INVESTIGATING THE SOCIOECONOMIC AND DEMOGRAPHIC PRINCIPAL COMPONENTS (SDPC) IMPACT ON THE REAL ESTATE MARKET PRICE IN FINLAND

(Focused Markets: Helsinki, Espoo, Vantaa, Tampere, and Turku, 2013-2020)

Lappeenranta-Lahti University of Technology LUT

Master’s Thesis

2022

Shima Edalatkhah

Examiner: Azzurra Morreale, PhD

Luigi Mittone, PhD

ABSTRACT

Lappeenranta-Lahti University of Technology LUT

School of Business and Management

Master's program, Business Analytics (MBAN)

Shima Edalatkhah

INVESTIGATING THE SOCIOECONOMIC AND DEMOGRAPHIC PRINCIPAL COMPONENTS (SDPC) IMPACT ON THE REAL ESTATE MARKET PRICE IN FINLAND (Focused Markets: Helsinki, Espoo, Vantaa, Tampere, and Turku, 2013-2020)

Master’s thesis, 2022

79 pages, 25 figures, eight tables, & one appendix

Examiners: Professor Azzurra Morreale, PhD

Professor Luigi Mittone, PhD

Keywords:

Real estate markets, Helsinki metropolitan area, Dwelling price, Per square meter price, Socioeconomic and demographic principal components, Housing price, Predictive model, PCR

This thesis investigates Finland's real estate market's socioeconomic and demographic principal components (SDPC) and their evolution in the five most populated cities, Helsinki, Espoo, Vantaa, Tampere, and Turku, from 2013 to 2020. The economy and demography evolution in the cities mentioned above has been investigated.

This thesis aims to create a predictive housing price model based on the socioeconomic and demographic factors influencing property prices in the focused cities. The Principal Component Regression model was applied in this thesis based on the study background and thesis goals.

The study results indicate that socioeconomic and demographic principal components significantly impact the property price in the focused markets. Accordingly, this study achieved its goal by developing a predictive housing price model that considers socioeconomic and demographic components. Running K-fold cross-validation, the result indicates that the predictive model with one principal component outperformed the other models.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Azzurra Morreale for her insightful review and recommendations during my thesis.

I also appreciate Professor Luigi Mittone and the entire LUT University community, particularly my professors. Thanks to the study program, I have extended my views and obtained outstanding support during my studies.

Moreover, my family is the inspiration for this master's thesis, with special thanks to my husband Mahyar for his unwavering support and belief in me throughout critical life challenges. For me, you are a never-ending source of inspiration.

In Espoo, on June 5th, 2022.

ABBREVIATIONS

AETM	Association for audio-visual & technology management
ANN	Artificial neural network
CBD	Central business district
CRT	Cube root transformation
GDP per capita	Gross domestic product per capita
GSEs	Government-sponsored enterprises
LT	Logarithmic transformation
ML	Maximum Likelihood
MREC	Mutual real estate company
MSE	Mean squared error
MSFEs	Mean square forecast errors
OLS	Ordinary Least Squares
PCA	Principal component analysis
PCR	Principal component regression
PSMP	Per square meter pricing
RMES	Root mean squared error
R^2	R-squared
SDPC	Socioeconomic and demographic principal components
SRT	Square root transformation
WLS	Weighted least squares

TABLE OF CONTENTS

1. INTRODUCTION

The real estate market trend has turned into one of the hottest titles in the press, specifically after Covid 19 Pandemic. This study investigates the impact of socioeconomic and demographic principal components (SDPC) on the Finnish real estate market price. The study is designed to meet the information needs of local and foreign parties interested in the Finnish real estate industry.

Accordingly, this research attempts to understand the dynamics in housing market prices in Finland, considering the target city's socioeconomic and demographic features. The socio-economic and demographic characteristics like the increase in population, average age, employment rate, income structure, tax, the share of foreign-language speakers, level of education, turnover of enterprises, etc., in that city have been taken into account based on the study literature. This study aims to develop a predictive housing price model based on principal components impacting socioeconomic and demographic characteristics in these cities. This study’s outcome will help possible interested parties such as real estate investors, buyers, business owners, and data analysts.

1.1 Background and motives of the study

Housing markets are significant in Finland. Following the Covid 19 pandemic, this market drew more investors and purchasers than before. The growth rate of the housing market price in Finland is the headline on different media. Finnish households are relatively leveraged, with mortgages accounting for most of their debt (PTT 2020.) Yet, according to Yle 2021, the real estate market in Finland’s capital region has heated up significantly. Since the third quarter of 2020, older detached house prices in cities with a population of over 100 thousand have more than 5.8% growth (Yle 2020.)

Dwelling prices in Espoo, Helsinki, and Vantaa rose over 4% in 2021, more than in any other year during the last decade. According to the Mortgage Society of Finland HYPO, Brotherus (2021), Tampere, and Turku are also experiencing a surge in real estate transactions. In 2020, the most significant overall profits were from residential, public use, and industrial assets (KTI 2021).

As cited in Helsinki Times and according to Huoneistokeskus, in January 2021, sales in Helsinki increased by 8.5 percent, in Espoo increased by 6.5 percent, while the number of sales in Vantaa, on the other hand, fell from the previous year. In Turku, the sales growth was particularly substantial (13 %). From January 2020 to January 2021, the real estate agency's sales escalated by 4.6 percent nationwide. (Helsinki Times 2021). The statistics above inspired the author to focus on influential factors in Finland's real estate market. According to the study background, there have been relatively few studies on the Finnish real estate market, most of them focusing on macroeconomic and economic variables influencing property market trends and investment in Finland. Thus, this study concerns Finland's real estate market's potential socioeconomic and demographic characteristics and expansion in large-scale Finnish cities such as Helsinki, Espoo, Vantaa, Tampere, and Turku from 2013 to 2020. The inspiration for the study stems from earlier research on the potential link between population transitions and property prices. For instance, Glaeser and Gyourko's research (2005) revealed an asymmetrical reaction in housing prices during population increase and decrease.

1.2 Thesis objectives and question

Earlier studies mainly focused on the influence of macroeconomic and economic variables on real estate investment and its price trends. This research aims to look at the principal socioeconomic and demographic components (SDPC) of the Finnish real estate industry. To have a homogeneous research market, this thesis concentrates on real estate in Finland's top five populated cities, including Helsinki, Espoo, Vantaa, Tampere, and Turku, with similar population densities and infrastructure, educational facilities, and dynamic construction trends. As a result, what differentiates this thesis from previous studies is:

- This thesis focuses on the Finnish real estate market's critical socioeconomic and demographic aspects regarding property price variations.

- This study concentrates on five of Finland's largest cities, including Helsinki, Espoo, Vantaa, Tampere, and Turku, from 2013 to 2020.
- The thesis aims to develop a predictive housing price model based on the potential socioeconomic and demographic principal components of the cities mentioned above.

Accordingly, the research question is formulated as follows:

1. Is it possible to create a predictive housing price model for Helsinki, Espoo, Vantaa, Tampere, and Turku based on socioeconomic and demographic principal components?

1.3 Limitations of the research

Given the small size of Finland's housing market and the low number of sales transactions in some areas, obtaining a reliable picture of the housing market's price trend and volatility is prone to statistical flukes (Aalto & Sinkkonen 2021). To avoid this and have a more realistic model, the data used in this study was confined to 2013-2020. To create a homogeneous market for study, it only investigated the housing markets of the five most populated cities in Finland rather than the entire country. The price of detached and terraced houses was excluded from the study due to many missing values, which might negatively influence the study's reliability.

1.4. Theoretical framework

The dominant point of this master's thesis is examining the Finnish real estate market price drivers. The theoretical section explores the elements that influence real estate market prices in other countries and Finland. Also, learning about Finland's economic, demographic, and real estate market structures and context.

As a result, the theoretical framework is based on three pillars. The research literature's general influential factors on property market price are first reviewed. Second, the socioeconomic and demographic elements that impact the real estate market are addressed.

Third, Finland's economy, demographic, and real estate market dynamics have been thoroughly researched, focusing on the target cities.

Figure 1. The thesis' theoretical framework

1.5 Thesis structure

This research consists of five main chapters, shown in Figure 1 as a progressive thesis. The thesis introduction and objectives are presented in the first chapter, followed by the question of research and study delimitations. The second chapter concerns the literature review and focuses on the background of the study. The first part of chapter two reviews the structure of Finland's economy. The second section focuses on the property market of Finland, including the trends, players, and framework of the property market in the mentioned Finnish cities. The third part focuses on the theoretical background.

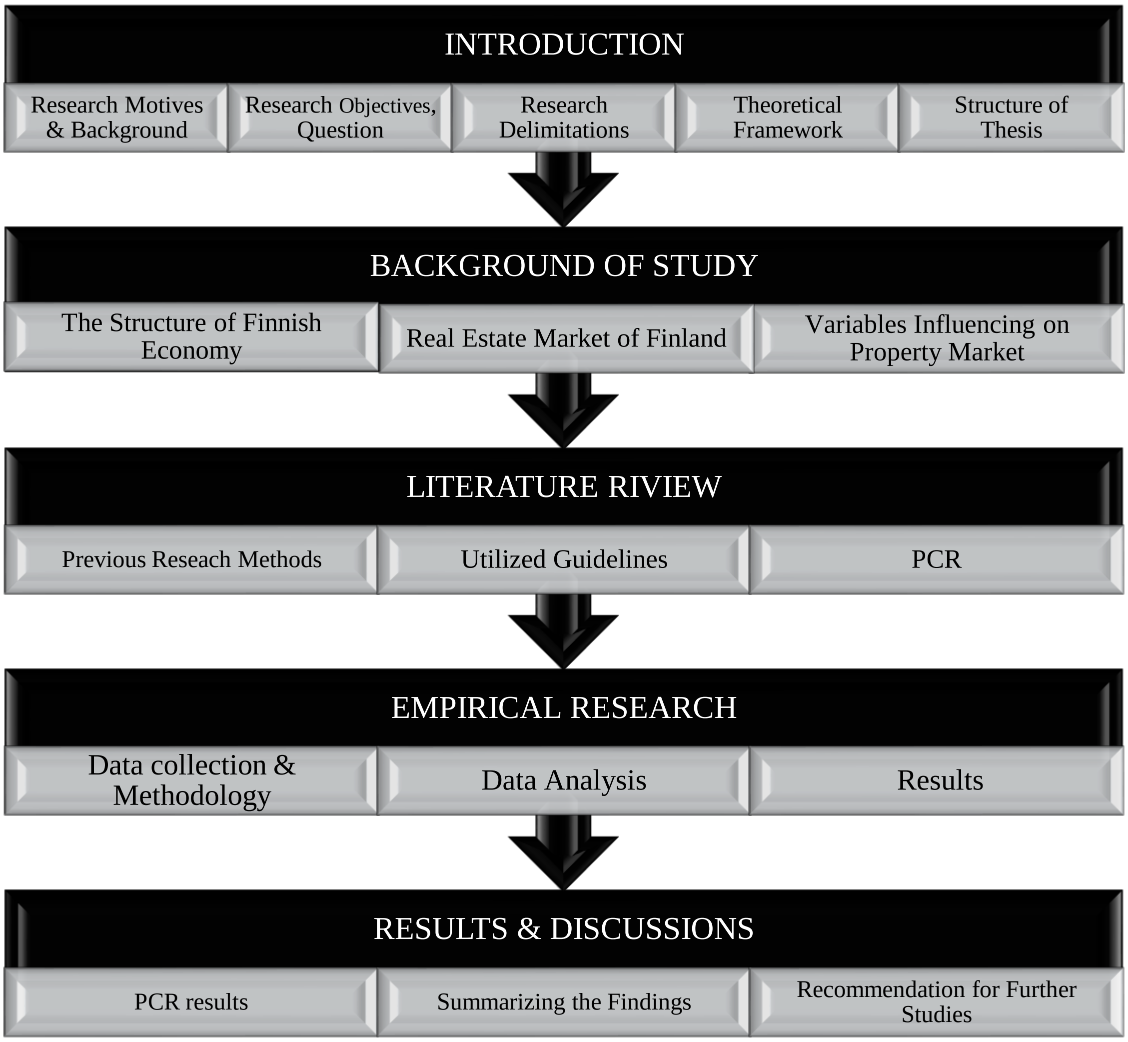

Figure 2. Thesis structure

Chapter 3 examines the theis literature, including the previously applied methods, and summarizes the guidelines as a benchmark. The last part reviews the PCR analysis method. Chapter four is specified empirical research. The first part describes the data collection, selected variables, and the methodology. In the second part, descriptive statistics and analysis are presented. In the final part, the findings and limitations are addressed. The thesis' conclusions are outlined in Chapter 5, along with recommendations for further research.

2. Background: Real estate market in Finland and the affecting factors

This chapter covers earlier research regarding Finland’s property market and the impacting elements on its price in many nations, including Finland, covering socioeconomic and demographic characteristics. The structure of Finland's economy, demography, and real estate market were then thoroughly investigated.

2.1 The structure of the Finnish economy

This part investigates the structure of the economy, GDP, and unemployment rate in Finland. Finland's economy is open and export-driven, and fluctuations in exports significantly impact the country's economic growth. (Bank of Finland 2021).

2.1.1 Finnish economy, GDP, and the unemployment rate

The overall value of Finnish exports climbed to €96 billion in 2019, accounting for about 40% of the country's GDP. In 2020, the COVID-19 epidemic affected Finnish exports. According to preliminary statistics from the Finnish Customs, the total number of commodities exported in Finland has decreased by almost 12%. As Figure 3 illustrates, based on the Statistics Finlands database, 51.1 percent of the Finnish economic structure is provided by private services. The proportion of public services is 18.3%. Interestingly, Construction has a 7.5 percent contribution.

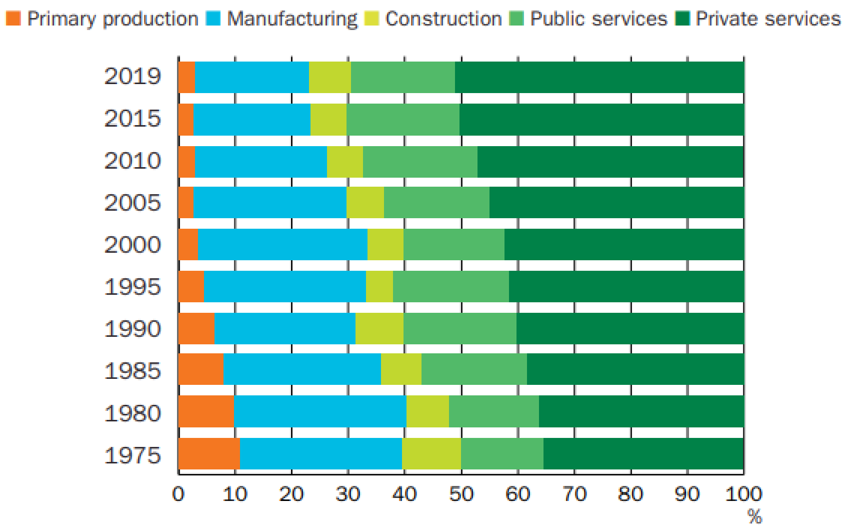

Figure 2. Finnish economic structure (Statistics Finland 2021)

On the other hand, Industrial production accounts for 20,5 percent of the economic structure. Primary production output represents only 2.7 percent of the contribution to the Finnish economy. (KTI 2021.) In 2019, the Finnish economy grew by approximately 1%, declining from last year's 1.7% and the lowest growth over four years (Lalaine 2022).

According to preliminary figures from Statistics Finland, the Finnish GDP declined by 2.9 percent in 2020. The GDP per capita in 2019 was €43 567. Following a severe drop in the second quarter, the economy recovered in the second half of the year, with private spending and exports driving the rebound. (World Bank 2021.)

Finland's unemployment rate increased from 6.7 percent in December 2019 to 7.9 percent in December 2020 after declining for several years due to solid economic development, as illustrated in figure 4. (KTI 2021.) By July 2021, the unemployment rate had fallen marginally to 7.6%. (Statistics Finland 2021.)

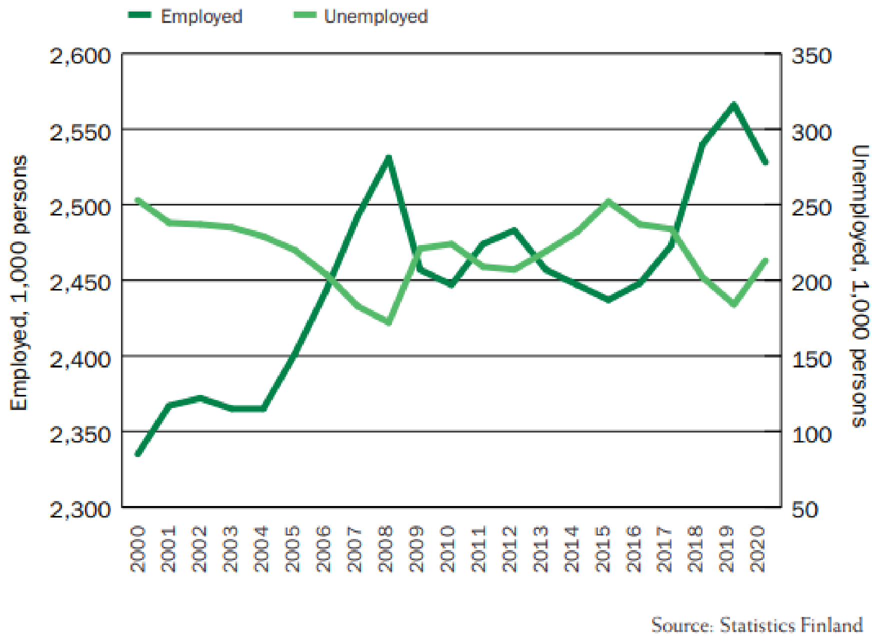

Figure 3. The employment rate in Finland (Statistics Finland 2021)

2.1.2 The structure of population and demographics in Finland

According to the latest UN statistics from Worldometer, Finland currently (as of Tuesday, 19 October 2021) has a population of 5,551,908 (Worldometer 2021). The population has grown by some 0.3 to 0.4% annually over the past ten years. Yet, the growth rate has declined in recent years. Currently, the death rate exceeds that of the births as the main strike of covid 19 pandemic. By 2020, the pandemic COVID-19 slowed down immigration, with a marked decrease from previous years, both in immigration and emigration. Finland is a relatively homogeneous country with a Finnish ethnic majority. Despite recent increases, the percentage of foreigners in the country remains among the lowest in the EU. Only 7.5 percent of the population did not speak Swedish or Finnish as their first language in 2019. Russian, Estonian, and Arabic speakers make up the largest groupings of foreign language speakers. (KTI 2021.) According to statistics Finland, Vantaa with 1,5% had the highest population growth in 2020 among the cities mentioned above, while Helsinki, with 0,5%, and Turku with 0,7%, had the lowest population growth. (Statistics Finland 2021.)

Finland is being struck harder than most other countries by the challenges of an aging population and a declining dependence ratio. Labour input does not rise as the population ages. Hence economic growth will be dependent on increased productivity. The pandemic's increased public debt burden also puts strain on the forecast. As a result, the GDP's long-term growth potential will be low, averaging 1-1.5 percent per year. The generations of baby booms are retired, increasing the overall dependency ratio, the total under 15 and over 65 years of age. The overall dependency proportion is nearly 62 and will rise to almost 65 by 2030. (KTI 2021.). Their share over 65 population is forecasted to rise to around 29% by 2050 (Findicator 2021).

The latest update in August 2021 shows us that the capital city, Helsinki, has a population of 656 920. There are 13 neighboring cities in the so-called Helsinki metropolitan area (Uusimaa), the largest of them being Espoo and Vantaa. This area has a population of 1,52 million. Espoo is the second most populated city in Finland with 292,796 inhabitants, followed by Tampere with 241,009 and Vantaa with 237,231 inhabitants. (Statistics Finland 2021.) Oulu is Finland's fifth-most populous city in northern Finland and is quite far from other cities (607 kilometers from Helsinki). With 207,327 inhabitants in 2020, Oulu remains out of the focused market of this thesis (Statista 2021.)

Table 1. Population in largest cities (Statistics Finland 2021)

	Municipality	Population 31.12.2020	Increase of population 2019-2020	%
1	Helsinki	656920	3085	0,5
2	Espoo	292796	3065	1,1
3	Tampere	241009	2869	1,2
4	Vantaa	237231	3456	1,5
5	Oulu	207327	1838	0,9
6	Turku	194391	1429	0,7
7	Jyväskylä	143420	1020	0,7
8	Kuopio	120210	928	0,8
9	Lahti	119984	161	0,1
10	Pori	83684	-250	-0,3

The sixth most populous city in Finland is the former capital city, Turku, with a 193,000 population that is located in the southwestern center of Finland (City of Turku 2021).

2.1.3 Public finances & security funds in Finland

The central government, municipal governments, and social security funds are Finland's general government. (KTI, 2021.)

Table 2. Finnish economy key figures (Statistics Finland 2021)

	2014	2015	2016	2017	2018	2019	2020	2021**	2022**	2023**
GDP(change in vol), %	-0.4	0.5	2.8	3.2	1.3*	1.3*	-2.9*	-2.5	2.0	1.4
Change in exports, %	-2.0	0.4	3.9	8.8	1.4*	6.7*	-6.3*	5.0	4.6	2.2
Inflation, %	1.0	-0.2	0.4	0.7	1.01	1.0	0.3	1.0	1.4	1.6
Unemployement rate, %	8.7	9.4	8.8	8.6	7.04	6.7	7.8	8.0	7.6	7.2
Private consumption, %	0.6	1.8	2.4	0.8	2.0*	0.9*	-4.9*	3.8	2.5	1.8

*Estimate **Forecast

Since the global financial crisis, Finland's general government finances have been in deficit (Valtioneuvosto 2021). As we can see in table 2, the deficit shrank to 0.7 percent and 0.8 percent of GDP in 2018 and 2019, respectively, thanks to the improving economy. Considering the Covid-19 pandemic, the mismatch between government spending and revenue will likely persist, especially as the population ages and imposes more pressure on government expenditures. (KTI, 2021.)

2.2 The real estate market in Finland

Real estate's physical features such as immobility, indestructibility, and uniqueness differentiate it from other assets. Economically speaking, real estate has four primary characteristics: scarcity, improvements, the permanence of investment, and location or area performance. There are five real estate types: residential, commercial, industrial, land, and particular purpose real estate. The latter is used mainly by the public, like public buildings, libraries, parks, schools, etc. (Chen & Scott & Logan 2021.)

2.2.1 Real estate market's players and structure in Finland

In March 2020, the breakout of the COVID-19 outbreak impacted negatively on the Finnish real estate market. The transaction was 5,6 billion euros, 13% less than the previous year. (KTI 2021.) The following parts discuss the Finnish real estate market players and their structure.

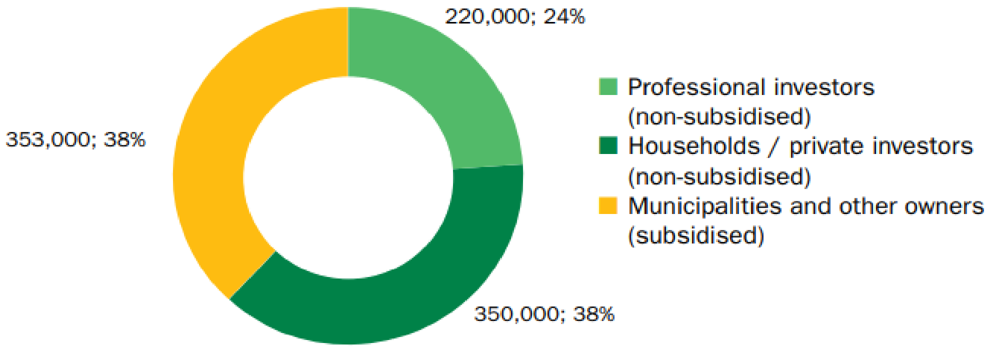

Figure 4. The ownership structure of rental flats for the year 2019(Statistics Finland 2021)

Heka Oy is one of the most prominent players, with approximately 50,000 housing units in this segment, owned by Helsinki. The city of Espoo owns about 15,000 housing units. The Turku, Vantaa, and Tampere are major municipal enterprises with 9,000-11,000 homes. In the current list of Finland's major property investors, Kojamo and SATO are ranked first and second, respectively. In 2020, there were about 600 completed or purchased housing holdings in Kojamo, and around 70 apartments were sold or demolished. Other significant players in residential rental are the leading pension funds such as Varma, Ilmarinen, Keva, and Elo. (KTI 2021.)

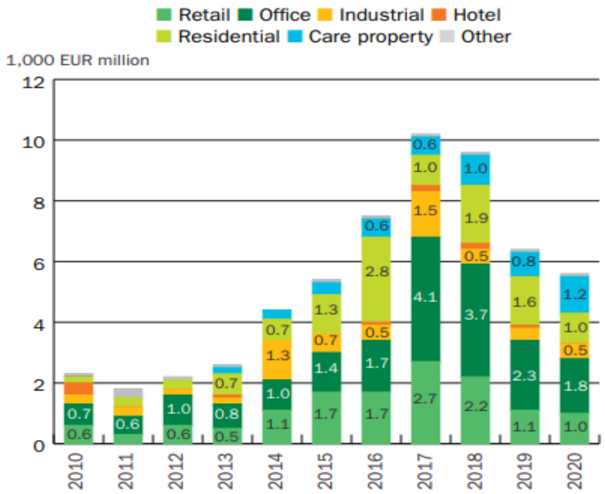

Figure 5. Transactions volume by property sector 2010-2020 (KTI 2021)

As figure 6 illustrates, offices with 32% of transactions were the most frequently traded during the last four years, followed by public use assets. Despite declining significantly over the previous couple of years, residential properties with an 18 percent trade share stand as the third most popular, followed by retail properties. In 2020, the total value of retail property sales was expected to exceed €1 billion. (KTI 2021.)

Finnish properties stood at the end of the year at €27 bln, equivalent to 33 percent of the total market. Their holdings totaled €25 billion at the end of 2019. (32 percent). Net sales of foreign investors totaled €1,5 billion in the transactions market.

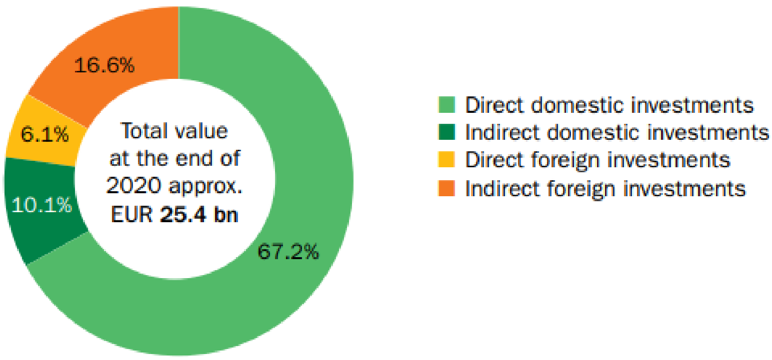

Figure 8. The structure of investment in the Finnish real estate market (KTI 2021)

Finnish institutional investors made 17.1 billion euros in direct domestic investments in 2019, up from 16.8 billion euros at the end of 2019. Institutional investors remained net sellers in the transactions market. Properties make up 9% of Finland's investment portfolios of pension schemes. Direct domestic investments account for 67 percent of the property of pension institutions. (KTI 2021.)

House prices have steadily grown in importance in the investment industry, presently accounting for 32% of all assets. Because of active new development and growing market values, the residential sector has grown. Figure 9 in the appendix shows the sector's Finnish property investment market structure. In 2020, the proportion of office buildings remained steady. The Covid-19 strike had a more significant impact on retail property. (KTI 2021.)

Figure 10. Total property sector returns in 2020 (KTI 2021)

However, even within the retail sector, the desirability of various types of retail property sectors varies significantly. In 2020, the most considerable overall profits were from residential, public use, and industrial assets.

2.2.2 Real estate market in Helsinki, Espoo, Vantaa, Tampere, and Turku

Because of the high turnover rate in the capital region, the number of properties has increased from the previous year, and the number of available properties has decreased. Hypo's top economist, Juhana Brotherus, stated. According to Helsinki Times (2021), at the beginning of 2021, there were 15% fewer dwellings available for purchase compared to the same time interval in 2020 and 2019. Hypo also predicts that house price growth would slow in the capital region in 2022 to around 2.5 percent, bringing down the national average. (Helsinki Times 2021).

Helsinki

Helsinki, with 560,000 employments available, follows a master plan based on a concept of greater inner-city development density and the zoning of additional areas for residential use. The expanding Pasila-Vallila-Kalasatama axis and neighborhoods in and around the CBD (central business district) are regions with rising commercial property supply. The Helsinki central business district is Finland's most prominent office and retail property sector. The most influential public sector administrative operations and cultural buildings are likewise housed in this physically limited region. According to the KTI rental index, rates in new office agreements declined somewhat in 2020, by 1.3 percent, and the occupancy rate fell to around 92 percent. (KTI 2021). The lack of the costliest buildings and premises in the KTI sample contributed to the drop in prices. In recent years, the rise of e-commerce and changing customer behavior have posed challenges to retail rental markets. Due to increasing remote working and travel and other constraints, the number of visitors to the city center fell substantially, reducing sales in the CBD's stores, restaurants, and hotels. In Jätkäsaari, several hotel developments are being built or planned. The Pasila region continues to see significant growth. In Pasila, the Pöllölaakso area will be developed for residential use. Kalasatama is still undergoing residential and commercial development. (KTI 2021.)

Espoo

The commercial property stock in Espoo, Finland's second-largest city, is dispersed. The metro of the west line serves Tapiola and Matinkylä-Olari. When the extension is finished in 2023, the metro will also be linked to the fifth center, Espoonlahti. Espoo's commercial property portfolio is diverse, as seen by the range of rental rates and vacancy rates in various districts and properties. (KTI 2021.) The office supplies in regions with metro access are primarily modernist. In the new station areas, retail and residential building is underway. Keilaniemi is a contemporary office neighborhood with numerous large corporate headquarters and innovative business park designs. Rents in the region have risen in tandem with the upgrading of the supply. According to the KTI database, average office rentals in the region were over €25 per sqm in 2020, with a high occupancy rate of 95%. Keilaniemi's current office stock draws tenants and investors. The Tapiola region also has a diverse office stock, with many older buildings being renovated or demolished and rebuilt for different uses. There are 84 retail malls Espoontori and Entresse are all located in Espoo's downtown. Residential development activity is still robust in Espoo districts. (KTI 2021.)

Vantaa

The most important commercial real estate market regions in Vantaa are centered around the airport and its environs and the old Tikkurila center, which also houses the municipal government. Tikkurila is Vantaa's major urban center and the most prominent office and retail district (and the airport area). The majority of the city's government buildings are likewise located in Tikkurila. Vantaa's infrastructure is being developed, and numerous property owners are making fresh investments in the region. There is still a lot of residential rental development in the area. Along with the Ring Rail Line, Kivistö is near Ring Road III and the Hämeenlinna highway. Kivistö is a fast-expanding region. In recent years, the area has seen a lot of residential construction. (KTI 2021.)

Tampere

As a popular student city, Tampere has a dynamic real estate market. In the Nordic countries, Tampere is the largest inland city. Tampere is a historic industrial city that has lately attracted high-tech and service-oriented enterprises. Tampere has a higher-than-average unemployment rate than the rest of the country. Tampere University and Tampere University of Applied Sciences have around 4,000 employees and 30,000 students. Tampere comprises eight cities with a combined population of over 400,000 people. Tampere alone is anticipated to have more than 250,000 people by 2030. (KTI 2021.)

Turku

Turku is a solid conventional university city home to Finland's primary Swedish-language university. Businesses surrounding the marine cluster and biotechnology are two of the region's current competency areas. The region's metal industries are bolstered by a thriving shipbuilding industry, with Meyer Werft leading participant. The Turku area is home to half of Finland's medical industry, and the healthcare sector employs more than 5,500 people. Valmet Automotive employs 6,000 workers in Uusikaupunki, around 60 kilometers north of Turku. Due to the COVID-19 epidemic, the shipbuilding cluster will lay off employees in

2020. A new University hospital facility is also being built at Kupittaa, with an expected completion date of 2021. Residential rentals continued to rise during 2020 in Truku. (KTI 2021.)

2.2.3 Taxation and VAT on real estate

In Finland, the real estate tax is estimated based on the assessed price of the property. Generally, this value is slightly below the property's market value. The revenues go to the city where the property is located. Forest land or agricultural land is excluded from property tax. Parliament shall regulate the minimum and maximum tax rates, and municipalities shall follow that accordingly. The tax rate range is between 0.93% and 2.0% of the value of the property for the year, which is on average 1.11% in 2021. This is while Tax rates may vary from 0.41% to 1.0% for permanent residences. At the moment, the average rate is 0.51%. There is also can a special tax rate for unbuilt lots that may vary from 2.0 to 6.0 percent. (KTI, 2021.) The share of operational costs in Helsinki in 2019 is illustrated in figure 11, available in the appendix. The total transaction prices cover the tax base, including the actual selling price and the mutual real estate company's potential debt. (KTI 2021.)

In Finland, the supplier of goods or services usually is responsible for paying VAT. Nevertheless, a reverse charge mechanism was applied to reduce the potential VAT fraud risk in the construction sector. VAT is the responsibility of the buyer of construction services. This mechanism applies to companies that continuously offer construction services. (Vero 2021.)

2.3 Theoretical background: variables and hypotheses influencing Finland's real estate market

Extensive research has been undertaken on the real estate market, with Finland having a modest share. The following section summarizes influential elements in the real estate market, including socioeconomic and demographic variables in both Finland and international markets.

2.3.1 Macroeconomic influential factors on Finland’s real estate market

When we focus on the property market of Finland, numerous factors might be classified as influencing house prices on a local level. However, some believe that national variables explain the evolution of home prices entirely, implying that the location of a housing market is irrelevant. Regardless of region, national influences are either rising or reducing house prices practically synchronously. As cited by Kokkinen (2020, pp 16-17), many researchers such as Abraham and Hendersat (1996), as well as Hort (1998) and Oikarinen (2007) pointed out that interest rates, inflation, and building prices all have an impact on the property price volatility. National variables that impact home price development include household debt. The other notable variable, as pointed out by Oikarinen (2009a), is the property price disparity across the country and government tax deductions. Housing market research is undertaken chiefly on a national basis due to the unavailability of data from local housing markets. There isn't a lot of global academic study on local housing market development, and much less in Finland. A few online research focuses on the fluctuations in home prices in the Helsinki area. (Kokkinen 2020.) Other studies show how the development of property prices in one local area is reflected in other local regions of Finland. The study demonstrated that Helsinki, Finland's central economic hub, will continue to be the country's guiding city. Changes in pricing in the Helsinki metropolitan region cause price increases in other local and provincial centers. (Oikarinen 2004.)

Urbanization, shifting supply and demand

Urbanization is transforming Finnish society in unimaginable ways: the country's regional structure is changing unprecedented in decades. The most influential element in the growth of the housing market is urbanization. Demand for housing is increasing in expanding areas, and there are many new-build buildings, resulting in rising home costs. On the other hand, the housing market in the recessive regions is deteriorating. (Keskinen et al. 2020.) The most critical factor influencing housing demand is the price of homes. As prices drop, so does the supply of housing, according to the law of demand. Housing cost is a crucial component in determining supply. As the price rises, the quantity supplied rises with it.

Changes in material prices and technological advancements can influence the construction and availability of new properties. The housing market equilibrium determines the amount and price of houses exchanged. Economists commonly believe that a highly competitive market will eventually attain balance. There will be a mismatch between supply and demand if the price differs from the equilibrium price. This encourages buyers and sellers to behave uniquely. The demand will outnumber supply if the price is lower than equilibrium. If the cost is excessively high, providers may be tempted to try to decrease prices; parallel to that, customers may hunt for better bargains. Empirical market data back up the market projections. When buyers and sellers meet independently to negotiate a surplus price, the outcome is frequently close to "market equilibrium." In general, the supply-and-demand paradigm produces good forecasts regarding price fluctuation. (Saylor Academy 2012.)

In their study regarding the housing market and forecasting prices, Augustyniak et al. (2015) mention that multiplier effects, even minor changes in core aspects, such as interest rates, result in demand shocks. Positive demand changes are difficult to meet since supply is constrained in the near run. In Finland, urbanization is concentrated in a few places, mainly in the metropolitan region and Tampere. In 2019, the urban area and Tampere had the fastest price increases. (Keskinen et al. 2020.)

Monetary policy and the area’s economy

The monetary policy, mortgage reference rates, and other mortgage parameters can impact housing demand, which changes the cost of borrowing. Furthermore, mortgage deductions or transfer taxes might impact the property market. (Lindblad et al.2019.) Municipal zoning choices have a significant impact on the property market. Changes in building laws or land taxes, labor availability, productivity changes, fiscal policies, and increased competition in the construction industry may affect the construction. (Lindblad 2019). Regional disparities have influenced the size of mortgages held by homeowners at property prices. Mortgages are soaring and have increased in value both in euros and concerning income level, particularly in metropolitan regions where home costs are higher and regional demand is more substantial. (Putkuri 2018.)

The metropolitan region's population and a few other university cities rapidly increase. Although population growth is modest in certain places, it is dropping in most areas, as seen by housing development. (Kokkinen 2020). Despite the development of the economy, the property market is just rising in the fastest-growing locations, while prices in small towns are declining (Kokkinen 2020; Manninen 2021). In Finland, the metropolitan area and Tampere saw the fastest price increases in 2019. Population growth is concentrated in these regions, evidenced by a more significant increase in price development. (Keskinen et al. 2020.) Previously, banks did not face any risks while issuing mortgages because, over time, home values in Finland rose in all locations. The imbalance value of the property might have a detrimental impact on the labor movement, negatively impacting the labor market. A rapid drop in property prices might result from shocks to the property market, such as depression or a recession. (PTT 2020.) Hornstein (2009) mentions that the market for single-family housing is divided into two parts: Existing and new dwelling demand. Modifications in dwelling demand diverse effects on the whole economy. An existing home's location and neighborhood features cannot be readily recreated. The growth in house prices is predicted to be concentrated in the central cities of developed areas over the 2020 decade.

2.3.2. Socioeconomic and demographic impact on the property market

Population, age, households, and property size

Demographic changes and the kind of home chosen impact total housing demand, especially in the long run. (Puturki 2018). In the early 2010s, a shift in price development occurred in regions with differing demographics (PTT Variables impacting housing demand and desired home style include demographic shifts and age structure. In places with older adults, smaller homes and apartments, rather than larger dwellings far from amenities, may be more common. Instead, the area's growing working-age population (20-64 years) maintains high demand (Lindblad 2019). Consumer Confidence Indicators from Statistics Finland, for example, can be used to identify demand-side constraints. The number of households interested in buying a property has increased. (Putkuri 2018.) Within areas and cities, polarization may be reflected in property prices. In Helsinki's most costly regions, housing prices continue to rise across the metropolitan area. Price differences between Espoo and Vantaa and Helsinki are becoming increasingly noticeable. New-build building, particularly in Vantaa, has kept price growth mild.

Between 2015 to 2019, the price of studio apartments throughout Finland rose by less than 2%, whereas in Helsinki increased by 18%. In addition, between 2015 and 2019, the price of a one-bedroom flat or larger climbed 10% higher in Helsinki than in other cities. Housing prices have risen at the most remarkable rate in Helsinki during the past ten years, overtaking income growth. (Keskinen et al. 2020.) However, in 2020, when the world was struck by the covid 19 epidemic, in Finland, the most significant housing price increases did not occur in conventional city center regions. Instead, there were plenty of surprising risers in the suburbs and beyond capital cities. Housing price growth occurred throughout Finland's western half, not merely in the Helsinki metropolitan area. Meanwhile noticeable share of the population chose to live on the outskirts, buying cottages or around lake villas rather than a dwelling in the city center. (Aalto & Sinkkonen 2021)

In their analysis, Levin, Montagnoli, and Wright (2009) found that population decrease and population aging placed downward pressure on prices; therefore, the long-run trend of growing real housing prices cannot be projected to continue in the future, particularly in future Scotland. Mankiw and Weil 1989, as cited in Levin, Montagnoli, and wright (2009)'s article, had discovered in their study on the US market that the demand for housing climbs dramatically until around the age of 30, then becomes essentially flat around the age of 40, and subsequently drops afterward. Peng (2017) investigated the effect of demographic changes on housing prices, especially the dependency rate. According to the empirical findings, housing prices are cointegrated with fertility and aging dependence rates. A rise in the fertility rate raises housing values in the long run. An increase in the old dependence rate, on the other hand, lowers housing values. The predicted demographic change in 2015 is a crucial indicator of future house price changes.

In another interesting study about US real estate market, Feng, Kim, and Lee (2018) investigated the potential influence of demographic shifts on US housing markets from 2010 to 2017 during and after the recession crisis. Their study results show that housing prices rise slower when the population grows than when prices fall when the population declines. When the population grows, housing price fluctuations become more elastic due to the endurance of housing stock.

Migration

The recent worldwide trend toward nationalism shows that the dramatic upward migration flow is a critical socio-economic challenge. According to Larkin et al. (2018), one important source of concern is migration's influence on dwelling availability. Provincial demand is modified through migrations. As a result, there is a clear distinction between Helsinki vs. the whole country. Outside of the broader Helsinki region, the working-age population will continue to decline in the 2020s. (Saifi 2021, p 23) As cited by Manninen (2021), the real estate market in Helsinki and other capital cities is expected to grow during the following years. (Putkuri 2018.) While immigration isn't the sole factor impacting property prices, it does contribute to the already high demand. For every three migrants, an additional dwelling is required. Migrants contribute to the economy's growth and compete for housing (Khadem 2021.)

In their research, Barbu, Vuta, Strachinaru, and Cioaca (2017) and Björklund (2019) highlight the association between variations in property prices as assessed by the housing price and migration patterns and other macroeconomic factors. This association was studied in more than 20 countries from 2007 to 2014. Researchers discovered a positive connection between the expansion of the home value index and migration, the percentage of market capital in GDP, and the pace of economic growth using panel data models. According to the findings, property prices rise by 0.045 percent for every one percent increase in migration. Although not significant, it demonstrates migration and home prices have a positive correlation. (Barbu et al., 2017.) According to Larkin et al. (2018), migration boosts home prices on average, based on a database including over 470 assessments of the influence of migration on house prices in 14 countries. However, sentiments toward immigrants have a moderating influence on this effect. Their study result shows that house price rises are more controlled in nations that are less welcoming to immigration.

Green (2018) uses data from 80 local authorities in England and Wales from 2010 to 2016 to investigate how immigration affects another avenue via which actual income and wealth might shift. The research's findings imply that immigration has a negative impact on property values. The study proves that native outmigration reaction to immigration flows is a significant element driving this outcome. The results are resilient to excluding London from the sample. Immigration has also been a modest influence on housing availability while decreasing crime within a local government. As a result, Green (2018) suggests that, while immigration severely affected locals through property prices in the years following the financial crisis, with net migration anticipated to reduce after Brexit, UK homeowners may benefit from a wealth effect in later years.

Education level, educational facilities

Feng investigates the impact of education on the Beijing rental housing market in his master's thesis. According to the data, first-class universities have a considerable impact on rising property prices, but second and third-class universities have no such influence. The study's findings show a positive relationship between property prices and higher education. However, the results vary in Beijing’s zones. (Feng et al. 2018.) The cost of living near a high-scoring school is 2.4 times the cost of dwelling near a lower-rated school. (Khadem 2021). Another study looked at the capitalization impacts of educational facilities on property prices in China, and the findings revealed that schools' locations had a strong influence on housing prices. The results also indicate that housing values in the area benefit from the proximity to high schools and colleges. The findings highlight the relevance of academic centers in China's property market in the Hangzhou area. (Wen & Zhang 2013.)

The other research on the housing rental market in China demonstrates the influence of higher education institutions and universities on the rental market has a positive and substantial association. Metz (2015) examines a study that considers elementary, middle, and high schools based on location. At the same time, most research on the association between learning and house price looks at the change in home value due to the level of education. The findings provide numerous vital insights. First, as the distance to educational centers and schools grows, so do housing costs, as predicted. The results show that parents favor residences within walking distance of an educational facility or school. The outcomes also demonstrate that shopping areas are commonly located near high schools, and residents value the convenience of being able to walk to these facilities. (Metz 2015.)

Enterprise’s turnover, income level, and the unemployment rate

Feng, Kim, and Lee (2018) examined factors such as stats of unemployment rates, growth of Gross Domestic Products, and mortgage rates in their research on housing prices using the regression model and new housing supply, represented by the number of construction licenses. Lowering the unemployment rate, rise in GDP, and declining mortgage rates all significantly influence housing demand and, as a result, property prices. (Feng, Kim, & Lee 2018). Hornstein's (2009) study results show that the opportunity to receive a loan is influenced by household credit history. Hornstein elucidates his point with an example that when family income rises, there will be greater demand for a mortgage; meanwhile, people’s saving behavior modifies. Higher savings rates allow them to finance home and enter the housing market earlier, which increases homeowner demand. (Hornstein 2009, p2) Likewise, the case of families paying less in advance for homes leads to an increase in housing demand. The growth of government-sponsored enterprises (GSEs) has altered the occupied housing market. GSEs purchase loans that fulfill specific criteria and then sell the securities they underlie. They've turned mortgages into a "commodity," which has reduced mortgage rates for homeowners..(Horstein 2009.)

Jin and Tang (2018) investigated the statistical link between metropolitan housing prices and the factors that influence them. They used real estate data for 90 cities in 2016 and the mean annual income for the population for the previous 17 years to explore the link between real estate prices and impacting elements in China. They discovered that the factors influencing property costs are the overall attractiveness of cities, their GDP, and their liveability. The city's GDP is essential, and interest rates are inversely connected with house prices. (Jin and Tang 2018)

Real estate properties represent the development potential. The density of social, educational, and cultural prospects, the region's living conditions, and commercial and industrial enterprises all influence the value of real estate property. Population density and social classification must also be considered for actual real estate valuation. (Nisanci, 2005.)

3. Literature review

3.1. Previous research methods

The study background shows that many methodologies are utilized regarding price prediction models of the housing market. No single model consistently outperformed the others because the amount of data, time frame, and variables varied between experiments. The real estate evaluation method is defined as determining the genuine value of real estate using experience (Dale and McLaughlin, 1988). The approach used is determined by the location and advertising parameters of the real estate. The following three are the most often used approaches in real estate evaluation.

- The Sales Comparison method
- The Income Method
- The Cost Method (Aclar and Cagdas, 2008)

The sales comparison method

The sales comparison techniques calculate the real estate value based on the actual transactions. The evaluated and compared real estate must have the exact attributes in this procedure, such as area, location, topographic features, social and economic aspects, transportation amenities, etc. This strategy is used to sell a property, land, or workplace, among other things. Aclar and Cagdas (2008.)

The income method

The income approach assesses the value of real estate based on the revenue generated. This strategy is utilized to rent residential, commercial, and recreational properties. (Aclar and Cagdas, 2008.)

The cost method

The cost approach relies on the building cost. Suppose the real estate valuation considers the structure's construction cost and other economic factors. This calculated value is used to calculate the expenses of depreciation, maintenance, repair, etc. (Nisanci, 2005).

Reviewing the literature of study, we can see that housing prices are forecasted using a variety of approaches. OLS and ML were applied by Dubin (1998), while Townsman, Payton, and Man likewise used OLS (2008). As cited by Manninen (2021), for housing price forecasts, Weighted Least Squares regression (WLS) and Artificial Neural Network (ANN) models were investigated by Limsobanchai, Gan, and Lee (2004). Regarding the analysis methodologies regarding housing price predictive models, Dubin (1998) found that Housing prices were predicted more accurately by ML regression than by OLS regression.

The ANN approach outperforms more substantially in the prediction of property prices than the WLS, according to Limsobuchay. (2004); On the other hand, they noted that the Hedonic pricing model's low performance might be due to a lack of environmental factors and a limited amount of data. They also utilized market prices instead of purchase prices. They determined that economic elements such as exchange rates or interest rates, which may change asset costs, were not included in the model. There are significant changes in the records applied in the forecasts. Also, the time scale of data collection, especially in Limsombunchai (2004), is relatively short. Ottensmann et al. (2008) used OLS to examine the performance of various metrics, even though data was not split in their study. The experiments revealed that accessibility to many work hubs is a strong indicator of house costs and should be accounted for in the model. Parameters concerning employment, mileage, travel duration, and access conditions to the work center must all be considered. Some international studies, such as (Limsombunchai et al. 2004), employed market prices instead of purchase prices or the mean cost of the flats having identical attributes (Bae.

2003). Boye, Mireku-Gyimal & Luguterah (2018) applied the PCR model to assess unit cost principal components in housing unit prices from 2003 to 2017.

As cited by Manninen (2021), some researchers such as Dubin (1998), Limsombunchai et al. (2004) as well as Ottensmann et al. (2008) applied the semi-log form is also utilized for home price projections to decrease error variance. However, according to Dubin (1968), even if logarithmic forms diminish heteroscedasticity, the results will be skewed when converted directly to functional forms for predicting housing prices. On the other hand, he claims that the log form's theoretical advantage overcomes biases leading to superior forecasts. The outcomes of his study are presented in a linear functional format.

Mostofi, Togan, and Basaga (2021) performed a house price prediction analysis using a deep neural network model principal component. Their research aims to look at the influence of skewness on the linked PCA and DNN as a “PCA-DNN” model's prediction accuracy. They employ 1381 price records with 19 attributes gathered by scraping real estate pricing units from Trabzon. Because some features had a high correlation, some of those were eliminated. A three-step procedure is used to analyze a high dimensioned and favorably skewed property pricing dataset. To begin, the data is normalized using one of three traditional skewness handling techniques. Second, PCA decreases the collinearity of the data. Third, the precision of the PCA-DNN approach is evaluated across data with varying degrees of skewness by examining their error values. The findings indicate that the CRT approach as cube root transformation may significantly enhance the forecasting precision as well as processing time.

Renee (2021) also applied principal component analysis and the regression model to a house price prediction. To estimate the ultimate price of a property, the Ames housing dataset was analyzed, containing 79 explanatory factors that practically describe residential dwellings' features. Renee chooses sales price as the target variable and divides the data into train and test sets to create a predictive model. After applying PCA, she picked the regression model and checked the accuracy of training data based on the test set. The result shows a 64% accuracy, indicating that the model is not overfitting.

Gupta (2010) investigates the ability of PCR and Stochastic regression approaches to anticipate actual home prices in the United States using 112 macroeconomic predictors. Bayesian regressions were applied to anticipate GDP in the United States during the time frame of January 1992 to December 2000. The PCA results reveal that one principal component model is optimal for predicting the property prices in the United Stated, according to Mean Square Forecast Errors. (Gupta 2010.)

Sisman (2016) uses PCA to assess house value in their study. They state that several factors influence both tangible and intangible real estate value. The study aims to provide insight into the dynamics of property market pricing in Samsun. Each sale data set comprised 13 components, and the 149 house prices were studied. Minitab16 software was used to perform the PCA on the data, and three components were found as principal components. Elevator, parking area, insulation, number of bathrooms, and number of rooms all had a high impact in all three categories.

In their study, Jolliffe, and Cadima (2016) state that in a variety of fields, massive data sets are becoming more prevalent. Approaches to analyzing big data sets must drastically restrict their scope in an interpretative manner while maintaining the core of the data. Many approaches have been developed for this goal, but the principal component analysis is one of the earliest and most commonly utilized today. It simply reduces the dimensionality of the dataset while preserving statistical information. (Jolliffe & Cadima, 2016)

In another study by Ismail, Warsame, and Wilhelmsson (2021), the research supplemented many underlying independent control variables, such as demographic characteristics and socioeconomic factors. The variables that assess housing and housing conditions in the municipality are the primary focus of this study. They employed a considerable number of highly correlated factors. They then began the exploratory analysis by looking at the correlations between the included variables. Then, a principal component analysis was conducted. Several new variables were constructed from all of the components involved in the study, but the new variables were not correlated. These additional variables are utilized to estimate direct and indirect effects in regression analysis in the last phase of the exploratory investigation. (Ismail, Warsame and Wilhelmsson, 2021)

Mostofi, Vedat, and Basaga (2021), in their house price prediction model, mention that the in-depth price prediction (DNN) of neural networks is influenced by two fundamental aspects of real estate data which are bulky and not normally distributed. To begin, the data is normalized using one of three traditional skewness handling techniques. Second, PCA decreases the bulk of the data variables. Third, the precision of the PCA-DNN approach is evaluated across data with varying degrees of skewness by examining their error values. The findings indicate that the CRT approach as cube root transformation may significantly enhance the forecasting precision as well as processing time. (Mostofi, Vedat, and Basaga, 2021.)

Li and Zhang (2011) conducted their research on 12 real estate listed businesses in China and calculated different financial indicators using PCA in the SPSS. They attempt to identify current challenges in China's real estate business to better comprehend the development scenario and its prospects and strengthen the real estate enterprise's overall competitive power. They chose PCA since the data was too large and the link between the indexes was somewhat complex. They used the PCA to evaluate the ten primary financial data. Using SPSS, Li and Zhang (2011) created a covariance matrix, a correlation coefficient matrix with eigenvalues and eigenvectors. To describe the complete information of primitive indexes, they used the first four primary components. To convey comprehensive information on primitive indexes t, shareholders' equity, inventory, and the property's entire turnover is a form of more considerable weight.

Oladunni & Sharma (2016), in their study, point out Hedonic Housing Theory as machine learning research. According to the hedonic pricing theory, a property's price is determined by its diversified features. The econometric concept is investigated using machine learning algorithms. As learning algorithms, SVR, K-Nearest Neighbor (K-NN), and PCR were employed in this study. The model's performance revealed that PCR has a modest advantage compared to Support Vector Regression and K-Nearest Neighbor. The study also confirmed the appropriateness and interchangeability of PCR, Support Vector Regression, and K-nearest neighbor in applying hedonic pricing theory (Oladunni & Sharma, 2016.)

3.2 Previous research summaries

The summary of prior investigations is shown in table 3 to provide a basic overview of the reviewed papers. The list is not comprehensive, but there is a noticeable concentration on data variables and techniques employed. This offers solid advice for the empirical portion of this thesis and acts as the thesis' foundation. Table 4 also outlines the applied methods, sample size, and observed findings from previous housing market studies.

Table 3. A compilation of prior research on the factors that can influence property price

Author	Year	Theoretical variables influencing on housing price
Abraham & Hendershott Hort Malpezzi Oikarinen	1996 1998 1999 2007	National influences, interest rsate, inflation, building cost
Abelson	2005	Disposable income, consumer price index, unemployment, real mortgage rates, housing supply...
Nisanci	2005	Density of social, educational,& cultural prospects, the region's living conditions, region's commercial & industrial enterprises
Abelson	2005	Macroeconomic factors (unemployment, interest rate, stock market, inflation, GDP, income ...)
Oikarinen	2004	Changes in pricing in the Helsinki metropolitan region cause price increases in other local & provincial centers all over Finland
Berg Oikarinen	2002 2009	National influences, household debts, price effect of construction all over country
Levin, Montagnoli & Wright	2009	Population shifts, and population ageing (Particularly in Scotland and Weils)
Montagnoli and wright	2009	US market that the demand for housing climbs dramatically until around the age of 30, then becomes essentially flat around the age of 40, and subsequently drops afterwards
Hornstein	2009	Household income, credit arrangements, shifts in owner-occupied housing supply and demand, government-sponsored enterprises and policies
Lönnqvist	2015	Infrastructure, transportation, working age
Wen & Zhang	2013	Education and educational facilities,proximity to high schools and colleges. (elementary and junior high schools had a strong school district influence on housing prices (Hangzhou area. Chinese market)
Augustyniak et.al.	2015	Demand shock by multiplier affects like interest rate change, construction industry policy
Metz	2015	Distance to educational centers & schools, residences within walking distance of an educational facility, shopping areas are extremely common near high schools, homeowners value accessibility to both sites.
Lönnqvist	2015	Dwelling size, quality, structural factors, accessibility to business & developed areas
Wen & Zhang	2013	Educational level, educational institutes, educational facilities
Peng	2017	Fertility rate & the aging dependence rate, as a result, a predictive demographic change is a key indicator for future price changes
Barbu, Vuta, Strachinaru & Cioaca	2017	Flow of immigrants, macroecjonomic indicators, pace of economic growth, by 1% increase in immigration, housing prices rise by 0.045 %
Feng, Kim & Lee	2018	Unemployement rate, GDP growth, and mortgage rate and population. Housing prices rise at a slower pace when population grows compared to the rate at which prices fall when population falls.(US housing markets from 2010 to 2017 during and after recession crisis)
Puturki	2018	Demographic demand & changes, types of housing, age structure, migration, working age, Consumer confidence, construction confidence
Larkin. et. al.	2018	Migration affects housing price and modifies regional demand, house price rises are more controlled in nations that are less welcoming to immigration
Green	2018	Immigration, financial crisis, macroeconomic policies (Brexit)
Jin and Tang	2018	Average annual income of the inhabitants, competitiveness of cities, city GDP, and the urban liveability index (The city's GDP is the most essential element, and interest rates are inversely connected with house prices)
Feng	2019	Educational level, educational institutes, educational facilities, (first-tier universities have a substantially higher pricing influence on the (Beijing's) housing market, but second- and third-tier universities do not)
Lindblad	2019	Monetary policy, mortgage rate, cost of loan, tax rate ...
Keskinen	2020	Population,urbanization, university cities, developed areas
Khadem	2021	Migration, access to businesses and educational centers quality (Residence near a high-scoring public-school costs 2.4 times as much as housing near a low-scoring public school)
Aalto & Sinkkonen	2021	Populated areas, living space, mean income, access to business services, popularity of outskirts areas after covid pandemic (change in customer preferences)

Table 4. The applied techniques, sample size, and findings from the study background

Researcher	Year	Predicting method	Observations	Type of housing price data	Time frame for data collection	Findings
Dubin	1998	OLS and ML regression	1493	Sale price	1978	ML regression predicted housing prices better than OLS regression, the theoretical superiority of the log form will overcome the bias to give superior forecasts
Limsombunchai, Gan & Lee	2004	WLS regression and ANN-model	200	Property market price	2003 (May)	ANN-model had a higher predictive capacity on housing prices than the WLS approach
Ottensmann, Pauton and Man	2008	OLS regression	8772	Sale price	1999	Accessibility to work hubs is a strong indicator of housing price and should be accounted for in the model. In location parameters concerning employment, either distances, travel times, or metrics of accessibility to the employment center should be considered
Li & Zhang	2011	PCA	1620	Enterprise's operational ability	2010	The first four major component accumulation contribution rate reached 86.19 percent. shareholders' equity turnover, inventory turnover, and total asset turnover occupy a higher load, demonstrating that the initial principal component indicates the enterprise's operational competence. The three indices of the factor loading amount are bigger in the third main component, net profit growth, net assets yield, and profits per share, implying profitability.
Oladunni & Sharma	2016	OLS, SVR, K-NN & PCR regression	...	Housing price	...	PCR has a modest advantage over SVR and K-NN models
Bork, Vinther & Mller	2018	PCA, PLS, SPLS	128 timeseries	Predictivity in future housing values	...	PLS models consistently outperform PCA models, PLS models produce strong out-of-sample predictive power in addition to the price-rent ratio, autoregressive benchmarks, and regression models based on limited datasets.
Boye, Mireku-Gyimal & Luguterah	2018	PCRM	420	Housing unit price	2003-2017	The principal components with small eigenvalues are eliminated to remove multicollinearity from the data. The PCs with higher eigenvalues explain as much variation as feasible in the original data set
Mostofi, Togan and Basaga	2021	PCA -DNN	1381	Housing unit price	...	The CRT approach significantly enhance the prediction accuracy & computational time of the PCA-DNN model.
Renee	2021	PCA & Regression	...	Price of property	2006-2010	The result shows a 64% accuracy, indicating that the model is not overfit.
Gupta	2010	PCA & Regression	...	Home price in US	1992-2012	According to the Mean Square Forecast Errors (MSFEs), the results show that a principal component regression with only one factor is best suitable for projecting real US house values.
Sisman	2016	PCA	1937	Price of property	...	They state that there are several factors that influence real estate value, both tangible and intangible. Elevator, parking area, insulation, number of bathrooms, and number of rooms all had high impact in all three categories.

As shown in table 4, as an overview of some past research, the methodologies, sample sizes, and periods vary. Applied methods differ, but almost all selected the price as their dependent or target variable. There are various methodologies, but the principal component analysis is the most commonly used in the research with a high number of predictors, such as real estate market ones. Some of the outcomes point out the model's accuracy, and others show the contribution level of variables, some of which are macroeconomic ones. The studies focused on property development aspects contain a pool of enormous parameters. The variable dimensionality challenge was apparent in some investigations. Data can be collected by choosing widely used variables and data availability. In addition, the findings of previous studies serve as a basis for the projected outcomes in the empirical section of this thesis.

3.3 Utilized guideline

According to the study background and as cited by Kokkinen (2020, pp 16-17), researchers such as Abraham and Hendershott (1996), as well as Hort (1998), Malpezzi 1999, Oikarinen (2009), and as table 3 shows, most previous studies imply the macroeconomy factors such as national and governmental influences on housing price (Abelson 2005; Hornstein 2009; Augustyniak et al. 2015). Monetary policies, inflation rate, GDP, interest rate, and unemployment rate are determinants of housing price (e.g., Feng, Kim & Lee 2018; Green 2018; Lindblad 2019). While some other researchers point out to increasing population, age structure, household size and property features (e.g., Nisanci 2005; Levin, Montagnoli & Wright 2009; Lönnqvist 2015; Peng 2017; Feng, Kim & Lee 2018; Puturki 2018; Keskinen 2020), immigration rate (Barbu, Vuta, Strachinaru & Cioaca 2017; Puturki 2018; Larkin et.al. 2018; Green 2018; Khadem 2021), educational level and facilities (Nisanci 2005; Wen & Zhang 2013; Metz 2015; Feng 2019; Keskinen 2020; Khadem 2021), and enterprises turnover & individual income level (Abelson 2005; Nisanci 2005; Hornstein 2009; Jin & Tang 2018; Feng, Kim & Lee 2018; Aalto & Sinkkonen 2021). As a result, the determinant factors considered in the empirical section are the population growth rate, age structure, educational level, immigration rate, household size and type, employment rate, income level, and sales price.

There have been a variety of methodologies for forecasting house prices that have been focused on different aspects, as shown in table 4. OLS was used by several researchers (e.g., Dubin, 1998; Ottensmann and Pauton, 2008). Comparing OLS and ML, Dubin's (1998) study findings show that housing prices were predicted more accurately by ML regression than by OLS regression. The ANN model has a more robust prediction capacity on housing values than the WLS approach, according to Limsombunchai, Gan, and Lee (2004). Oladunni & Sharma (2016), in their study, point out Hedonic Housing Theory, based on which a house is a unique property with several attributes that influence its pricing. Applying

(SVR), (K-NN), and (PCR), the learning algorithms were compared. The model's performance revealed that in comparison to SVR and K-NN, PCR has a modest advantage. The study also confirmed that PCR, SVR, and K-NN are all acceptable and interchangeable when applying

hedonic pricing theory. (Oladunni & Sharma, 2016.)

Principal component analysis and principal component regression analysis are other popular methods for predicting property prices and determining which influential variables contribute the most to fluctuations in housing prices, which has become an even more common approach in recent years (Li & Zhang, 2011; Gupta, 2010; Sisman, 2016; Boye, Mireku-Gyimal & Luguterah, 2018; Mostofi, Togan and Basaga, 2021; Renee, 2021). When dealing with multicollinear data, PCA and PCR significantly manage the dimensionality problem while assessing many variables and conserving a large amount of data throughout the analysis.

According to the research background and the study goal of developing a predictive housing price model, PCR is considered the most suitable analysis method. It is aligned with the thesis goal to develop a property price predictive model to assess the significance level of SDPC principal components in the Finnish real estate market. PCR solves multicollinearity issues while preserving as many predictive variables as feasible in the original data set without missing much information. Thus, according to the background of the study and inspired by Boye, Mireku-Gyimal & Luguterah's (2018) PCRM model, PCR is selected for this thesis and discussed in the next section.

3.4 Principal Component Regression

Considering the thesis focuses on the selected socioeconomic and demographic components (SDPC) in Per Square Meter Pricing (PSMP) of the housing market in ten main cities in Finland from 2013 to 2020, to implement the study, the Principal Components Regression Analysis is applied. The PCR approach solves multicollinearity issues while preserving as many predictive variables as feasible in the original data set without missing much information. This goal is accomplished by combining the information in the original data set into a new set that is the Principal Components, with a fewer number of variables (PCs). Inspired by Boye, Mireku-Gyimal & Luguterah's (2018) PCRM model assesses unit cost principal components in housing unit price. This study considers X the log-transformed data matrix containing the SDPC data matrix of dimension (n.p) and Y log-transformed observed PSMPs of size (n x 1). The low eigenvalues represent high multicollinearity (Hintze 2007). The principal components with small eigenvalues are eliminated to remove multicollinearity from the data. The PCs with higher eigenvalues explain as much variation as feasible in the original data set (Boye, & Mireku-Gyimal & Luguterah 2018).

The initial step to start the analysis is performing PCA to minimize the dimension of the predictors. The idea is to have the main components that contain most of the dataset's helpful information in new predictors regarding the variance. The use of PCA also contributes to the management of collinearity when there is a high correlation among several predictors. The data is therefore grouped to the corresponding predictors by applying PCA. (Goonewardana 2019.) The approach is made by selecting features that are not correlated to have a new matrix with fewer predictors. This helps implement the analysis in an easier way to describe and visualize. The output first displays the coefficient as each predictor's contribution to the new transformed variable component. It also shows the so-called score, which each observation coordinates in a new feature. The output also indicates how much information is captured by each component, explained as a percentage of variance. (Li & Zhang 2011, P 5.)

The MSE is calculated to determine the predictability's accuracy. MSE measures the average squared difference between predicted and actual values. (Binieli 2018.) RMES, the root mean squared error squared units of predictions, is an additional measure for model performance (Brownlee 2020). R^2 is the other standard metric anticipated by the proportion of variance of the dependent variable. The other factor to consider is preventing model overfitting and underfitting. Overfitting is the tendency of model for complicated models, making the interpretation and generalization of model challenging. To trade-off between the complexity of the overfitting and underfitting, the data is divided into train and test data. Thus, firstly the dataset is randomly partitioned into train and test sets, accounting for 80% of the training data and 20% of the test data. The process is referred to as non-stratified partitioning. The training data is further divided into training and validation data.

Principal component regression is a commonly used and popular technique. The following are some of the advantages of PCR:

- PCR can do regression when explanatory variables are highly correlated or collinear.
- PCR is simple to understand: it substitutes the X1, X2,..., Xp basis with an alternating basis of principal components, removes components that don't explain the substantial variance, and then regresses the target variable over selected components.
- The PCR is implemented automatically and needs to be done to decide the number of components to preserve.
- Principal components that have been dropped demonstrate which linear combinations of variables are causing collinearity.
- The number of components maintained in PCR is a discrete parameter.
- In case there a broad data, when there are more variables than observations, PCR can be used (Wicklin 2017.)

When determining which principal components to eliminate, the amount of the variation of the components is used as selection criteria. (Wicklin 2017.)

Accordingly, for the thesis, the Principal Component Regression (PCR) model depicts the relationship between the socioeconomic and demographic parameters of cities known as X and the target variable, the property price per square meter of blocks of flats represented by Y in associated towns. The predictive model estimates the housing price for the above-specified cities following the application of the algorithm. The initial sample for this study was selected and aggregated from six yearly data sets from Statistics Finland, including the main numbers, based on the influential variables proposed by the study backdrop. The data sets contained key figures of socioeconomic and demographic factors such as population by region, education level of people aged 15 and over by the city, household income type and structure by city, dwelling price by city, volume and value of dwelling sales by city, and establishment and turnover of businesses by city. As a result, only information regarding the target markets was chosen. Followingly, the thesis data includes 40 observations related to the target cities' information from 2013 to 2020 regarding 46 features. The target variable in the thesis data is the per square meter price in Helsinki, Espoo, Vantaa, Tampere, and Turku, which is numerical and indicated in 1000 euros. The following section includes the Empirical research.

4. Empirical research

Following stating the research problem, the first step in implementing the empirical part was to collect and understand the data, the target and exploratory variables, the outliers, and missing values using MATLAB software. The z-score method was used to normalize the thesis data. Once the data was prepared, correlation analysis was conducted. Considering high dimensionality, the principal component analysis was performed at the next step to detect the optimal number of components considering how much data they capture. After PCA, the next step was data modeling. The final step was evaluating the model accuracy and deploying the results. Figure 12 shows the process of empirical research.

Figure 12. Empirical research process

The findings of the analytics models will be described in the following section. First, the research design is explained according to the background of the study. Then data collection and methodology are described. Next, some descriptive statistics and the correlation between the variables are assessed. Finally, the results of analysis methods and observations are discussed and interpreted in detail. The empirical research is wrapped up by pointing out the limitations of implemented tests and observed outcomes.

4.1. Research method

The following part describes the thesis data structure. Following the background of the study, there hasn't been a lot of global academic research on the evolution of local housing markets. Kokkinen (2020) indicates how the growth of property prices in one local region reflected the development of property prices in other regional areas in Finland. Price changes in the Helsinki metropolitan region, as Finland's economic core, induce price hikes in other local and provincial centers (Oikarinen 2004). Among the few studies on Finland's housing market, most have concentrated on either the predictive pricing model taking macroeconomic factors into account, mainly on investment and risk mitigation regarding the Finnish housing market. The study background also indicates that the most common focused market has been Helsinki metropolitan area, yet there are a few studies considering the whole country. The other critical issue is Finland's housing market is not significant, and there are only limited sales transactions in some places. Thus, having a reliable picture of market price trends and fluctuations is prone to statistical flukes (Aalto & Sinkkonen 2021).

Considering the above points, there are gaps in the research background concerning demographic and, more precisely, socioeconomic and demographic aspects influencing home prices in Finland's property market. To cover the gap in the foundation of study and to diverge from previous studies:

- The research aims to analyze the Finnish real estate market's socioeconomic and demographic principal components in terms of price variations and to develop a predictive pricing model.
- To maintain reliability, the study's scope has been confined to 2013-2020 to focus on current data sets and maintain balance by not considering too old data or dealing with very recent data that contains many missing values.
- Considering that Finland's real estate market and transaction volume are limited, the study is also confined to Finland's five most populated cities, including Helsinki, Espoo, Vantaa, Tampere, and Turku, to avoid the conclusion being influenced by statistical flukes.

4.2 Data collection and study design

The thesis data is compiled and integrated from six yearly data groups from Statistics Finland, one of Finland's most reliable and official statistics sources. Following the insights from the study's background, the variables were chosen.

Supply and demand are the main determinants of the property price. Several things affect supply and demand. Many of these elements are susceptible to policy changes. Some factors, such as interest rates or credit availability, impact demand and supply. (Lindblad et al.2019.) According to Abraham & Hendershott (1996), as well as Hort (1998), Malpezzi (1999), and Oikarinen (2004), the study background and cited by Kokkinen (2020, pp 16-17), and as shown in table 3, as well as some previous studies by Abelson (2005) cited by Manninen (2021, pp 2-5), Hornstein (2009) and Augustyniak et al. (2015), macroeconomic factors such as national and governmental influences are determinants in housing price. Some other studies also mention monetary policies, inflation rate, GDP, interest rate, and unemployment rate are key drivers in housing prices (e.g., Abelson 2005; Feng, Kim & Lee 2018; Green 2018; Lindblad 2019).

According to Lönnqvist (2015), age structure affects property market prices significantly. Puturki (2018) and Keskinen (2020) also argue that one of the driving forces in the property market is the age of the population, particularly the working-age structure of that area. On the other hand, some studies emphasized demographic features, including the increase in population, age structure, family size, and property features (e.g., Nisanci 2005; Levin, Montagnoli & Wright 2009; Lönnqvist 2015; Peng 2017; Feng, Kim & Lee 2018; Puturki 2018; Keskinen 2020) significantly can impact on real estate market price. Other studies also point out that immigration rate (Barbu, Vuta, Strachinaru & Cioaca 2017; Puturki 2018; Larkin et al. 2018; Green 2018; Khadem 2021), educational level, and facilities (Nisanci 2005; Wen & Zhang 2013; Metz 2015; Feng 2019; Keskinen 2020; Khadem 2021), and enterprises turnover & individual income level (Abelson 2005; Nisanci 2005; Hornstein 2009; Jin & Tang 2018; Feng, Kim & Lee 2018; Aalto & Sinkkonen 2021) can significantly impact the price of the property market. Lindblad (2019) and Aalto & Sinkkonen (2021) also point out that mean income, employment status, customer confidence level, and access to businesses and developed areas are determinants of property market price. Wen & Zhang (2013) and Feng (2019) discuss their studies regarding the property market price in China, the role of the educational level of residents, educational facilities, and the number of educational institutions and their potential influence on the price rate of the housing market.

Puturki (2018), Keskinen (2020), and Aalto & Sinkkonen (2021) also discuss the population diversity and structure, as well as the number of immigrants, may shape and adjust housing market demand and, as a result, the housing prices. Therefore, the other predictors in this thesis data include the overall number of Finnish, Swedish, and Sami speakers vs. the number of foreign speakers who indicate the number of immigrants in each city. Other driving determinants in property market pricing discussed by Aalto & Sinkkonen (2021) include construction confidence, supply, home size, dwelling structure, and quality criteria.

As a result, the population growth rate, age structure, educational level, immigration rate, household size and type, employment rate, income level, and sales price are the determining variables addressed in the empirical section (table 3). Consequently, this research examines five of Finland's most populous cities during 2013-2020 by the above variables, which were selected based on the study background and aligned with the goal of this research. The study examines the socioeconomic and demographic principal components (SDPC) in Finland's real estate market and constructs a predictive pricing model based on that data. All data variables, calculation procedures, units, and other necessary information can be found in table 5 of the Appendix.

The initial sample for this research was collected and combined from six yearly data sets from Statistics Finland, one of Finland's most reliable and official statistics sources (Table 9). The variables were chosen based on the insights gained from the study's background, including key figures of the target cities regarding population, age, the ratio of immigration, educational level, employment rate, income level, household size, type, etc.

The thesis's data includes 46 variables in 40 observations. Yet the data suffers multicollinearity as some of the variables are strongly correlated (Table 4 appendix).

Table 6. List of selected data groups for this study

1. Key figures on population by region (%) :
· Increase of population
. Share of persons by age group (under 15-15-64, above 65)
. Demographic & economic dependency ratio
. Share of Finnish, Swedish, Sami, & foreign- language speakers
.Average age of both sexes

2. Education 15 years old or over by city:
. The number of population with upper secondary degree
. The number of population with Bachelor's or equivalent degree
. The number of population with Master's or equivalent degree
. The number of population with doctoral or equivalent degree

3. Income and income structure of households (mean)
. Total earned income
. Total entrepreneurial income
. Total entrepreneurial income of business activities
. Total property income
.Total divident income & income from sales profit

4. Dwellings & price by city
. Total # of terraced houses, flats & 1,2, 3,+4-person households
. Old dwellings in housing companies
. Old terraced houses in housing companies
. Old blocks of flats in housing companies
. New terraced houses in housing companies

5. Number and value of dwellings sales
. Number of transactions of new terraced houses in housing companies
. Number of transactions of new blocks of flats in housing companies
. Number of sales of total building types
. Average area of sold properties
. Total sales value

6. Establishment of enterprises by city
. The number of established enterprises
. The number of personnel in enterprises
. The turnover of enterprises (1000€)

The data contained missing values specifically for the year 2020. The price of detached and terraced houses was excluded from the study due to many missing values, which might negatively influence the study's reliability. However, the data had missing values for variables such as occupants' education level, capital income, sales profit, and unemployment security. Since the target variable, PSMP, missing values were managed using MATLAB software functions like linear interpolation and nearest value. Outliers also were dropped above the upper quartile (75%) and below the lower quartile (25 %).

4.3 Descriptive statistics

It's critical to understand the data structure before implementing machine learning algorithms. The data contains 40 rows and 46 variables. The target variables of the thesis data are the average PSMP of blocks of apartments in Helsinki, Espoo, Vantaa, Tampere, and Turku, in the time frame 2013-2020 (figure 13, appendix), which is numerical and indicated in 1000 euros. The initial step in implementing the analysis was understanding the data. Table 7 shows the data summary statistics in the appendix.

4.3.1. Data visualization

To understand the insights from the numbers, this section is allocated to visualizing the demographic situation in target cities, including the share of foreign-language speakers, average age of inhabitants by sex, increase of population, working-age population proportions, educational level of inhabitants, total sales transactions of whole buildings, entire established enterprises, average per square meter price of blocks of flats, the annual change in the price of each city is explored. In the next part, the results of correlation analysis are also discussed to see which variables in data are positively or negatively correlated. The population growth rate of Finland's top 5 inhabited cities fluctuated from 2013 to 2020. Helsinki has the highest proportion, followed by Espoo and Vantaa. In 2020, their population growth slowed substantially

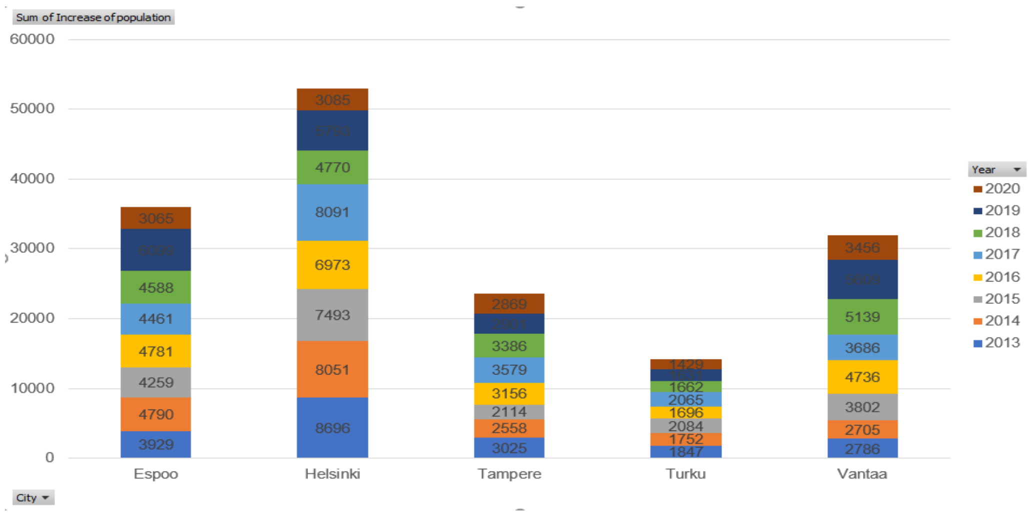

Figure 14. Increase of population in five most populated cities of Finland during 2013-2020

The share of foreign-language speakers has interestingly been growing in all ten cities from 2013 to 2020, as seen in figure 14. Yet this growth trend has accelerated in Helsinki, Espoo, and Vantaa. Helsinki has the highest share of foreign-language speakers compared to other target cities, which rose from nearly 85,000 in 2013 to roughly 110,000 in 2020, making up almost 17% of its population.

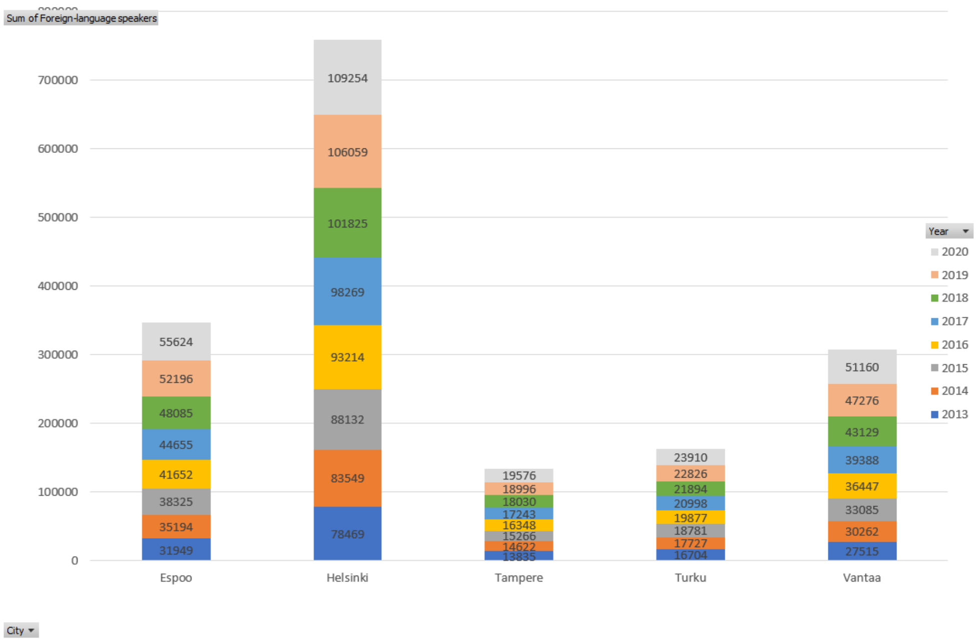

Figure 15. Foreign-language speakers in target cities

Considering the age of the population from 2013 to 2020, Espoo, with an average age of 37,7 in 2013 and 38,8 in 2020, has the youngest age structure among other cities. The working population age group, on the other hand, is dropping in all cities except Vantaa, as can be seen in Figure 16. This decline is particularly apparent in Helsinki, where the working-age population has decreased from over 208,000 in 2015 to just over 204,000 in 2020, even though the proportion of people over 65 is rising in all mentioned cities.

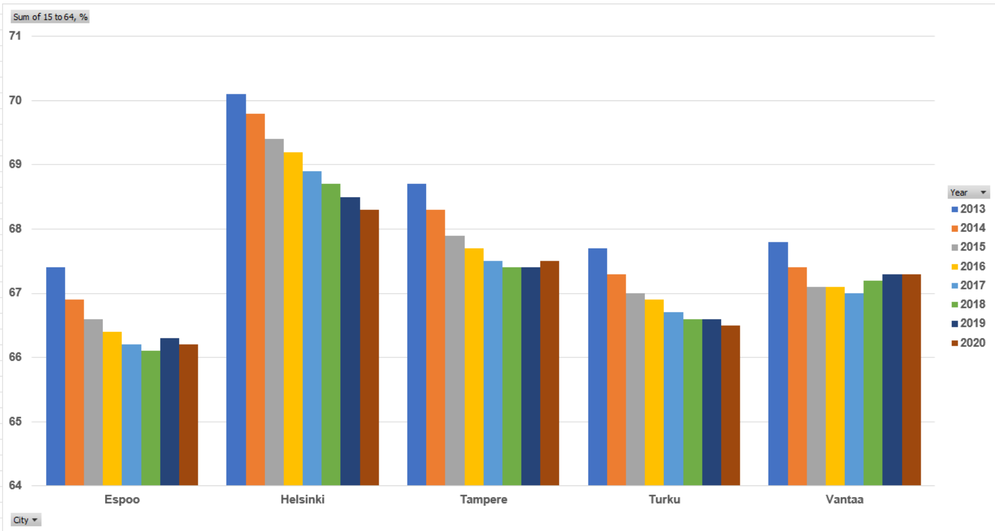

Figure 16. Working population age group in target cities (2015-2020)

Comparing the educational levels of its residents by the end of 2020, Helsinki stands out among comparable cities with the most educated population, with 558,690 people holding bachelor's degrees, 690,808 master's degrees, and 84,492 Ph.D. Espoo, Tampere, and Vantaa, respectively, are ranked next regarding the higher educated population.

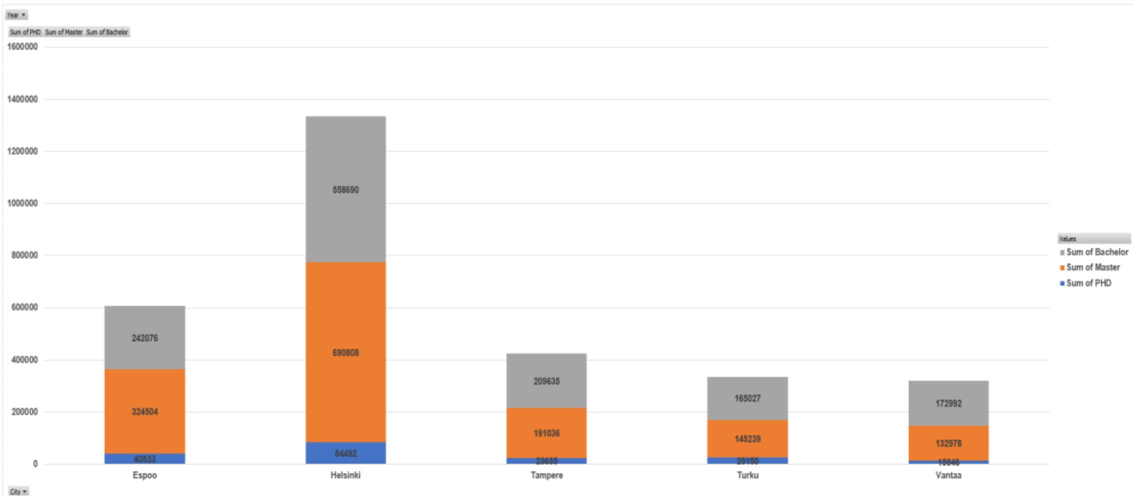

Figure 17. Sum of educational level by the city

There are 242,076 bachelor's degree holders, 324,504 master's degree holders, and 40,533 Ph.D. holders in the Espoo by 2020.

Helsinki has been the busiest market during the last years in terms of quantity of sales. From 2013 to 2017, there was an upward trend, reaching 38859 transactions. In 2018, the sales volume declined to 38,202 transactions; however, by 2020, it rebounded to 46,398 total trades. After Helsinki, the hottest real estate sales transactions markets are Tampere, Espoo, Turku, and Vantaa. Figure 18 shows the total number of founded businesses in each city during 2015-2020 and their total earned income. Despite being ranked second in terms of the total number of established companies, Espoo, with over 860 million euros, has the highest rated income proportion of all cities. Vantaa, on the other hand, with nearly 670 million euros, is the second-highest earned enterprise income city, slightly ahead of Helsinki, even though Vantaa is placed fourth in terms of established enterprises.

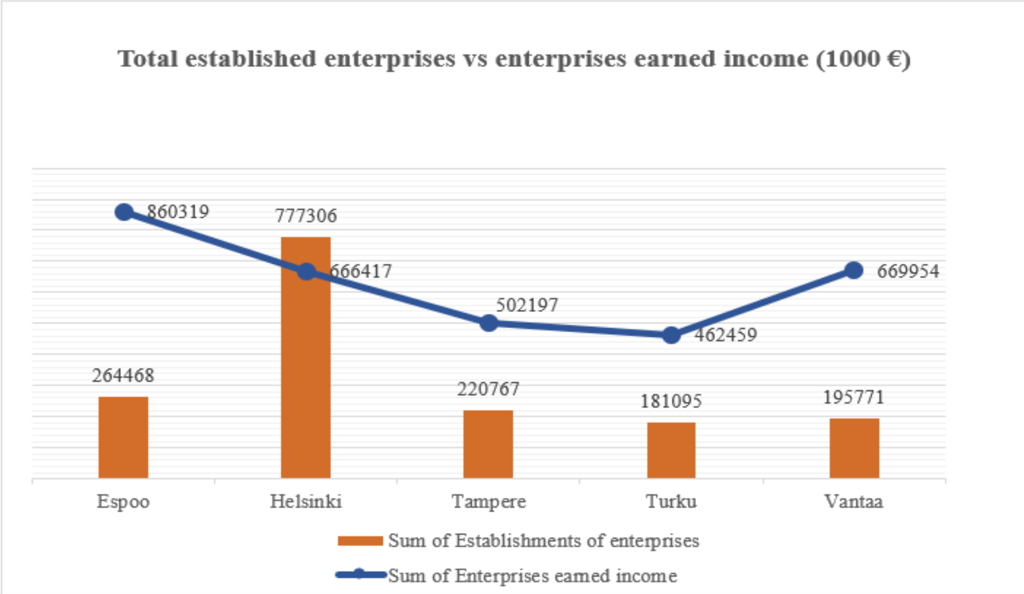

Figure 18. Total established enterprises and their earned income by city, 2015-2020

In 2013, Espoo's average gross income (78,082€) was higher than comparable cities. Espoo retained its position as the city with the highest average property income until 2020.

Figure 19. Average gross income in target cities

In 2013, the average gross income in Helsinki was 59,824€ and lagged behind Espoo. From 2016 to 2020, the average gross income gap between Espoo and Helsinki narrowed with an accelerated growth trend in Helsinki, and 2020 overpassed Espoo with an average gross income of 46,176€. As figure 19 depicts, Espoo, Helsinki, and Vantaa stand out regarding the average gross income of comparable cities.

Concerning the target variable, the average PSM price of old apartments had a steady upward trend in Espoo, Helsinki, Tampere, and Turku. Out of a PSMP of 3,897€ in 2013 to 4,845€ in 2020, Helsinki has seen the highest growth. The average PSMP in Espoo rose from 3,080€ in 2013 to 4,434€ in 2020. Tampere saw a similar upward trend, rising from 2,248€ in 2013 to 3,455€ in 2020.

Figure 20. Average blocks of flats per square meter price in target cities

Helsinki, as anticipated, experienced the most significant price increase from 2013 to 2020, with an almost 8% increase. Helsinki and Espoo have relatively similar pricing growth. Vantaa saw a considerable price surge in 2018, followed by a decline in 2019 before rebounding in 2020.

4.3.2. Correlation analysis and insights

Correlation analysis (table 8, appendix) shows a strong correlation between some variables. For instance, an increase in population is negatively correlated with the economic dependency ratio. It is also positively correlated with the number of people with bachelor’s and master’s degrees, enterprises' establishment, turnover, and PSM price of dwellings.

The proportion of the working-age group (15-64) is strongly negatively correlated with the demographic dependency ratio. It is also highly positively correlated with upper secondary and bachelor’s degree holders, Finnish and Sami speakers, sales of all buildings, and enterprises' establishment. Over 65 population is negatively correlated with foreign-language speakers and enterprise earned income, average gross income, municipal tax, and PSM price of old terraced houses. The economic dependency ratio is negatively correlated with enterprise turnover and earned income, average gross income, disposal cash income, and municipal tax. The average age is negatively associated with enterprise earned income, average gross income, income tax, and municipal tax. Educational levels are highly positively correlated with other academic levels, Swedish, Sami, and foreign speakers, and sales of entire buildings, establishment, and turnover of established enterprises.

PSM blocks of flats are highly positively correlated with a Ph.D. education degree, Finnish, Swedish, Sami, and foreign-language speakers, the total household dwelling units, total number of terraced houses and flats, 2,3 and over 4-persons household dwelling units, employed labor force, the total number of sales of all building types and the value of sales, as well as PSM price of old dwellings. PSM blocks of flats are also moderately positively correlated with the increase of population and proportion of the population that are 15-64 years old, upper secondary, Bachelor’s, and Master’s degree holders. It is also linked with 1-person household dwelling units, the number of the establishment of enterprises, enterprises’ turnover, property income, dividend income, the average capital income tax, and the PSM price of old terraced houses. The target variable is also moderately negatively correlated with the demographic dependency ratio. Therefore, the abovementioned predictors, especially those with a high correlation with PSMP, are expected to contribute to the principal component’s regression model.

Since correlation does not necessarily imply causation (Lee, 2021), the further analysis explores the significant relation between explanatory predictors and the target variable. The PCA and PCR analysis and results are presented in the following sections. First, the principal component analysis is performed, the principal components are observed, and the regression model is conducted.

4.4. PCR

When the data contains many features, and the goal is to reduce the number of features or predictors, principal component regression is particularly effective. When the variables are highly correlated, PCR comes into the picture.

There are two steps to follow. The first is to conduct PCR to minimize the feature's dimensions. As a result, we may run a regression model after determining the principal components. The data set includes 46 variables related to socioeconomic and demographic statistics on Finland's top five most populated cities in 2013-2020, with 40 rows. The data is cleaned and processed using Matlab software before performing the PCA. The only categorical variable in the data, “City,” was converted to numeric format. The city and year were taken out of the analysis to concentrate only on socioeconomic and demographic variables. The pairwise correlation is conducted followingly.

Before implementing the PCA, the target variable, which is the average per square meter price, was separated from other exploratory variables. The principal component analysis contains three main outputs:

- The first output of PCA on Matlab is the coeff, which is the contribution of each original predictor to the principal components. The coefficients of the first principal component, p1, are found in the first column of coef, while the coefficients of the second principal component, p2, are located in the second column, and so on... till the very last column.
- • The second output, the score, contains the original data coordinates in the new coordinate system established by the principal components. The original data is contained in the score matrix, which has the same size as the input data matrix observations represented in the corresponding point of the principal components p1, p2, ... pn.
- The Latent vector containing the variance explained by each principal component is the third output. Latent indicates the amount of information taken by the components.
- The final output is a vector holding the percentage variance explained by the corresponding principal component. Both latent and explained convey the amount of

captured information from the original data, present it in the variance format, and explain it in the percentage format.

The target or dependent variable is segregated from independent/predictor variables to implement data modeling and run the PCR. Because the response or target variable is the price of old blocks of flats, it is separated from the other variables. Since the PCA is performed on a subset of the training data predictors, this separation is essential. Then, modeling was performed between these components and the dependent/target variable. Once the target variable is separated from the other variables, the new matrix of predictors, called X, has 43 predictors. The data was then converted from table to array format in the next phase. A data array is required to employ the prediction matrix in principal component analysis. In the first step, all data is divided into training and test data to avoid overfitting and later implement the model. Following random assignment, 20% of the data is retained for test data and the remainder for training data. In the second step, the training data is subdivided into training2 and validation sets for applying the modeling.

Thus, only the training data is used to create the model; the validation data is used to assess the trained models. The model with the best number of components is then chosen using K-fold cross-validation. Once the best model (the number of principal components) is selected, the test data comes in to assess the final model. In other words, the final step is to apply the select model to test data to see how the predictive model works with new data. Figure 21 shows the modeling for PCR.

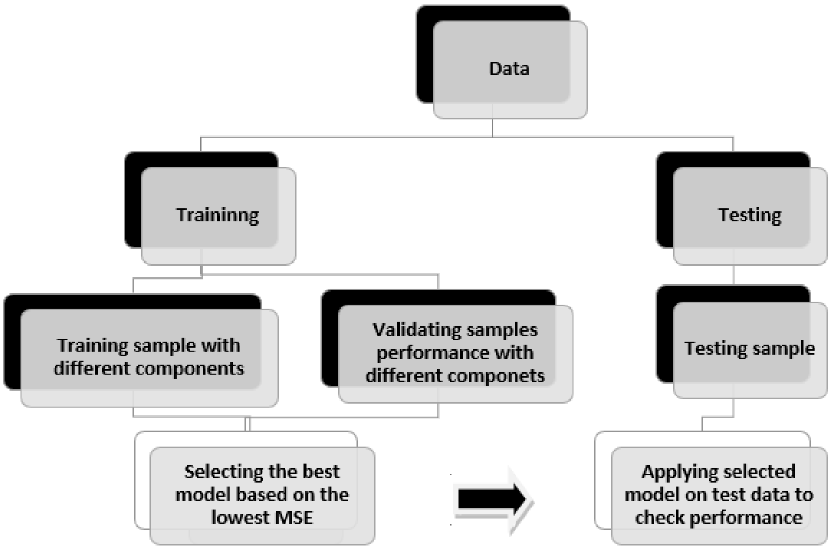

Figure 21. Data modeling & validation procedure for PCR

We can observe the explained variance in figure 22 after running the PCA. The first seven principal components explain more than 95% of the variance. Figure 22 indicates that the first principal component explains almost 48% of the variance, the second roughly 23%, the third about 10%, etc.

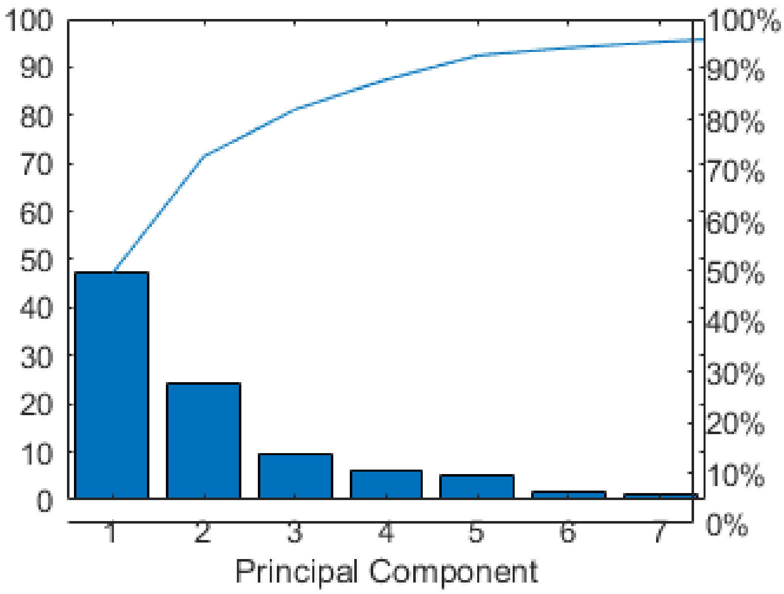

Figure 22. The proportion of variance explained by principal components

The heatmap (figure 23) lets us see how each predictor contributes to the principal components. The result depicts a fair contribution of many variables in the first and second principal components, and no variable is dominant compared to others.

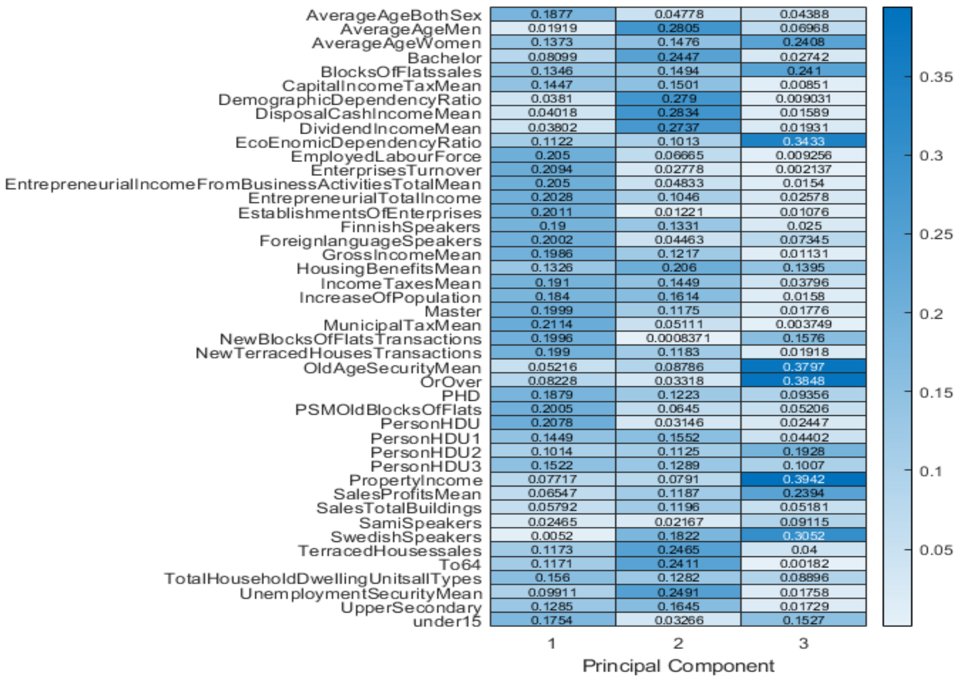

Figure 23. The contribution of predictors in each principal component

The K-fold cross-validation result (figure 24, appendix) shows that the optimal model includes one principal component. Figure 25 in the appendix illustrates the model. The model contains one principal component with an R-square of 0.74 and a P-value of (1.9921e-08), nearly zero, which significantly impacts the price, as the target variable. Running the test data to assess the model accuracy shows an estimated mean squared prediction error of 2.6119e+05.

5. Results and Discussion

Inspired by the high growth in Finland’s real estate market following the covid pandemic, this study aims to examine the socio-economic and demographic principal components (SDPC) and develop a predictive model based on those. The study's scope was confined to Finland's top five populated cities. The research was created to address the information demands of both domestic and international stakeholders interested in the property market of Finland. The following section includes the results of previous investigations and a comparison with the findings of this thesis.

5.1 The influential SDPC on housing prices in Finland

There are various theories regarding the impact of socioeconomic and demographic features on property prices. While there are limited studies on the possible socioeconomic and demographic influences on the price of real estate in Finland, numerous studies have focused on various housing markets in China, Australia, and the United States.

5.1.1 Monetary policies, GDP, inflation, interest, and the unemployment rate

Many prior studies either concentrated on Finland or other countries such as the US, UK, and China real estate markets. As cited by Kokkinen 2020, for instance, Abraham & Hendershott (1996), Hort (1998), Malpezzi (1999), and Oikarinen (2004), as well as some previous studies by Abelson (2005), Hornstein (2009), and Augustyniak (2015), suggest macroeconomic elements such as monetary policies, inflation rate, GDP, interest rate, and unemployment rate are key drivers in property price.

5.1.2. Population growth, age structure, immigration ratio, and household size

Other researchers point to parameters like population increase, age structure, household size, and property attributes. (e.g., Nisanci (2005); Levin, Montagnoli & Wright (2009); Lönnqvist (2015); Peng (2017); Feng, Kim & Lee (2018); Puturki (2018); Keskinen (2020)). The other variable which has been taken into account specifically in research regarding China and the UK property market is the immigration rate, which can modify the demand balance of the area and eventually impact the housing market price both for sale and rental markets (Barbu, Vuta, Strachinaru & Cioaca (2017); Puturki (2018); Larkin et al. (2018); Green (2018); Khadem (2021). The results of the correlation analysis on the thesis data indicated that the target variable, PSMP of blocks of flats, is moderately correlated with the population growth rate and, specifically, the working-age ratio. It is also highly positively correlated with the total household dwelling units, the total number of terraced houses and flats, 2,3 and over 4-persons household dwelling units, employed labor force, the total number of sales of all building types, and the value of sales, as well as PSM price of old dwellings, which is aligned with the study's background.

5.1.3. Educational level and facilities, Income level of individuals and businesses

According to Nisanci (2005); Wen & Zhang (2013); Metz (2015); Feng (2019); Keskinen (2020); Khadem (2021), whose studies mainly were conducted regarding the property markets of China and the United States and Turkey, the educational level of inhabitants as well as access to educational facilities such as schools and colleges significantly can impact the property price in the district. This is especially noticeable if the school or university is well-known and considered first-tier. Some studies have highlighted the proximity to academic institutions and retail malls or commercial hubs, which can substantially influence the area's property price. For example, if shopping and commercial hubs are close to schools and universities, the price increase in the neighborhood will be more visible. Aside from the importance of educational level and academic facilities, some studies such as Abelson

(2005); Nisanci (2005); Hornstein (2009); Jin & Tang (2018); Feng, Kim & Lee (2018); Aalto & Sinkkonen (2021) place more emphasis on the income level of city dwellers, both individuals and businesses, considering their saving and eligibility to receive mortgage and make the property market more competitive and as a result impact on the city's property market price. Running a correlation analysis on the thesis data revealed that PSMP of blocks of flats, as the target variable, is substantially positively connected with the population with Bachelor’s and Master’s education Ph.D. degrees. Education degree, which is consistent with the study's background. It also suggested that the PSMP is substantially linked with the ratio of Finnish, Swedish, Sami, and foreign-language speakers. The correlation analysis results also imply that the PSMP of blocks of flats positively correlates with enterprise turnover, property revenue, dividend income, average capital income tax, and the PSM price of ancient, terraced dwellings. The demographic dependence ratio is also adversely linked with the target variable. However, since correlation does not always indicate causation, extensive research is conducted to determine the impact of various factors on property values in target markets.

5.2 PCR results

The principal component regression PCR analysis assessed and answered the research question. The results indicate that the optimal model contains one principal component with an R-square of 0.74 and a P-value of nearly zero, indicating it significantly impacts the price as the target variable. The chosen model was then tested against the test data to see how accurate it was, which showed an estimated mean squared error of 2.6119e+05. As a result, the thesis outcome shows sufficient evidence to show the possibility of creating a predictive property price model containing socioeconomic and demographic principal components.

5.3. Summarizing the findings

Diverged from earlier studies that mainly focused on macroeconomic factors in the housing market, this study aimed to analyze the Finnish real estate market's socioeconomic and demographic principal components in terms of property price variations and construct a predictive pricing model based on that information. Considering the small scale of Finland's property market and the limited number of sales transactions across many areas of the country, to have a homogeneous market regarding population, infrastructure, and sales transactions volume, this study limited its focuses to the housing markets in five of Finland's most populated cities from 2013 to 2020: Helsinki, Espoo, Vantaa, Tampere, and Turku, to develop a predictive housing price model based on the potential socioeconomic and demographic factors. This study met its aims by addressing the research question as follows:

Is it possible to create a predictive housing price model for Helsinki, Espoo, Vantaa, Tampere, and Turku based on socioeconomic and demographic principal components?

The results of the correlation analysis indicated that the PSM blocks of flats are moderately positively correlated with the *increase of population Bachelor's and Master's degree holders*. It is also correlated with *1-person household* dwelling units, the number of *the establishment of enterprises, enterprises' turnover*, *property income*, *the average capital income tax*, and the *PSM price of old terraced houses*. The target variable is also *moderately negatively correlated with the demographic dependency* ratio. Since correlation does not necessarily imply causation, further analysis was conducted.

- The PCA result heatmap (figure 23) depicts the fair contribution of different socio-economic and demographic variables in the first three principal components. As a result, the PCA analysis findings imply that no specific socioeconomic and demographic variable dominantly influences the price.
- The first principal component captures almost 48 percent of the data variance and contributes significantly to price prediction models with an R-squared of 0.74 and a

P-value of nearly zero (1.9921e-08). Consequently, the thesis outcome provides adequate evidence to demonstrate that it is possible to create a predictive property price model based on socioeconomic and demographic principal components.

- Running K-fold cross-validation, the result indicates that the predictive model with one principal component outperforms other models. The selected model was then tested against the test data to see how accurate it was, which showed an estimated mean squared error of 2.6119e+05.

Accordingly, this study developed a predictive property price model. The predictive model with one socio-economic and demographic principal component outperformed other models.

5.4 Limitations and recommendations for further studies

This research investigated the real estate market dynamics in Finland to see if socioeconomic and demographic parameters significantly contribute to a predictive property price model regarding Helsinki, Espoo, Vantaa, Tampere, and Turku housing markets. This study accomplished its goal; however, there are a few points to consider when referring to the study's findings:

- Socio-demographic statistics can be challenging to investigate in connection to the housing market since they are subject to various macroeconomic and social changes, particularly during times of crisis, such as the Covid 19 pandemic. Another issue with demographic data is that it has a snapshot-like character. The most up-to-date data comes from recent years, while it might not be available when conducting analysis. Year-wise covariates would improve accuracy and make analyzing short-term changes in covariate coefficients safer. Each year, the model has its error variance, and demographics are expected to shift more slowly than apartment prices. For this reason, more investigation is needed, specifically regarding larger market scales and more extended time frames.

- Considering that residential, commercial, industrial, land, and particular purpose real estate are the five real estate forms, this study solely focused on residential apartments. Other research might explore the commercial, industrial, or general real estate markets in Finland.

- Further studies might also consider the rental market in the residential and commercial sectors, as this study concentrated on sales data from the residential sector.

- This research concentrated on the five most populated cities in Finland. A similar study can be implemented in a market with a larger sales transaction scale.

- Due to the lack of available data, this study could not compare housing market prices before and after the covid 19 pandemic; Besides, for the same reason, the scope of the study is limited to statistics from 2013 to 2020. Other studies might examine the influence of SDPC on house prices before and after the Covid 19 epidemic.

- This research examined selected socioeconomic and demographic parameters. Other studies may focus on other factors, such as job titles, or the industrial field of enterprises, to examine their contribution to the property price predictive model. Comparing different predictive models in the applied predictors is another possibility for further investigations.

References

Abraham, J.M., Hendershott, P.H. 1996. Bubbles in metropolitan housing markets. Journal of Housing Research. Vol. 7, pp.190-210.

Acral, A., Cadges, V., 2008, "Tasinmaz Degerlemesi", TMMOB Granita ve Dadastro M" Mühendisleri Odası, 2. Baskı, Ankara. pp 1-47.

Hort, K. 1998. The determinants of urban house price fluctuations in Sweden 1968-1994. Journal of housing economics. Vol. 7. pp 94-120.

Malpezzi, S. 1999. A simple error correction model of house price, Journal of Housing economics. Vol. 8, pp 25-63.

Newbold, P. & Carlson, W. & Thorne, B. (2006) Statistics for Business and Econometrics. 6th edition. Prentice-Hall. New Jersey.

Oikarinen, E. (2004). The diffusion of housing price movements from the center to surrounding areas. Journal of Housing Research. Vol. 15, issue 1. Pp. 3-38.

Oikarinen, E. (2007) Studies on housing price dynamics. Doctoral theses. Turku School of Economics. Series A-9:2007.

Oikarinen, E. (2009a) Household borrowing and metropolitan housing price dynamics-Empirical evidence from Helsinki. Journal of Housing Economics. Vol. 18, issue 2. Pp, 126-139.

Oikarinen, E. (2009b) Interaction between housing price and household borrowing: the Finnish case. Journal of Banking and Finance. Vol. 33, issue 4. pp. 747-756.

Online resources:

Aalto, H., Sinkkonen, J. 2021. How to estimate housing prices with demographic data. [Accessed October 28th, 2021] Available on:

https://www.reaktor.com/blog/how-to-estimate-housing-prices-with-demographic-data/

Abelson, P., Joyeux, R., Milunovich, G. and Chung, D. 2005. House price in Australia: 1970 to 2003 facts and explanations. [Accessed Nov 5th, 2021] Available on: https://www.proptechnow.com.au/wp-content/uploads/uploads/houseprices-1.pdf

Augustyniak, H., Leszczynki, R., Kaszek, J., Olszewski, K., and Waszczuk, J. 2015. On the dynamics of the primary housing market and the forecasting of house prices. [Accessed, Nov 14th, 2021], available on: https://www.bis.org/ifc/events/ws_micro_macro/olszewski_paper.pdf

Barbu, T.C., Vuta, M., Strachinaru, A.I.and Cioaca,S.I. (2017). An assessment of the immigration impact on the international housing price. Amfiteatru Economic Journal, ISSN 2247-9104, The Bucharest University of Economic Studies, Bucharest, Vol. 19, Iss.46, pp. 682-695 [Accessed, Dec 21st, 2021], available on: https://www.econstor.eu/bitstream/10419/169098/1/aej-v19-i46-p682.pdf?cv=1

Bank of Finland, 2021. Finnish economy takes off, but rapid growth just temporary. [Accessed September 28th, 2021] Available on:

https://www.suomenpankki.fi/en/media-and-publications/releases/2021/finnish-economy-takes-off--but-rapid-growth-just-temporary/

Binieli, M. 2018. Machine learning: an introduction to mean squared error and regression lines. freeCodeCamp Forum. [A accessed October 27th, 2021] Available on: https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/

Björklund, M. 2019. Be careful with your principal components. Evolution. International Journal of Organic Evolution. Vol 73. Issue 10. [Accessed April 4th, 2022] Available on: https://onlinelibrary.wiley.com/doi/10.1111/evo.13835

Boye, P. Mireku-Gyimah, D. 2018. Principal component regression model for estimating the price of a housing unit. Ghana Journal of Technology, Vol. 3., No. 1, pp. 17-23. [Accessed November 17th, 2021] Available on: https://www.researchgate.net/publication/336253846_Principal_Components_Regression_Model_for_Estimating_the_Price_of_a_Housing_Unit

Brotherus, J. 2021. HYPO. Hypon asuntomarkkinakatsaus. [Accessed September 24th, 2021] Available on: https://www.hypo.fi/wp-content/uploads/2021/02/Hypon-Asuntomarkkinakatsaus_01_2021.pdf

Brownlee, J. 2020. Machine learning mastery. Time series forecasting with Python, Root mean squared error. [Accessed October 27th, 2021] Available on: https://machinelearningmastery.com/time-series-forecasting-performance-measures-with-python/

Brownlee, J. 2020. How to configure k-fold cross-validation. [Accessed October 27th, 2021] Available on: https://machinelearningmastery.com/how-to-configure-k-fold-cross-validation/

Chen. J; Scott. G; Logan. M, 2021. Real Estate. Real Estate Investigating Guide. [Accessed September 25th, 2021] Available on: https://www.investopedia.com/terms/r/realestate.asp

City of Turku, 2021. [Accessed September 25th, 2021] Available on: https://www.turku.fi/en/europeade-2017/arriving-turku

Findicator, 2021. The age structure of the population. [Accessed September 27th, 2021] Available on: https://findikaattori.fi/en/14

Feng, M. 2019. How will proximity to higher education affect the rental housing market? A case study in Beijing. KTH Royal Institute of Technology. [Accessed Nov 17th, 2021] Available on: https://www.diva-portal.org/smash/get/diva2:1335037/FULLTEXT01.pdf

Feng, Y. Taewon, K. & Lee,D. 2018. Housing price and population changes: growing vs. shrinking cities. Journal of Accounting and Finance Research. Vol.7, No:4. [Accessed October 18th, 2021] Available on:

https://www.sciedu.ca/journal/index.php/afr/article/view/14139/8811

Goonewardana, H. 2019. Apprentice Journal. PCA: application in machine learning.

[Accessed October 27th, 2021] Available on: https://medium.com/apprentice-journal/pca-application-in-machine-learning-4827c07a61db

Green, L. 2018. The effect of immigration on UK house prices. [Accessed October 24th, 2021] Available on:

https://www.nottingham.ac.uk/economics/documents/research-first/luke-green.pdf

Gupta, R. Kabundi, A. 2010. Forecasting real U.S. house prices: Principal Components versus Bayesian regression. University of Pretoria, South Africa. International Business & Economic Research Journal. Vol. 9, No. 7. [Accessed Oct 6th, 2021]. Available on:

https://clutejournals.com/index.php/IBER/article/download/605/591/2377

Hornstein, A. 2009. Problems for a fundamental theory of house prices. Economic Quarterly. Volume 95. Pp 1-24. [Accessed October 24th, 2021] Available on: https://www.richmondfed.org/-/media/RichmondFedOrg/publications/research/economic_quarterly/2009/winter/pdf/hornstein.pdf

Helsinki Times, 2021. Hypo expects the Finnish capital region to witness a rapid rise in house prices. [Accessed September 25th, 2021] Available on: https://www.helsinkitimes.fi/finland/finland-news/domestic/18754-hypo-expects-finnish-capital-region-to-witness-rapid-rise-in-house-prices.html

Ismail, M. Warsame, A and Wilhelmsson, M. 2021. An exploratory analysis of housing and the distribution of covid-19 in Sweden. [Accessed December 26th, 2021] Available on: https://www.diva-portal.org/smash/get/diva2:1607730/FULLTEXT02

Jin, P. and Tang, Q.G. 2018. Research on the influence factors of housing price based on a semi-parametric functional linear model. Proceedings of the 4th annual international conference on management, economics, and social development (ICMESD 2018). Advances in Economics, Business, and Management Research (AEBMR), Vol. 60, Atlantis Press. [Accessed November 24th, 2021] Available on:

https://www.atlantis-press.com/proceedings/icmesd-18/25898747

Jolliffe, I.T. and Cadim, J. 2016. Principal component analysis, a review, and recent developments. Publisher: The Royal Society. [Accessed October 24th, 2021] Available on: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202

Khadem, N. 2021. Housing is expensive now, imagine a market with more migrants. Economists see rent house prices lift in 2023. [Accessed Nov 19th, 2021]. Available on:

https://www.abc.net.au/news/2021-11-17/migration-house-prices-rent-housing-affordability-property/100625806

Keskinen, P., Karikallio, H. & Kiviholma, S. 2020. Housing market 2020 forecast. Regional housing market forecast for 2020: Housing price developments are becoming increasingly polarized- rent developments are becoming more consistent. [Accessed September 25th, 2021] Available on: https://www.ptt.fi/ajankohtaista/uutiset/asuntomarkkinat-2020-ennuste.html

KTI, 2021. The Finnish Property Market. 2021. [Accessed September 23rd, 2021] Available on: https://kti.fi/wp-content/uploads/The_Finnish_Property_Market_2021.pdf

Kokkinen, E. 2020. Development of housing prices outside and at the vertices of the Finnish growth triangle. [Accessed Nov 8th, 2021]. Available on: https://lutpub.lut.fi/bitstream/handle/10024/162001/progradu_kokkinen_eetu.pdf?cv=1&isAllowed=y&sequence=1

Laakso, S., Kostiainen, E. and Metsäranta, H. (2016). Liikennehankkeiden laajemmat taloudelliset vaikutukset. Kansantaloudellinen aikakauskirja -112. - 4/2016. [Accessed November 4th, 2021] Available on: https://docs.iza.org/dp11497.pdf

Lalaine C. Delmendo. 2022. Finnish housing market strengthening. [Accessed Feb 6th, 2022] Available on:

https://www.globalpropertyguide.com/Europe/Finland/Price-History

Larkin, M.P. Askarov, Z. Doucouliagos, C. Dubelaar, C. Klona, M., Newton, J., Stanley, T.D. 2018. Do house prices sink or ride the wave of immigration? IZA Institute of LAbour Economics. [Accessed Nov 26th, 2021] Available on: https://docs.iza.org/dp11497.pdf

Lee, I. 2021. Four reasons why correlation does not imply causation. Towards Data Science. [Accessed November 16th, 2021] Available on: https://towardsdatascience.com/4-reasons-why-correlation-does-not-imply-causation-f202f69fe979

Li, Y. & Zhang, Q. 2011. The application of Principal Component Analysis on Financial Analysis in Real Estate Listed Company. [Accessed November 16th, 2021] Available on: https://www.researchgate.net/publication/271881798_The_Application_of_Principal_Component_Analysis_on_Financial_Analysis_in_Real_Estate_Listed_Company/citations

Lindblad, A., Sariola, M. & Viertola, H. (2019). What factors influence house prices and residential construction? Bank of Finland Bulletin [Accessed November 4th, 2021] Available on: https://www.bofbulletin.fi/en/2019/3/what-factors-influence-house-prices-and-residential-construction/

Lönnqvist, H. 2015. On the effect of urban natural amenities, architectural quality, and accessibility to workplaces on housing prices. An empirical study on the Helsinki Metropolitan Area. The city of Helsinki. Urban Facts. [Accessed November 4th, 2021] Available on: https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/16_02_04_Tutkimuksia_5_2015_Lonnqvist.pdf

Manninen, M. 2021. Predicting the development of housing prices in Westmetro's phase 2 areas. [Accessed November 15th, 2021] Available on: https://lutpub.lut.fi/bitstream/handle/10024/162939/Masters%20thesis_Manninen_Melisa.pdf?cv=1

Mostofi, F., Togan, V. and Basaga. H.B. 2021. House price prediction: A data-centric aspect approach on performance of combined principal component analysis with deep neural network model. Karadeniz Technical University. Department of civil engineering. Journal of Construction Engineering, Management, and Innovation, Vol 4, Issue 2. Pages 106-116. [Accessed Jan 11th, 2022]. Available on: https://www.researchgate.net/publication/352869140_House_price_prediction_A_data-centric_aspect_approach_on_performance_of_combined_principal_component_analysis_with_deep_neural_network_model

Metz, N.E. 2015. Effect of distance to schooling on home prices. University of Central Oklahoma, USA. Department of Economics. The Review of Regional Studies, The Official Journal of the Southern Regional Sciences Association. [Accessed Nov 5th, 2021]. Available on:

https://www.researchgate.net/publication/284746261_Effect_of_Distance_to_Schooling_on_Home_Prices

Nisanci, R. Yomralioglu, T. and Yildirim, V. 2007. An Implementation of Nominal Asset Based Land Readjustment. [Accessed Nov 5th, 2021]. Available on:

https://www.researchgate.net/publication/312289572_An_Implementation_of_Nominal_Asset_Based_Land_Readjustment

Oladunni, T. & Sharma, Sh. (2016), Hedonic Housing Theory_ a machine learning investigation. [Accessed Nov 15th, 2021]. Available on: file:///C:/Users/shima/Downloads/HedonicPricingTheory.pdf

Peng. C.W., Tsai, I.CJin. 2017. The effect of demographic structure change on housing price. Taiwan Economic Review, 45. Pp, 163-192. [Accessed Nov 15th, 2021]. Available on: https://econ.ntu.edu.tw/ter/new/data/new/TER45-1/TER451_4Abstract%E8%8B%B1.pdf

Putkuri, H. 2018. Wide regional disparities in Finnish house prices and household indebtedness. [Accessed November 20th, 2021] Available on: https://www.bofbulletin.fi/en/2018/2/wide-regional-disparities-in-finnish-house-prices-and-household-indebtedness/

PTT (2020) Pellervo Economic Research Center. Asuntomarkkinat 2020 erikoisteema – Asuntomarkkinoiden polarisaatio jatkuu 2020-luvulla. [online document]. [Accessed September 22nd, 2021] Available on:

https://www.ptt.fi/julkaisut-ja-hankkeet/uutiset/asuntomarkkinat-2020-erikoisteema-asuntomarkkinoiden-polarisaatio-jatkuu-2020-luvulla.html

Renee, T. 2021. How to impose principal component analysis on house price regression. [Accessed Dec 2nd, 2021] Available on:

https://medium.com/mlearning-ai/how-to-impose-principal-component-analysis-on-a-house-price-regression-254b8dfd6ddc

Saifi, A. 2021. An investigation of economic drivers of the Finnish housing market. Metropolia University of Applied Sciences. [Accessed Jan7th, 2022] Available on: https://www.theseus.fi/bitstream/handle/10024/508212/Saifi_Adibullah.pdf.pdf?cv=1&isAllowed=y&sequence=2

Saylor Academy. 2012. Macroeconomics theory through applications. Housing supply and demand. [Accessed September 16th, 2021] Available on:

https://saylordotorg.github.io/text_macroeconomics-theory-through-applications/s08-01-housing-supply-and-demand.html

Scott, G, and Logan, M, 2021. Investopedia. Real Estate. [Accessed September 25th, 2021] Available on: https://www.investopedia.com/terms/r/realestate.asp

Sisman, Y. and Sisman, A. 2016. Principal component analysis approach in the determination of house value. [Accessed Oct 4th, 2021] Available on: https://www.researchgate.net/publication/303915471_Principal_Component_Analysis_Approach_in_the_Determination_of_House_Value

Statista, 2021. Largest cities in Finland in 2020 by the number of inhabitants. [Accessed September 26th, 2021] Available on: statista.com/statistics/327469/largest-cities-in-Finland

Statistics Finland, 2021. Establishment of Enterprises by city 2013-2019. [Accessed Oct 14th, 2021] Available on:

https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__yri__alyr/statfin_alyr_pxt_11dc.px/

Statistics Finland, 2021. Income and consumption. [Accessed September 26th, 2021] Available on:

https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__tul__tjt__asuntokuntien/statfin_tjt_pxt_118w.px/table/tableViewLayout1

Statistics Finland, 2021. Key figures on population by region. [Accessed October 12th, 2021] Available on:

https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__vrm__vaerak/statfin_vaerak_pxt_11ra.px/table/tableViewLayout1/

Statistics Finland, 2021. Number and value of house sales 2015-2020. [Accessed Oct 16th, 2021] Available on:

https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__asu__ashi__vv/statfin_ashi_pxt_136a.px/

Statistics Finland, 2021. Number, income, and income structure of household-dwelling units by city, 1995-2019. [Accessed Oct 17th, 2021] Available on:

https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__tul__tjt__asuntokuntien/statfin_tjt_pxt_118w.px/

Statistics Finland, 2021. Population. [Accessed September 25th, 2021] Available on:

https://www.stat.fi/tup/suoluk/suoluk_vaesto_en.html

Statistics Finland, 2021. Population aged 15 or over by the level of education, city, gender, and age. [Accessed Oct 15th, 2021] Available on:

https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__kou__vkour/statfin_vkour_pxt_12bq.px/

Statistics Finland, 2021. Price of dwellings in housing companies. [Accessed Oct 13th, 2021] Available on:

https://www.stat.fi/til/ashi/2021/08/ashi_2021_08_2021-09-29_tie_001_en.html

Valtioneuvosto, Ministry of Finance, 2021. Expected return in Finland's economy now at hand. [Accessed September 29th, 2021] Available on: https://valtioneuvosto.fi/en/-/10623/expected-upturn-in-finland-s-economy-now-at-hand

Vero, 2021. Income and deductions used for calculating the tax rate for 2021. [Accessed September 25th, 2021] Available on:https://www.vero.fi/en/individuals/tax-cards-and-tax-returns/tax_card/tax-base-details-for-2020/

Vero, 2021. List of income tax rates of cities and church parishes in 2020. [Accessed September 25th, 2021] Available on:

https://www.vero.fi/contentassets/653c812344f84cd1ab790e899e989305/kunnat2019s.pdf

Vuori. P, 2020. Helsinki area housing production breaks records. An overview of construction and population trends in 1961-2019. [Accessed October 5th, 2021] Available on: https://www.kvartti.fi/en/articles/helsinki-area-housing-production-breaks-records

https://www.worldometers.info/world-population/finland-population/

Wen, H., Zhang, Y. and Zhang, L. 2013. Do educational facilities affect the housing price? An empirical study in Hangzhou, China. Elsevier. Vol 42. Pp. 155-163. [Accessed Nov 18th, 2021] Available on:

https://www.sciencedirect.com/science/article/abs/pii/S019739751300129X

Wicklin, R. 2017. Should we use principal component regression? [Accessed Nov 18th, 2021] Available on: https://blogs.sas.com/content/iml/2017/10/25/principal-component-regression-drawbacks.html

World Bank, 2021. GDP and GDP per capita- Finland. [Accessed September 25th, 2021] Available on: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=FI

Worldometer, 2021. Finland population (live). [Accessed Oct 19th, 2021] Available on:

Yle NEWS, 2021. Prices of older detached housed up 5% during Q3. [Accessed Nov 18th, 2021] Available on: https://yle.fi/news/3-12192983

Appendix 1: Figures and tables

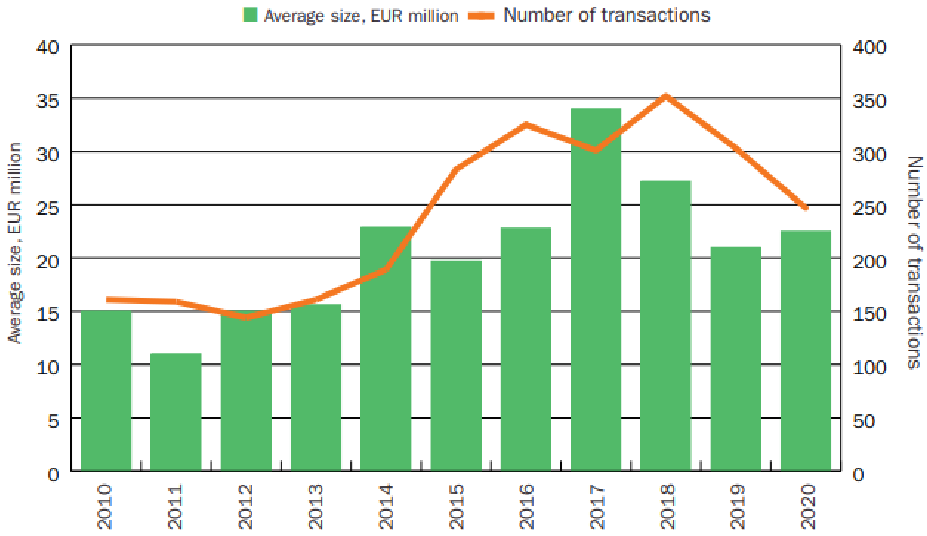

Figure 7. Number and the average size of property transactions 2010-2020 (KTI 2021)

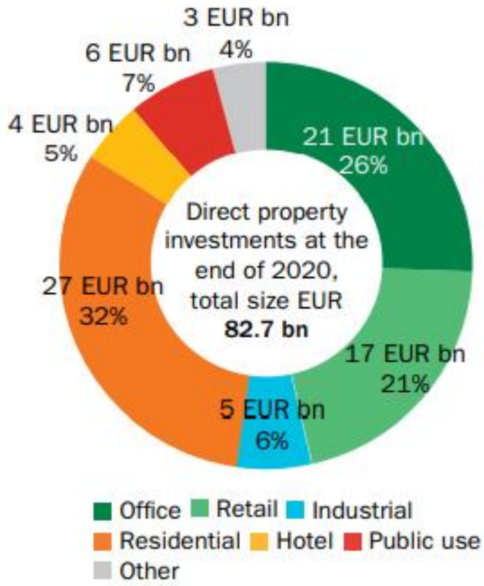

Figure 9. Finnish investment market structure by sector (KTI 2021)

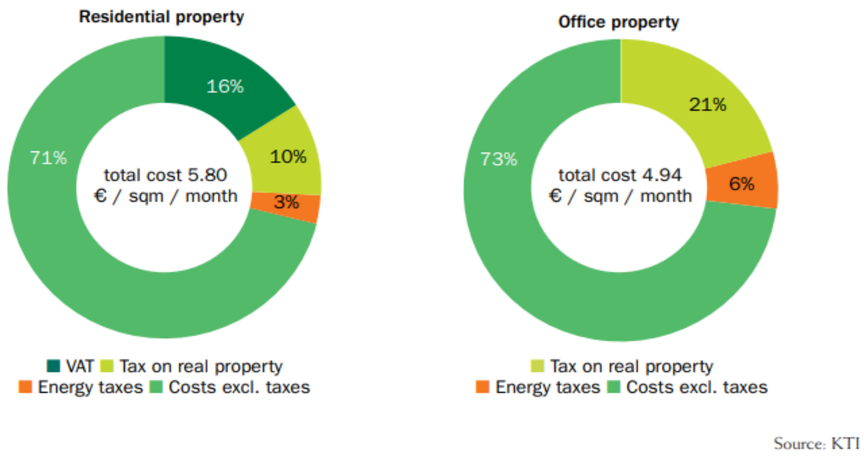

Figure 11. The share of taxes of operational costs in Helsinki in 2019(KTI 2021)

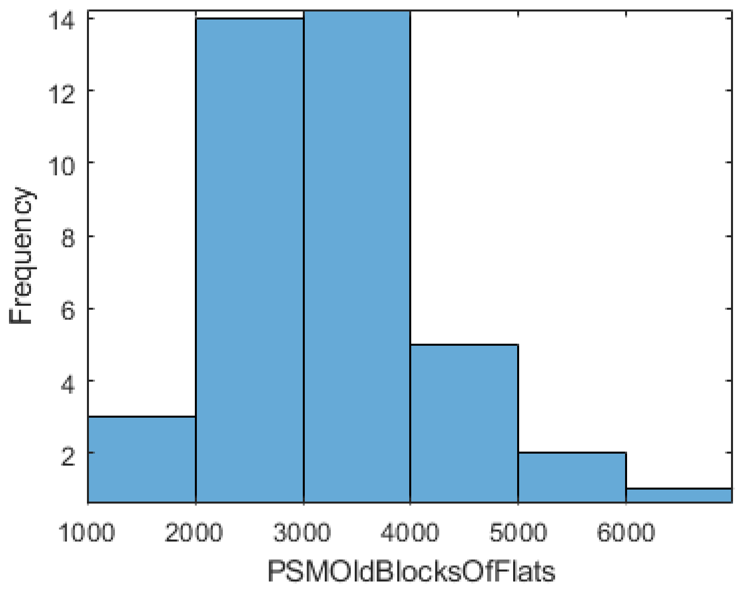

Figure 13. Histogram of the target variable

Table 5: Data variables and explanations

1	**Increase of population**	Increase of population, or total change is the difference between the population of two consecutive statistical reference periods. The difference has been divided by the population of the previous statistical reference period and then multiplied by one hundred.
2	**under 15, %**	Share of persons aged under 15, %
3	**15 to 64, %**	Share of persons aged 15 to 64, %
4	**65 or over, %**	Share of persons aged 65 or over, %
5	**Demographic dependency ratio**	The demographic dependency ratio is calculated as the total number of persons age 0-14 and the number of persons aged 65 and older divided by the number of persons aged 15-64. The figure obtained is multiplied by one hundred.
6	**Economic dependency ratio**	The percentage of economic dependency following the above method
7	**Average age, both sexes**	The average age is calculated by summing the ages of the persons and dividing the amount by the number of persons. The age of a person is defined as a person's age in whole years plus 0.5 years.
8	**Average age, men**	The average age is calculated by summing the ages of the persons and dividing the amount by the number of persons. The age of a person is defined as a person's age in whole years plus 0.5 years.
9	**Average age, women**	The average age is calculated by summing the ages of the persons and dividing the amount by the number of persons. The age of a person is defined as a person's age in whole years plus 0.5 years.
10	**Upper secondary degree holders #**	Upper secondary education
11	**Bachelor degree holders #**	Total number of people with Bachelor's or equivalent level
12	**Master degree holders #**	Total number of people with Master's or equivalent level
13	**PhD degree holders #**	Total number of people with doctoral or equivalent level
14	**Finnish speakers #**	The total number of Finnish speakers in each city
15	**Swedish speakers #**	The total number of Swedish speakers in each city
16	**Sami speakers #**	The number of Sami speakers in each city
17	**Foreign-language speakers #**	Languages (ISO 639-1) The foreign languages are other than Finnish, Swedish and Sami
18	**Total household dwelling units (all**	Total number of all types of household units in each city
19	**Terraced houses #sales**	Total number of terraced houses sales in each city
20	**Blocks of flats #sales**	Total number of blocks of flats sales in each city
21	**1 person HDU**	Total number of one-person household dwelling units
22	**2 person HDU**	Total number of two-persons household dwelling units
23	**3 person HDU**	Total number of three-persons household dwelling units
24	**4+ person HDU**	Total number of four-persons household dwelling units
25	**Employed labour force**	Total number of employed labour force in each city
26	**New terraced houses transactions**	The total number of transactions of new terraced houses in housing companies
27	**New blocks of flats transactions**	Number of transactions of new Blocks of flats in housing companies
28	**Sales (Total buildings)**	Total number of sales of total building types
29	**Establishments of enterprises**	Number of establishment of new enterprises in total
30	**Enterprises turnover**	Sales profit from actual activity, after deduction of granted discounts, value added tax and other taxes (EUR 1,000)
31	**Entrepreneurial total income**	Average annual total earned income of entrepreneurs (EUR 1,000)
32	**Entrepreneurial income from business activities total, mean**	Average annual total earned income of entrepreneurs from business activities
33	**Property income**	Property income total, mean (EUR 1,000)
34	**Dividend income, mean**	Average annual income from dividend, (EUR 1,000)
35	**Sales profits, mean**	Average annual income from profit of sales, (EUR 1,000)
36	**Old age security, mean**	Average annual old age security, (EUR 1,000)
37	**Unemployment security, mean**	Average annual unemployement security, (EUR 1,000)
38	**Housing benefits, mean**	Average annual housing benefits, (EUR 1,000)
39	**Gross income, mean**	Average annual gross income, (EUR 1,000)
40	**Income taxes, mean**	Average annual taxes on earned income, (EUR 1,000)
41	**Capital income tax, mean**	Average annual taxes on capital income, (EUR 1,000)
42	**Municipal tax, mean**	Average annual municipal tax, (EUR 1,000)
43	**Disposal cash income, mean**	Average annual disposal cash income, (EUR 1,000)
44	**PSM old blocks of flats**	weighted geometric averages of square meter prices of old blocks of flats in housing companies (EUR/m2)

Table 7. Summary statistics for data's numerical variables

	Mean	Mode	Median	Std	Min	Max
Increase of population	3965	...	3518	1938	1429	8696
Share of under 15	15,5	13,7	14,2	2,6	12,6	19,6
Share of 15 to 64	67,5	67,4	67,4	1	66,1	70,1
Share of 65 or over	17	14,5	16,8	2,3	13,2	20,9
Demographic dependency ratio	48,2	48,4	48,5	2,2	42,6	51,4
EcoEnomic dependency ratio	120,91	135,8	114,9	12,9	103,2	142
Average Age both sex	40,1	41,7	40,5	1,3	37,7	40,2
Average age, men	38,6	38,7	38,9	1,1	36,4	40,4
Average age, women	41,5	43,4	42	1,5	39	43,5
Upper secondary #	65956	68620	67789	52094	61	178283
Bachelor degree holders #	33711	23227	24576	20588	17486	86162
Master's degree holders #	37114	19895	22346	29190	14136	108127
PHD degree holders #	4722	3452	3412	3427	1302	13144
Finnish speakers	253507	...	209353	129764	155573	511043
Swedish speakers	14699	...	10243	12638	1172	36754
Sami speakers	26	17	17	20	9	71
Foreign-language speakers	42784	...	34140	29159	13835	109254
Total household dwelling units (all type	156320	...	117751	88088	95399	344898
Terraced houses#sales	1055	913	995	297	660	1704
Blocks of flats#sales	3985	9047	2788	2920	1886	10365
1 person HDU	72678	...	52867	45250	36354	172044
2 person HDU	48646	...	37960	26550	31111	103697
3 person HDU	16480	...	13162	9029	8221	34031
4+ person HDU	17162	11659	15235	9106	6090	35391
Employed labour force	170392	413677	119156	114457	93946	413677
New terraced houses transactions	367	...	153	423	13	1441
New blocks of flats transactions	1659	...	1124	1972	162	9883
Sales (Total buildings)	6423	...	5133	3398	2802	15466
Establishments of enterprises	18648	...	13769	14608	812	52184
Enterprises turnover	35438806	15506824	32476403	24200216	11595581	90992707
Entrepreneurial total income	1393	1321	1328	145	1130	1710
Entrepreneurial income from business	1091	1010	1056	93	979	1317
Property income	5335	...	4602	2075	2828	9309
Dividend income, mean	1041	...	823	753	490	3382
Sales profits, mean	1073	772	905	685	471	4305
Old age security, mean	7283	...	7856	2390	934	9515
Unemployment security, mean	1568	1723	1613	370	494	2186
Housing benefits, mean	896	884	916	143	562	1111
Gross income, mean	56887	...	58799	11944	40060	78716
Income taxes, mean	2899	...	2456	1568	806	6101
Capital income tax, mean	1077	1811	846	458	490	1983
Municipal tax, mean	7203	...	7671	1753	3753	10287
Disposal cash income, mean	42009	36379	40418	7556	32573	56082
PSM old blocks of flats	3241	...	3089	1059	1782	6288

Table 8. Correlation analysis

Increase of popder 15, to 64, or overc depedepene Age bge agee age, r secoracheloMaster PHD sh specish spei speanguagelwelling house of flatserson Herson Herson Herson hd labo houseof flats otal bents of ises tuurial tom busierty incl incomrofits, securitent secbenefincome taxes,come tpal taxish incl blocks of flats

Increase o	1
under 15,	0,24	1
15 to 64, %	0,57	-0,43	1
65 or over	-0,51	-0,92	0,04	1
Demograp	-0,56	0,43	-1,00	-0,04	1
EcoEnomi	-0,73	-0,58	-0,21	0,73	0,20	1
Average A	-0,34	-0,98	0,27	0,96	-0,28	0,61	1
Average a	-0,35	-0,96	0,25	0,95	-0,25	0,59	0,99	1
Average a	-0,34	-0,98	0,27	0,97	-0,28	0,64	1,00	0,98	1
Upper sec	0,34	-0,18	0,18	0,12	-0,18	-0,23	0,20	0,23	0,18	1
Bachelor#	0,78	-0,16	0,69	-0,12	-0,68	-0,53	0,07	0,06	0,07	0,62	1
Master#	0,80	-0,07	0,63	-0,20	-0,62	-0,55	-0,02	-0,04	-0,01	0,57	0,99	1
PHD#	0,74	-0,13	0,61	-0,11	-0,59	-0,47	0,06	0,04	0,07	0,58	0,97	0,99	1
Finnish sp	0,73	-0,21	0,78	-0,11	-0,78	-0,43	0,11	0,11	0,11	0,53	0,90	0,88	0,88	1
Swedish s	0,69	0,05	0,48	-0,26	-0,47	-0,50	-0,08	-0,10	-0,08	0,42	0,81	0,85	0,87	0,86	1
Sami spea	0,70	-0,26	0,80	-0,06	-0,80	-0,43	0,19	0,20	0,17	0,54	0,88	0,82	0,82	0,97	0,80	1
Foreign-la	0,74	0,07	0,55	-0,32	-0,54	-0,58	-0,10	-0,08	-0,13	0,57	0,82	0,81	0,81	0,91	0,90	0,90	1
Total hous	0,69	-0,26	0,77	-0,05	-0,76	-0,39	0,18	0,18	0,17	0,55	0,89	0,86	0,87	1,00	0,86	0,97	0,91	1
Terraced h	0,60	0,71	-0,09	-0,75	0,11	-0,64	-0,74	-0,76	-0,72	0,26	0,41	0,51	0,47	0,33	0,54	0,19	0,49	0,28	1
Blocks of f	0,66	-0,37	0,82	0,05	-0,81	-0,34	0,28	0,28	0,28	0,54	0,90	0,86	0,87	0,98	0,83	0,97	0,85	0,99	0,18	1
1 person h	0,61	-0,40	0,79	0,10	-0,78	-0,28	0,32	0,32	0,31	0,55	0,87	0,82	0,84	0,97	0,81	0,96	0,86	0,99	0,15	0,99	1
2 person h	0,71	-0,25	0,78	-0,07	-0,77	-0,40	0,16	0,16	0,15	0,54	0,90	0,87	0,87	1,00	0,86	0,97	0,91	1,00	0,29	0,99	0,98	1
3 person h	0,80	0,00	0,70	-0,30	-0,69	-0,56	-0,08	-0,07	-0,08	0,51	0,89	0,88	0,87	0,97	0,91	0,94	0,96	0,97	0,48	0,93	0,91	0,97	1
4+ person	0,74	0,20	0,42	-0,41	-0,41	-0,60	-0,24	-0,22	-0,25	0,66	0,80	0,81	0,80	0,86	0,87	0,80	0,95	0,85	0,67	0,77	0,77	0,85	0,92	1
Employed	0,70	-0,24	0,78	-0,08	-0,77	-0,43	0,16	0,16	0,15	0,54	0,89	0,86	0,86	0,99	0,86	0,98	0,92	1,00	0,27	0,98	0,98	1,00	0,97	0,85	1
New terra	0,22	0,18	0,35	-0,35	-0,35	-0,28	-0,28	-0,32	-0,26	-0,52	0,15	0,20	0,17	0,16	0,19	0,11	0,08	0,12	0,13	0,13	0,08	0,13	0,18	-0,04	0,14	1
New block	0,45	-0,17	0,64	-0,09	-0,63	-0,19	0,04	0,01	0,06	-0,45	0,35	0,35	0,34	0,43	0,34	0,41	0,25	0,40	-0,03	0,42	0,40	0,41	0,39	0,08	0,41	0,62	1
Sales (Tot	0,64	-0,21	0,63	-0,04	-0,62	-0,36	0,16	0,18	0,15	0,66	0,83	0,79	0,81	0,95	0,83	0,92	0,92	0,96	0,34	0,93	0,95	0,96	0,93	0,89	0,95	-0,05	0,24	1
Establishn	0,79	-0,18	0,75	-0,13	-0,74	-0,54	0,08	0,06	0,08	0,53	0,97	0,96	0,94	0,89	0,80	0,87	0,78	0,87	0,35	0,90	0,84	0,88	0,87	0,74	0,88	0,23	0,43	0,77	1
Enterprise	0,87	0,15	0,59	-0,43	-0,58	-0,77	-0,24	-0,25	-0,25	0,48	0,93	0,94	0,90	0,81	0,80	0,78	0,80	0,78	0,58	0,77	0,71	0,79	0,86	0,80	0,80	0,29	0,36	0,70	0,92	1
Entrepren	0,56	0,38	0,10	-0,46	-0,08	-0,59	-0,44	-0,49	-0,40	0,23	0,55	0,66	0,62	0,41	0,58	0,27	0,43	0,37	0,80	0,33	0,27	0,38	0,48	0,56	0,36	0,26	0,13	0,34	0,55	0,67	1
Entrepren	0,35	0,38	-0,19	-0,33	0,21	-0,32	-0,35	-0,34	-0,34	0,27	0,35	0,44	0,47	0,26	0,49	0,15	0,42	0,26	0,67	0,17	0,19	0,25	0,35	0,52	0,24	-0,06	-0,06	0,33	0,20	0,39	0,56	1
Property i	0,53	0,35	0,03	-0,39	-0,01	-0,47	-0,38	-0,41	-0,35	0,33	0,58	0,69	0,70	0,46	0,70	0,30	0,52	0,44	0,79	0,38	0,35	0,44	0,54	0,64	0,42	0,13	0,07	0,45	0,49	0,64	0,80	0,82	1
Dividend i	0,36	0,11	0,41	-0,29	-0,40	-0,24	-0,22	-0,28	-0,19	-0,52	0,25	0,33	0,31	0,27	0,37	0,18	0,15	0,23	0,17	0,23	0,19	0,24	0,28	0,03	0,24	0,82	0,78	0,04	0,32	0,36	0,30	0,08	0,34	1
Sales profi	0,24	0,26	0,12	-0,33	-0,12	-0,21	-0,33	-0,36	-0,30	-0,35	0,16	0,25	0,26	0,16	0,34	0,04	0,13	0,13	0,32	0,10	0,08	0,13	0,20	0,09	0,13	0,64	0,44	0,03	0,14	0,28	0,30	0,35	0,55	0,83	1
Old age se	0,33	0,17	0,06	-0,21	-0,05	-0,47	-0,21	-0,26	-0,19	0,06	0,28	0,34	0,27	0,05	0,14	0,03	-0,01	0,01	0,31	0,07	-0,03	0,03	0,08	0,07	0,02	0,20	0,08	-0,09	0,41	0,43	0,66	-0,07	0,26	0,18	0,00	1
Unemploy	-0,03	-0,04	0,13	-0,01	-0,14	0,06	-0,01	-0,06	0,01	-0,08	-0,07	-0,06	-0,13	-0,11	-0,21	-0,09	-0,25	-0,14	-0,08	-0,08	-0,15	-0,12	-0,14	-0,18	-0,14	0,04	-0,04	-0,23	0,10	-0,03	0,15	-0,57	-0,30	-0,07	-0,33	0,64	1
Housing b	-0,07	-0,47	0,00	0,52	0,00	0,17	0,53	0,58	0,51	0,70	0,20	0,10	0,13	0,18	-0,04	0,25	0,18	0,22	-0,18	0,23	0,30	0,21	0,09	0,20	0,21	-0,63	-0,39	0,37	0,14	-0,04	-0,14	-0,09	-0,17	-0,68	-0,59	-0,10	-0,01	1
Gross inco	0,57	0,74	-0,10	-0,77	0,11	-0,77	-0,77	-0,79	-0,75	0,08	0,34	0,46	0,39	0,20	0,45	0,10	0,34	0,15	0,86	0,09	0,02	0,16	0,36	0,48	0,16	0,27	0,03	0,13	0,34	0,61	0,85	0,54	0,74	0,31	0,38	0,62	0,11	-0,39	1
Income ta	0,54	0,68	-0,06	-0,72	0,07	-0,66	-0,73	-0,77	-0,70	0,02	0,34	0,47	0,41	0,21	0,48	0,08	0,31	0,16	0,85	0,11	0,03	0,17	0,35	0,45	0,16	0,32	0,09	0,12	0,35	0,58	0,87	0,54	0,77	0,40	0,45	0,62	0,14	-0,47	0,98	1
Capital inc	0,55	0,32	0,06	-0,38	-0,05	-0,50	-0,36	-0,40	-0,33	0,34	0,61	0,72	0,71	0,47	0,69	0,32	0,51	0,45	0,79	0,40	0,37	0,45	0,54	0,64	0,43	0,14	0,08	0,44	0,55	0,67	0,86	0,72	0,98	0,34	0,51	0,41	-0,15	-0,13	0,77	0,80	1
Municipal	0,54	0,73	-0,06	-0,78	0,07	-0,75	-0,78	-0,80	-0,76	-0,03	0,26	0,37	0,28	0,14	0,35	0,05	0,24	0,08	0,77	0,03	-0,05	0,10	0,28	0,38	0,09	0,32	0,07	0,03	0,31	0,55	0,79	0,34	0,57	0,32	0,30	0,71	0,31	-0,45	0,97	0,95	0,63	1
Disposal c	0,59	0,51	0,12	-0,61	-0,11	-0,75	-0,55	-0,55	-0,55	0,22	0,46	0,52	0,46	0,33	0,48	0,29	0,44	0,30	0,66	0,26	0,20	0,31	0,45	0,50	0,32	0,24	0,04	0,24	0,47	0,66	0,65	0,34	0,56	0,29	0,33	0,52	0,07	-0,24	0,77	0,74	0,60	0,75	1
PSM old b	0,63	0,08	0,36	-0,25	-0,35	-0,44	-0,10	-0,08	-0,12	0,64	0,74	0,74	0,74	0,77	0,77	0,73	0,84	0,77	0,52	0,71	0,72	0,77	0,81	0,86	0,76	0,00	0,07	0,82	0,65	0,69	0,39	0,53	0,58	0,04	0,14	-0,09	-0,35	0,22	0,30	0,28	0,54	0,17	0,36	1

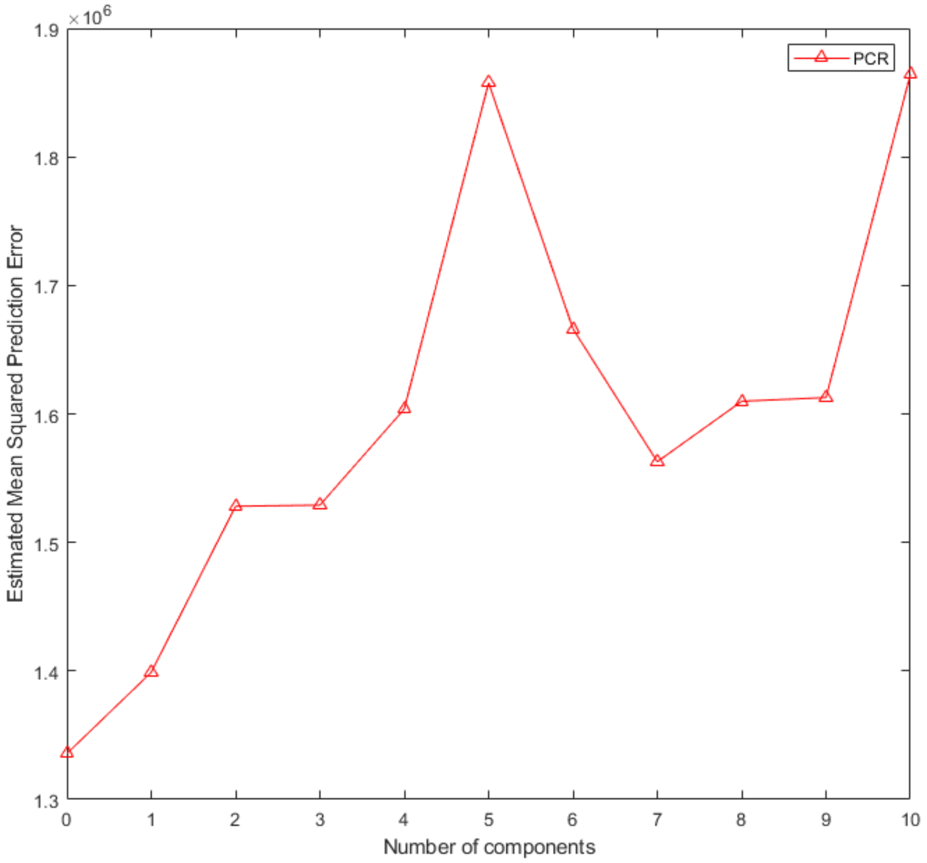

Figure 24. K-fold cross-validation result

PCR

	Estimate	SE	Tstat	pValue	R-squared	F-stat	P-value
Model1					0.74	67.4	1.99e-08
(Intercept)	3281.4	124.05	26.452	2.8952e-19			
x1	201.54	24.55	8.2093	1.9921e-08			

Figure 25. PCR model's result