early hedonic study. Land economics, 1999, Vol. 75 (4), 620-626.

Cordera, R., Coppola, P., dell'Olio, L. & Ibeas, Á. (2019). The impact of accessibility by public transport on real estate values: A comparison between the cities of Rome and Santander. Transportation research. Part A, Policy and Practice, Vol. 125, 308-319.

Debrezion, G., Pels, E. & Rietveld, P. (2011) The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market. Urban studies (Edinburgh, Scotland), Vol. 48 (5), 997-1015.

Debrezion, G., Pels, E. & Rietveld, P. (2007) The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-analysis. The journal of real estate finance and economics, Vol. 35 (2), 161-180.

Deng, T., Nelson, J. D. (2011) Recent Developments in Bus Rapid Transit: A Review of the Literature. Transport Reviews, Vol. 31 (1), 69-96.

Dorantes, L. M., Paez, A. & Vassallo, J. M. (2011). Analysis of House Prices to Assess Economic Impacts of New Public Transport Infrastructure: Madrid Metro Line 12. Transportation research record, Vol. 2245 (1), 131-139.

Dubin, R. A. (1998). Predicting house prices using multiple listings data. The journal of real estate finance and economics, 1998, Vol. 17 (1), 35-59.

Gao, L., Chong, H., Zhang, W. & Li, Z. (2023). Nonlinear effects of public transport accessibility on urban development: A case study of mountainous city. Cities, Vol. 138, 104340.

Gihrin, T. A. & Smith, J. J. (2006). Financing Transit Systems Through Value Capture. The American Journal of Economics and Sociology, Vol. 65 (3), 751-786.

Goodmann, A. C. (1997). Andrew Court and the Invention of Hedonic Price Analysis. Journal of Urban Economics, Vol. 44 (2), 291-298.

Henneberry, J. (1998). Transport investment and house prices. Journal of property investment & finance, Vol. 16 (2), 144.

Hess, D. B. & Almeida, T. M. (2007). Impact of Proximity to Light Rail Rapid Transit on Station-area Property Values in Buffalo, New York. Urban studies (Edinburgh, Scotland), Vol. 44 (5/6), 1041-1068.

Hutcheson, G. D. & Sofroniou, N. (1999). The multivariate social scientist introductory statistics using generalized linear models. 1st edition. London: SAGE Publications.

Hypo (2023). Hypon asuntomarkkinakatsaus Q4/2023. [Online]. [Accessed 23.1.2024]. Available at: https://www.hypo.fi/wp-content/uploads/2023/12/Hypon-Asuntomarkkinakatsaus_joulukuu-2023.pdf

KVKL (2024). KVKL Hintaseurantapalvelu. [Online]. [Accessed 23.1.2024]. Available at: https://kiinteistonvalitysala.fi/hintaseurantapalvelu/

Lancaster, K. J. (1966). A New Approach to Consumer Theory. The Journal of Political Economy, Vol. 74 (2), 132-157.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. Housing economics and public policy, Vol. 1, 67-89.

Mulley, C. (2014). Accessibility and Residential Land Value Uplift: Identifying Spatial Variations in the Accessibility Impacts of a Bus Transitway. Urban studies (Edinburgh, Scotland), Vol. 51 (8), 1707-1724.

Munoz-Raskin, R. (2010). Walking accessibility to bus rapid transit: Does it affect property values? The case of Bogotá, Colombia. Transport policy, Vol. 17 (2), 72-84.

OP (2024). Euribor and other reference interest rates. [Online]. [Accessed 23.1.2024]. Available at: https://www.op.fi/en/private-customers/loans-and-homes/interest-rates-and-prices/euribor

Ottensmann, J. R., Payton, S., & Man, J. (2008). Urban location and housing prices within a hedonic model. Journal of Regional Analysis and Policy, Vol. 38(1).

Raitiotiealliansi (2023a). Raitiotien ensimmäinen osuus valmistuu etuajassa ja 34 miljoonaa euroa alle budjetin, liikenne alkaa 9.8. . [Accessed 25.10.2023]. Available: https://raitiotiealliansi.fi/tiedotteet/raitiotien-ensimmainen-osuus-valmistuu-etuajassa-ja-34-miljoonaa-euroa-alle-budjetin-liikenne-alkaa-9-8/

Raitiotiealliansi (2023b). Osa 2 Pyynikintori - Santalahti - Lentävänniemi. [Accessed 25.10.2023]. Available: https://raitiotiealliansi.fi/rakentaminen/osa-2/

Rathwell, S. & King, M. (2011). Considerations for Median BRT on Arterial Roads. ITE Journal, Vol. 81 (1), 44-48.

Rode, P., Floater, G., Thomopoulos, N., Docherty, J., Schwinger, P., Mahendra, A. & Fang, W. (2014). IDEAS Working Paper Series from RePEc.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. The Journal of Political Economy, Vol. 82 (1), 34-55.

Ryan, S (1999). Property Values and Transportation Facilities: Finding the Transportation-Land Use Connection. Journal of planning literature, Vol. 13 (4), 412-427.

So, H. M., Tse, R.Y.C. & Ganesan, S. (1997). Estimating the influence of transport on house prices: evidence from Hong Kong. Journal of property investment & finance, Vol. 15 (1), 40.

Sommervoll, D. E. (2006). Temporal Aggregation in Repeated Sales Models. The journal of real estate finance and economics, Vol. 33 (2), 151-165.

Tampereen Ratikka (2023a). Tampereen ratikan suunitelmia osalta 1 ja 2. [Accessed 25.10.2023]. Available: https://www.tampereenratikka.fi/osien-1-ja-2-suunnitelmia/

Tampereen Ratikka (2023b). Ratikan vaiheet. [Accessed 25.10.2023]. Available: https://www.tampereenratikka.fi/ratikan-vaiheet/

Tampereen Ratikka (2023c). Raitiotien hankesuunnittelu Pirkkalasta Linnainmaalle alkaa. [Accessed 25.10.2023]. Available: https://www.tampereenratikka.fi/suunnittelu-ajankohtaista/pirkkala/raitiotien-hankesuunnittelu-pirkkalasta-linnainmaalle-alkaa/

Tampereen Ratikka (2023d). Tampereen ja Pirkkalan valtuustot päättivät Pirkkala-Linnainmaa -raitiotien toteutussuunnittelun aloittamisesta. [Accessed 25.10.2023]. Available: https://www.tampereenratikka.fi/suunnittelu-ajankohtaista/pirkkala/tampereen-ja-pirkkalan-valtuustot-paattivat-pirkkala-linnainmaa-raitiotien-toteutussuunnittelun-aloittamisesta/

Tilastokeskus (2024a). 11xb -- Kuluttajahintaindeksi (2015=100), kuukausitiedot, 2015M01-2023M12. [Online]. [Accessed 23.1.2024]. Available at: https://pxdata.stat.fi/PxWeb/pxweb/fi/StatFin/StatFin__khi/statfin_khi_pxt_11xb.px/

Tilastokeskus (2024b). 135z -- Työvoimatutkimuksen tärkeimmät tunnusluvut, niiden kausitasoitetut aikasarjat sekä kausi- ja satunnaisvaihtelusta tasoitetut trendit, 2010M01-

2023M12. [Online]. [Accessed 23.1.2024]. Available at: https://pxdata.stat.fi/PxWeb/pxweb/fi/StatFin/StatFin__tyti/statfin_tyti_pxt_135z.px/

Tilastokeskus (2024c). 12fy -- Rakennus- ja asuntotuotanto, 1995M01-2023M12. [Online]. [Accessed 23.1.2024]. Available at: https://pxdata.stat.fi/PxWeb/pxweb/fi/StatFin/StatFin__ras/statfin_ras_pxt_12fy.px/

Vuchic, V. R. (2002). Bus Semirapid Transit Mode Development and Evaluation. Journal of Public Transportation, Vol. 5 (2), 271-95.

Wei, F., Zhang, D. & Wen, H. (2020). Analysis of the Impacts of Urban Public Transport on the Housing Price in Hangzhou, China. Journal of Physics. Conference series, Vol. 1616 (1), 12028.

Wilson, J. H., Keating, B. P. & Beal-Hodges, M. (2016). Regression analysis: understanding and building business and economic models using Excel. 2nd edition. Business Expert Press, New York.

Wirasinghe, S. C., Kattan, L., Rahman, M. M., Hubbell, J., Thilakaratne, R. & Anowar, S. (2013). Bus rapid transit - a review. International Journal of Urban Sciences, Vol. 17(1), 1-31.

YIT (2024). Oma, vuokratontti vai valinnainen vuokratontti? [Online]. [Accessed 25.1.2024]. Available at: https://www.yit.fi/asunnot/myytavat-asunnot/asunnon-osto/rahoitus/vuokratontti

Yiu, C. Y. & Wong, S. K. (2005). The Effects of Expected Transport Improvements on Housing Prices. Urban Studies, Vol. 42 (1), 113-125.

Zhang, M. & Yen, B. T. H. (2020). The impact of Bus Rapid Transit (BRT) on land and property values: A meta-analysis. Land use policy, Vol. 96, 104684.

Zhou, J. & Yang, Y. (2021). Transit-based accessibility and urban development: An exploratory study of Shenzhen based on big and/or open data. Cities, Vol. 110, 102990.

Appendices

Appendix 1. Regression results and Ramsey RESET-test for linear CPI model

Source	SS	df	MS	Number of obs	=	11,101
				F(9, 11091)	=	2488.73
Model	8.9611e+13	9	9.9567e+12	Prob > F	=	0.0000
Residual	4.4372e+13	11,091	4.0007e+09	R-squared	=	0.6688
				Adj R-squared	=	0.6686
Total	1.3398e+14	11,100	1.2071e+10	Root MSE	=	63251

Price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Area	2903.803	22.30462	130.19	0.000	2860.081	2947.524
BuiltYear	1266.411	22.86998	55.37	0.000	1221.582	1311.241
CentrumDist	-22534.33	276.2733	-81.57	0.000	-23075.88	-21992.79
TramDist1	-38749.19	2388.617	-16.22	0.000	-43431.3	-34067.07
Stage2						
Before Decision	-20065.81	5005.149	-4.01	0.000	-29876.8	-10254.83
During Construction	-10263.73	3729.178	-2.75	0.006	-17573.59	-2953.881
StartedApartLag3	9.080448	1.596837	5.69	0.000	5.950363	12.21053
Euribor3	-14222.3	1465.242	-9.71	0.000	-17094.43	-11350.16
CPI	2828.359	451.4374	6.27	0.000	1943.461	3713.257
_cons	-2702507	66827.2	-40.44	0.000	-2833501	-2571514

```
Ramsey RESET test using powers of the fitted values of Price
       Ho:  model has no omitted variables
                 F(3, 11088) =     508.92
                    Prob > F =     0.0000
```

Appendix 2. Regression results and Ramsey RESET-test for linear employment model

Source	SS	df	MS	Number of obs	=	11,101
				F(9, 11091)	=	2483.89
Model	8.9553e+13	9	9.9503e+12	Prob > F	=	0.0000
Residual	4.4430e+13	11,091	4.0059e+09	R-squared	=	0.6684
				Adj R-squared	=	0.6681
Total	1.3398e+14	11,100	1.2071e+10	Root MSE	=	63293

Price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Area	2903.841	22.31927	130.10	0.000	2860.092	2947.591
BuiltYear	1266.744	22.88477	55.35	0.000	1221.886	1311.602
CentrumDist	-22548.91	276.4628	-81.56	0.000	-23090.83	-22007
TramDist1	-38820.06	2390.714	-16.24	0.000	-43506.28	-34133.83
Stage2						
Before Decision	-32778.78	3796.524	-8.63	0.000	-40220.65	-25336.92
During Construction	-22531.03	2414.536	-9.33	0.000	-27263.95	-17798.12
StartedApartLag3	9.218247	1.710991	5.39	0.000	5.864401	12.57209
Euribor3	-6458.764	765.8609	-8.43	0.000	-7959.988	-4957.541
Employment	2981.71	599.318	4.98	0.000	1806.94	4156.479
_cons	-2620051	64089.56	-40.88	0.000	-2745678	-2494424

```
Ramsey RESET test using powers of the fitted values of Price
       Ho:  model has no omitted variables
                 F(3, 11088) =     509.04
                    Prob > F =     0.0000
```

Appendix 3. Regression results and Ramsey RESET-test for logarithmic CPI model

Source	SS	df	MS	Number of obs	=	11,101
				F(6, 11094)	=	3409.03
Model	1377.3676	6	229.561266	Prob > F	=	0.0000
Residual	747.061323	11,094	.067339221	R-squared	=	0.6483
				Adj R-squared	=	0.6482
Total	2124.42892	11,100	.191389993	Root MSE	=	.2595

ln_Price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_Area	.7320013	.0062361	117.38	0.000	.7197774	.7442253
ln_BuiltYear	10.05159	.1826711	55.03	0.000	9.69352	10.40966
ln_CentrumDist	-.2561964	.0028386	-90.25	0.000	-.2617605	-.2506322
ln_TramDist1	-.0349026	.0033435	-10.44	0.000	-.0414564	-.0283488
ln_StartedApartLag3	.0810068	.0062001	13.07	0.000	.0688536	.0931601
ln_CPI	.8875716	.0514824	17.24	0.000	.7866569	.9884863
_cons	-71.69502	1.405065	-51.03	0.000	-74.4492	-68.94084

```
Ramsey RESET test using powers of the fitted values of ln_Price
       Ho:  model has no omitted variables
                 F(3, 11091) =     50.31
                    Prob > F =      0.0000
```

Appendix 4. Regression results and Ramsey RESET-test for logarithmic Euribor model

Source	SS	df	MS	Number of obs	=	1,348
				F(5, 1342)	=	545.22
Model	174.120667	5	34.8241334	Prob > F	=	0.0000
Residual	85.7165409	1,342	.063872236	R-squared	=	0.6701
				Adj R-squared	=	0.6689
Total	259.837208	1,347	.192900674	Root MSE	=	.25273

ln_Price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_Area	.7506438	.0178109	42.15	0.000	.7157035	.7855841
ln_BuiltYear	11.25516	.5466964	20.59	0.000	10.18269	12.32763
ln_CentrumDist	-.2709958	.0080028	-33.86	0.000	-.2866951	-.2552965
ln_TramDist1	-.0368538	.0095857	-3.84	0.000	-.0556583	-.0180492
ln_Euribor3	-.0149504	.0048323	-3.09	0.002	-.0244302	-.0054706
_cons	-76.14388	4.151898	-18.34	0.000	-84.28879	-67.99896

```
Ramsey RESET test using powers of the fitted values of ln_Price
       Ho:  model has no omitted variables
                 F(3, 1339) =     14.55
                    Prob > F =      0.0000
```

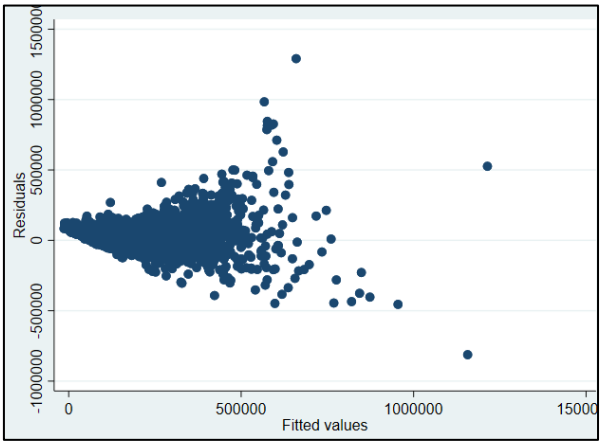
Appendix 5. Testing for homoscedasticity in linear CPI model: White's test, Breusch-Pagan-test, and scatterplot

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(51)     =   3049.43
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	3049.43	51	0.0000
Skewness	128.93	9	0.0000
Kurtosis	-1.39e+12	1	1.0000
Total	-1.39e+12	61	1.0000




```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of Price

         chi2(1)      = 22865.80
         Prob > chi2  =   0.0000
```

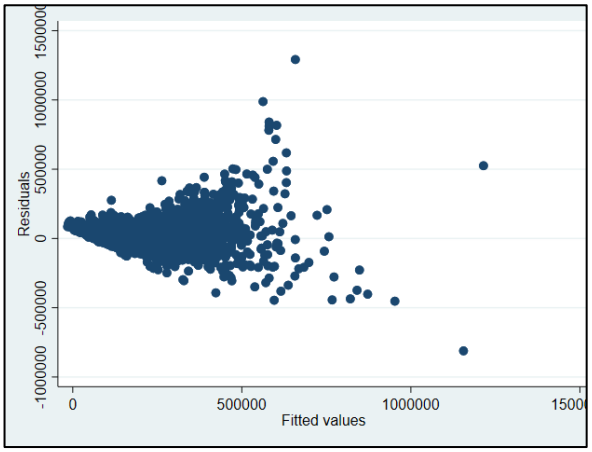
Appendix 6. Testing for homoscedasticity in linear employment model: White's test, Breusch-Pagan-test, and scatterplot

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(51)     =   3039.61
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	3039.61	51	0.0000
Skewness	127.93	9	0.0000
Kurtosis	-1.33e+12	1	1.0000
Total	-1.33e+12	61	1.0000




```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of Price

         chi2(1)      = 22768.22
         Prob > chi2  =   0.0000
```

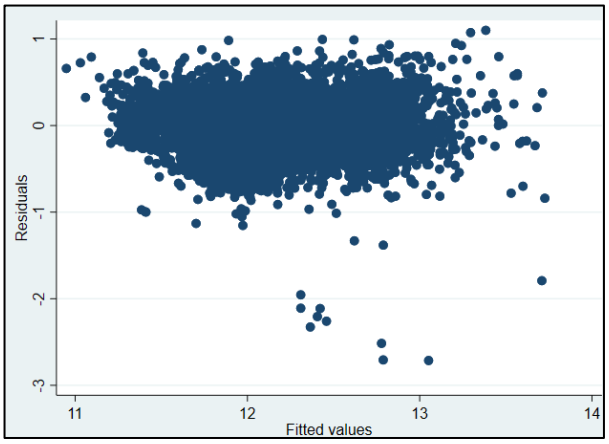
Appendix 7. Testing for homoscedasticity in logarithmic CPI model: White's test, Breusch-Pagan-test, and scatterplot

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(27)     =    734.34
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	734.34	27	0.0000
Skewness	69.03	6	0.0000
Kurtosis	10.56	1	0.0012
Total	813.93	34	0.0000




```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_Price

         chi2(1)      =   148.97
         Prob > chi2  =   0.0000
```

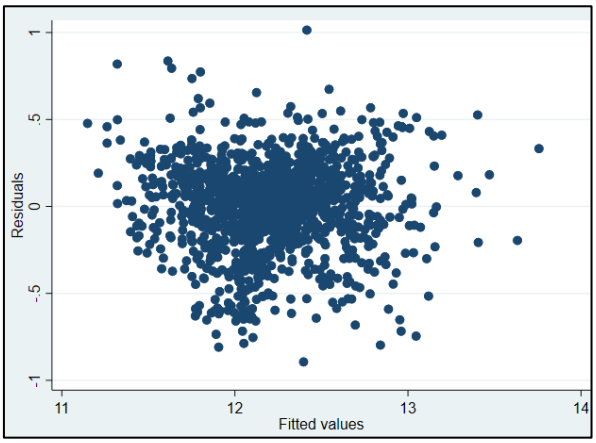
Appendix 8. Testing for homoscedasticity in logarithmic Euribor model: White's test, Breusch-Pagan-test, and scatterplot

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(20)     =    300.04
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

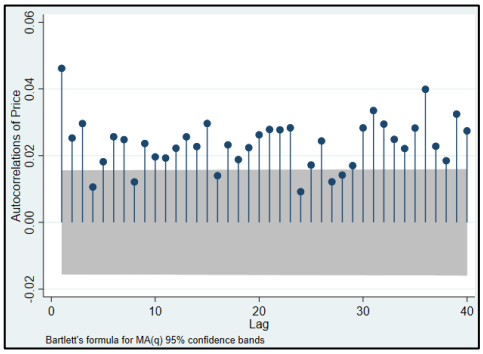
Source	chi2	df	p
Heteroskedasticity	300.04	20	0.0000
Skewness	67.55	5	0.0000
Kurtosis	13.99	1	0.0002
Total	381.59	26	0.0000

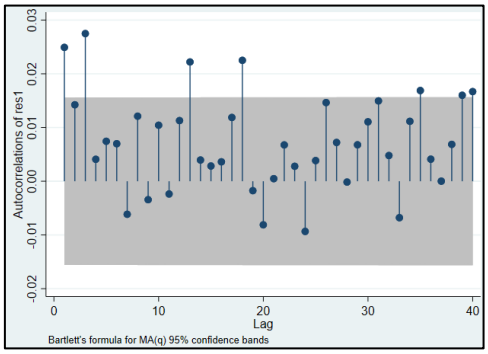



```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of ln_Price

         chi2(1)      =     0.14
         Prob > chi2  =   0.7078
```

Appendix 9. Testing for autocorrelation in linear CPI model: Correlograms, Breusch-Godfrey-test and Durbin-Watson-test





Number of gaps in sample: **3334**

Breusch-Godfrey LM test for autocorrelation

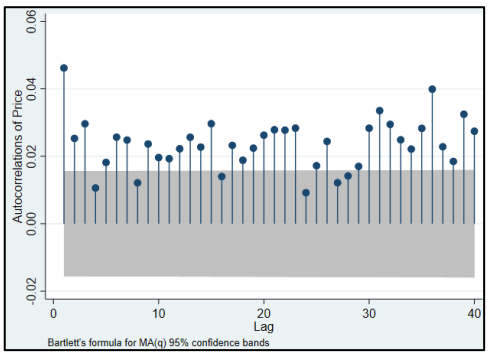
lags(*p*)	chi2	df	Prob > chi2
1	**1.723**	**1**	**0.1893**

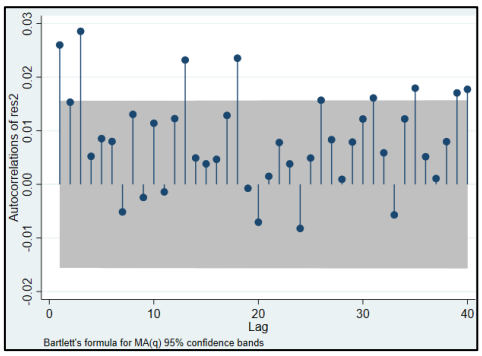
H0: no serial correlation

Number of gaps in sample: **3334**

Durbin-Watson d-statistic(**10, 11101**) = **1.337267**

Appendix 10. Testing for autocorrelation in linear employment model: Correlograms, Breusch-Godfrey-test and Durbin-Watson-test





Number of gaps in sample: **3334**

Breusch-Godfrey LM test for autocorrelation

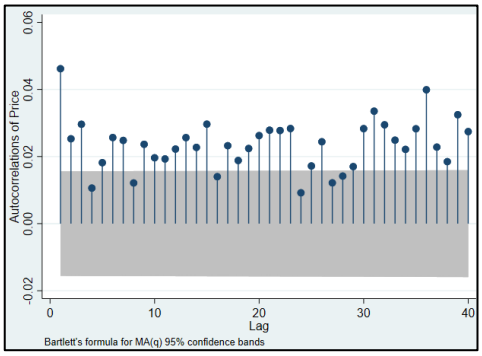
lags(*p*)	chi2	df	Prob > chi2
1	**1.825**	**1**	**0.1767**

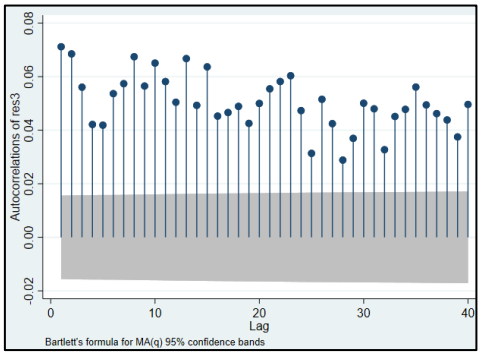
H0: no serial correlation

Number of gaps in sample: **3334**

Durbin-Watson d-statistic(**10, 11101**) = **1.335482**

Appendix 11. Testing for autocorrelation in logarithmic CPI model: Correlograms, Breusch-Godfrey-test, and Durbin-Watson-test





Number of gaps in sample: 3334

Breusch-Godfrey LM test for autocorrelation

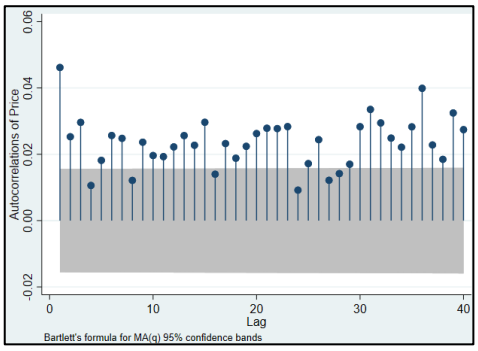
lags(*p*)	chi2	df	Prob > chi2
1	53.979	1	0.0000

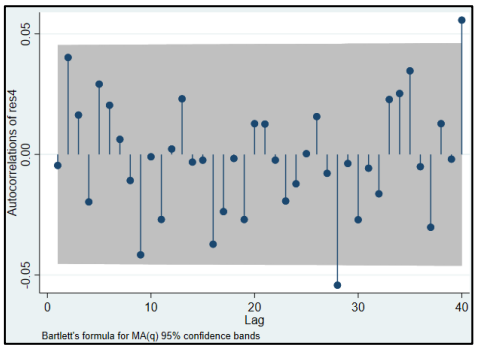
H0: no serial correlation

Number of gaps in sample: 3334

Durbin-Watson d-statistic(7, 11101) = 1.290787

Appendix 12. Testing for autocorrelation in logarithmic Euribor model: Correlograms, Breusch-Godfrey-test, and Durbin-Watson-test





Number of gaps in sample: 360

Breusch-Godfrey LM test for autocorrelation

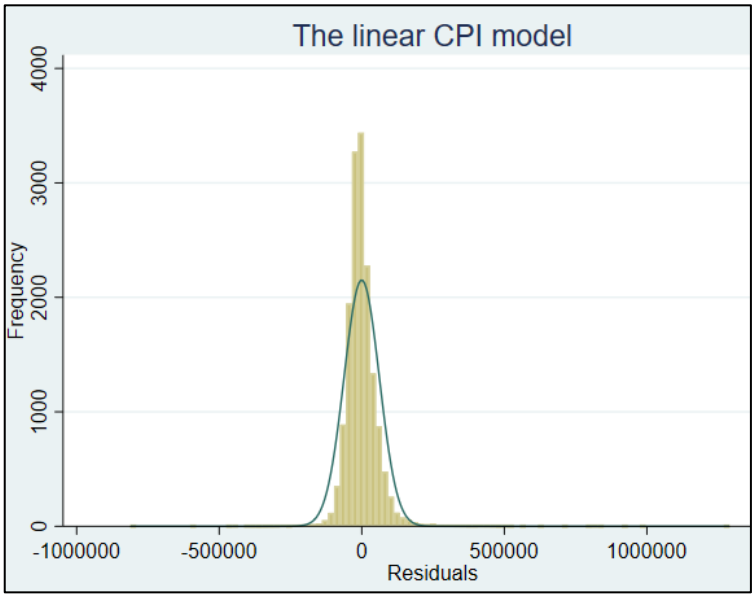
lags(*p*)	chi2	df	Prob > chi2
1	0.094	1	0.7587

H0: no serial correlation

Number of gaps in sample: 360

Durbin-Watson d-statistic(6, 1348) = 1.466048

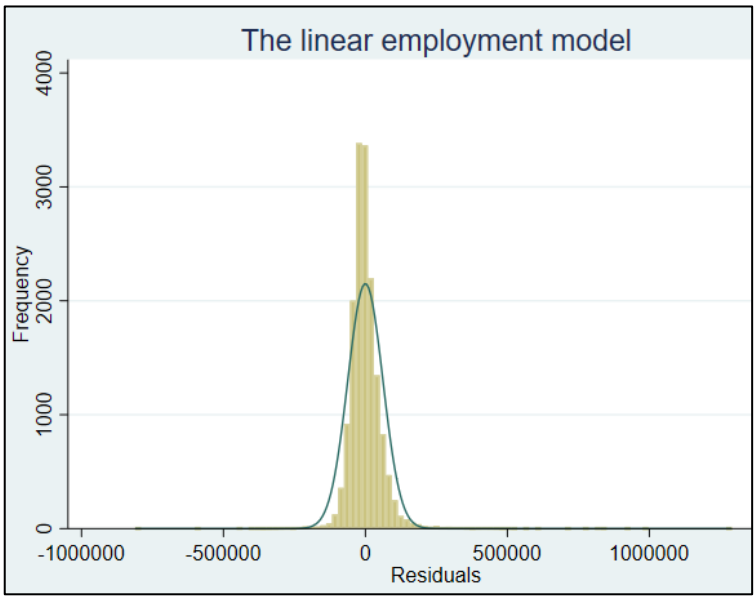
Appendix 13. Testing for normality in linear CPI model: histogram and Shapiro-Wilk-test



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res1	15,844	0.79285	1526.186	19.856	0.00000

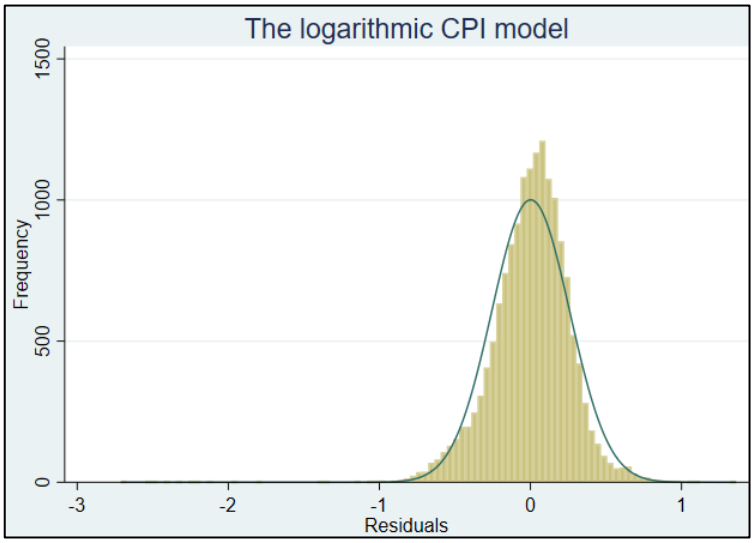
Appendix 14. Testing for normality in linear employment model: histogram and Shapiro-Wilk-test



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res2	15,844	0.79380	1519.179	19.843	0.00000

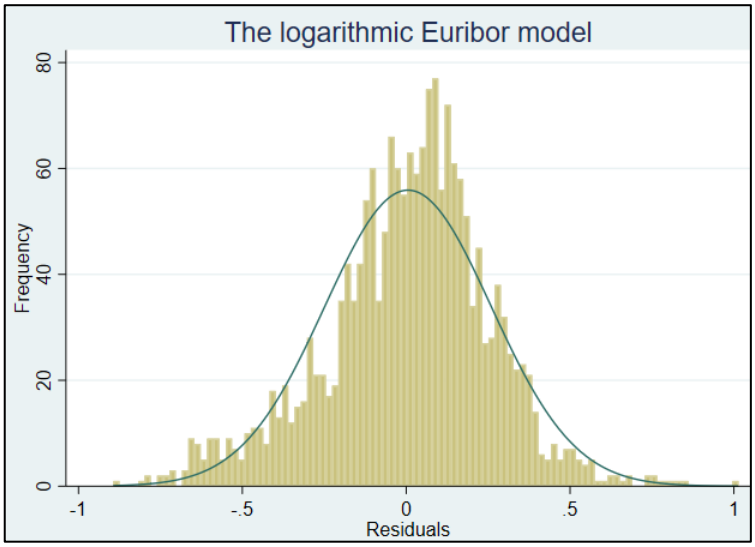
Appendix 15. Testing for normality in logarithmic CPI model: histogram and Shapiro-Wilk-test



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res3	15,843	0.96407	264.728	15.111	0.00000

Appendix 16. Testing for normality in logarithmic Euribor model: histogram and Shapiro-Wilk-test



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
res4	1,867	0.98588	15.724	6.991	0.00000

Appendix 17. Testing for multicollinearity in linear CPI model: VIF-test and correlation matrix

Variable	VIF	1/VIF
Area	1.05	0.949973
BuiltYear	1.12	0.892036
CentrumDist	1.15	0.873311
TramDist1	1.05	0.948848
Stage2		
2	10.38	0.096349
3	9.44	0.105964
StartedApa~3	1.55	0.645951
Euribor3	5.04	0.198229
CPI	18.25	0.054802
Mean VIF	5.45	

	Area	BuiltY~r	Centru~t	TramDi~1	Starte~3	Euribor3	CPI
Area	1.0000						
BuiltYear	-0.0628	1.0000					
CentrumDist	0.1293	0.3006	1.0000				
TramDist1	0.1747	0.0870	0.1641	1.0000			
StartedApa~3	-0.0159	0.0110	0.0174	0.0096	1.0000		
Euribor3	-0.0005	-0.0134	0.0362	0.0063	0.1055	1.0000	
CPI	-0.0173	0.0242	0.0587	0.0050	0.3400	0.7503	1.0000

Appendix 18. Testing for multicollinearity in linear employment model: VIF-test and correlation matrix

Variable	VIF	1/VIF
Area	1.05	0.949964
BuiltYear	1.12	0.892045
CentrumDist	1.15	0.873251
TramDist1	1.05	0.948420
Stage2		
2	5.96	0.167678
3	3.95	0.253095
StartedApa~3	1.78	0.563367
Euribor3	1.38	0.726527
Employment	5.47	0.182669
Mean VIF	2.55	

	Area	BuiltY~r	Centru~t	TramDi~1	Starte~3	Euribor3	Employ~t
Area	1.0000						
BuiltYear	-0.0628	1.0000					
CentrumDist	0.1293	0.3006	1.0000				
TramDist1	0.1747	0.0870	0.1641	1.0000			
StartedApa~3	-0.0159	0.0110	0.0174	0.0096	1.0000		
Euribor3	-0.0005	-0.0134	0.0362	0.0063	0.1055	1.0000	
Employment	-0.0243	0.0507	0.0546	0.0124	0.5425	0.3248	1.0000

Appendix 19. Testing for multicollinearity in logarithmic CPI model: VIF-test and correlation matrix

Variable	VIF	1/VIF
ln_CPI	1.18	0.844900
ln_Started~3	1.18	0.845785
ln_Centrum~t	1.16	0.864928
ln_BuiltYear	1.09	0.917930
ln_TramDist1	1.08	0.927218
ln_Area	1.05	0.949784
Mean VIF	1.12	

	ln_Area	ln_Bui~r	ln_Cen~t	ln_Tra~1	ln_Sta~3	ln_CPI
ln_Area	1.0000					
ln_BuiltYear	-0.0656	1.0000				
ln_Centrum~t	0.1526	0.2675	1.0000			
ln_TramDist1	0.1628	0.0886	0.2366	1.0000		
ln_Started~3	-0.0265	0.0179	0.0220	0.0152	1.0000	
ln_CPI	-0.0174	0.0255	0.0465	0.0160	0.3915	1.0000

Appendix 20. Testing for multicollinearity in logarithmic Euribor model: VIF-test and correlation matrix

Variable	VIF	1/VIF
ln_Centrum~t	1.19	0.843059
ln_BuiltYear	1.14	0.876675
ln_TramDist1	1.05	0.953567
ln_Area	1.04	0.963454
ln_Euribor3	1.00	0.998202
Mean VIF	1.08	

	ln_Area	ln_Bui~r	ln_Cen~t	ln_Tra~1	ln_Eur~3
ln_Area	1.0000				
ln_BuiltYear	0.0062	1.0000			
ln_Centrum~t	0.1459	0.3613	1.0000		
ln_TramDist1	0.1393	0.1058	0.1709	1.0000	
ln_Euribor3	-0.0054	-0.0318	0.0010	-0.0069	1.0000