OULUN YLIOPISTO
UNIVERSITY of OULU

OULU BUSINESS SCHOOL

Samu Heikkilä

SIMPLE VS COMPLEX MODELS IN HOUSING MARKET FORECASTING: EMPIRICAL EVIDENCE FROM HELSINKI METROPOLITAN AREA

Master's Thesis

Finance

Spring 2020

UNIVERSITY OF OULU
Oulu Business School

ABSTRACT OF THE MASTER'S THESIS

Unit			
Faculty of Finance			
Author: Heikkilä Samu		Supervisor: Koivuranta Matti	
Title: Simple vs Complex Models in Housing Market Forecasting: Empirical Evidence from Helsinki Metropolitan Area			
Subject	Type of the degree	Time of publication	Number of pages
Finance	Master of Science	May 2020	69

Abstract

This study seeks to examine whether it is possible to gain similar forecasting performance from simple forecasting models compared to more complex specifications in housing market context. Evaluation is conducted by comparing the predictive power of five common modelling techniques out-of-sample: Autoregressive Integrated Moving Average (ARIMA), Simple Regression (SR), Multiple Regression (MR), Vector Autoregression (VAR) and Autoregressive Integrated Moving Average with a vector of explanatory variables (ARIMAX).

A set of macroeconomic variables is used with these different modelling techniques to generate *ex-post* (out-of-sample) forecasts for the housing market of Helsinki Metropolitan Area. The dataset employed in this study is gathered from public sources and covers a period from 1999 to 2018. The *ex-post* forecasts are generated one, two, three, four and five steps ahead, i.e. from 2016 H2 to 2018 H2, and the forecasting accuracy is assessed by calculating Theil´s U and root-mean-square error (RMSE) values for each of the forecasts.

The obtained results imply that added model complexity does not necessarily yield better results, as the more complex run the risk of overfitting small data samples. What is more, the results indicate that while the complex models tend to fit historic data with greater accuracy, the higher historical fit does not always translate into superior forecasting results. However, it seems probable that the shortcomings of the more complex models in this study are aggravated by the very specific features of the utilized dataset. Hence, market participants should acknowledge that the obtained forecasting results are always not only largely dependent on the chosen methodology, but also on the utilized dataset.

Keywords
Real Estate, Housing Market, Forecasting, Simple, Complex
Additional information

CONTENTS

TABLES

FIGURES

1 INTRODUCTION

Real estate forecasting has become an indispensable tool for strategic decision making in the real estate sector. In addition to strategic asset allocation and portfolio management decisions, forecasting outcomes are also used in developers´ estimations of demand and construction costs when validating their business plans (see e.g. Brooks & Tsocalos, 2010, p. 2). The increased emphasis on forecasting is a natural evolvement as more economic data has become readily available. As a result, the forecasting of housing prices is also a rather popular area of research and there are numerous studies trying to predict price patterns of various housing markets. Despite advancements in modelling, limited success has been achieved in finding reliable and consistent models to predict movements in real estate markets (Tonelli et al., 2004).

Although there are difficulties in generating consistent forecasts, housing markets, and more broadly real estate markets, are generally seen as forecastable to a certain degree by researchers. One of the first papers highlighting the forecastability of housing prices was written by Case and Shiller (1989). They show that various information variables predict future housing prices. Following these initial studies, numerous other studies, including Muellbauer and Murphy (1997), Crawford and Fratantoni (2003) and Bork and Møller (2015) have suggested that housing prices are in fact forecastable.

As researchers build more and more complex models trying to more accurately capture housing market movements, there is simultaneously, however, some conflicting housing forecasting literature that suggests that simpler models could actually outperform or perform equally to their more complex counterparts. Brown, Song and McGillivray (1997), in one of the first studies comparing housing price forecasts, consider UK housing prices and find that a simpler univariate model outperforms other more complex multivariate models when it comes to the accuracy of the forecasts.

This phenomenon is also broadly observed in the field of property forecasting. Chaplin (1999) discovered that simple models produced more accurate forecasts of the UK property market rents, compared to the more complex econometric models, despite their lower historical fit. Similarly, Patrick, Okunev, Ellis, and David (2000) observed that simple exponential smoothing models were highly comparable, and generally outperformed, other more complex structures when forecasting property market movements in three different countries. Jadevicius and Huston (2015) provide a recent study concerning UK property market rent forecasting. They find that while more complex models, such as Vector Autoregression (VAR), had a significantly higher historical goodness-of-fit, their forecasting performance was comparable with much simpler models with a significantly lower historical fit.

In Finnish context, the earlier literature regarding real estate has primarily focused on understanding the market dynamics (see e.g. Kuismanen et al., 1999; Oikarinen, 2007) rather than comparing the performance of various forecasting methods. However, one particular study of Helsinki office rents by Karakozova (2004) evaluates the forecasting performance of three different models. In a similar manner to Chaplin (1999) and Jadevicius and Huston (2015), Karakozova finds that the market is indeed forecastable, and that the explanatory power of the model is not correlated with the actual forecasting capability of the model.

The forecasting performance of complex models compared to the simpler counterparts is largely unexplored in the Finnish housing market context, despite the significance of the Finnish housing industry. In Finland, the total value of the housing stock is about 320 billion euros which represents approximately 30% of the national wealth.[1] Consequently, as the housing stock constitutes a significant share of the overall wealth, housing is one of the most important sectors in the Finnish economy. Changes in the housing market can even be a signal of the evolvement of GDP, due to the "wealth-effect of housing" in addition to the influences that housing market has on financial and construction activity (see e.g. Belsky & Prakken, 2004; Case et al., 2005). Thus, given the significance of the housing market, it is highly important to designate which forecasting methods lead to the most reliable forecasting results.

Real estate markets are local, and as a result local price forecasting is more appropriate and accurate compared to, for instance, forecasting a country-wide index (Al Marwani,

[1] ROTI 2019 -report (from www.ril.fi) showcases that buildings form 45% of the national wealth of Finland. Approximately two-thirds of the wealth tied up in buildings is comprised of housing.

2014). Therefore, this study focuses on the housing market of the Helsinki metropolitan area (HMA), which includes cities of Helsinki, Espoo, Kauniainen and Vantaa. HMA is the fastest growing area of Finland, and although it only represents 0,4% of the land area of the whole country, it corresponds to 21% of Finland´s total population.[2]

This study intends to broaden the empirical research on housing market forecasting and build on the existing literature in two main ways. First of all, this study investigates whether it is possible to gain similar forecasting performance from simple models compared to more complex specifications in housing market context. Majority of the previous literature regarding this issue has concentrated on commercial and industrial markets. Secondly, this study contributes to the research on the Helsinki metropolitan area housing market and its forecastability.

In total, five different modelling and forecasting techniques are employed in this study, including Simple Regression (SR), Multiple Regression (MR), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average with a vector of an explanatory variables (ARIMAX) and Vector Autoregression (VAR). The model selection is based on a common classification of real estate forecasting models by Lizieri (2009). A set of macroeconomic variables, based on previous academic literature, is used with these different modelling techniques to generate *ex-post* (out-of-sample) forecasts for five different time periods. In total 25 forecasts are generated. To compare the accuracy of these forecasts, different metrics such as Theil´s U statistic and root-mean-square-error (RMSE) are used.

Given the heterogeneity of housing as an asset class, it is desirable to use quality-adjusted price indices in empirical studies. The use of non-quality-adjusted price indices could result in statistical inaccuracies, such as exaggerated short-term volatility (Oikarinen, 2007). To reduce the impact of heterogeneity, this study is conducted using quality-adjusted price indices based on privately financed multi-storey building

[2] Statistics Finland (www.stat.fi) reports the regional division of the Finnish population. As of spring 2020, the HMA area corresponds to 21% of the total population of Finland.

apartments sold in the secondary market in the HMA area. These quality-adjusted i.e. hedonic price indices for the HMA area are provided by Statistics Finland starting from the year 1988. Other variables used in this study, such as GDP, do not all have regional HMA data available, so national data is used.

The results of this study suggest that Simple Regression model (SR) is able to outperform other more complicated model structures, such as Vector Autoregression (VAR) and Autoregressive Integrated Moving Average with a vector of an explanatory variables (ARIMAX) specifications, despite its lower historical fit. Moreover, the results suggest that developments in the HMA housing market can be forecasted, but due to data scarcity it can be problematic to consistently form reliable forecasts. Thus, the empirical results of this study imply that simpler forecasting models seem to flourish in data scarcity, as more complex models run the risk of overfitting the available data. However, it seems probable that the shortcomings of the more complex models in this study are aggravated by the very specific features of the utilized dataset.

The rest of this study is structured in a following way. First, a background describing the special characteristics of Finnish housing markets is given. This is followed by a review of the relevant forecasting literature. Chapter 4 discusses the specific methodologies and datasets used in this study. Then, in Chapter 5, the results from the empirical analysis are presented. Finally, the conclusions are derived.

2 PRACTICAL BACKGROUND

In order to study the forecastability in the Helsinki metropolitan area (HMA) housing market context, it is important to understand how real estate markets work in Finland. Understanding the dynamics and structural changes of the Finnish housing markets is also helpful later in this study when choosing the appropriate variables and attributes for the forecasting models. This chapter lays out the practical context for this study by discussing some of the special features and the historical developments of Finnish housing markets, especially in the context of the HMA apartment market. Firstly, the current structure and the distinctive characteristics of the housing market in Finland are presented. Then, an overview of the historical development of the Finnish apartment market over the past decades is given.

2.1 Structure and characteristics of Finnish housing markets

The Finnish housing market is characterized by the substantial and growing role of its capital region. In the end of the year 2017, there were around 3 003 000 housing units in Finland, of which 2 680 000 were currently occupied according to data from Statistics Finland. From the year 1990 to 2017, approximately 29 000 housing units have been built per annum. The larger Helsinki region, which consists of Helsinki and 13 surrounding municipalities, formed 26% of the total occupied housing units in Finland, while the Helsinki metropolitan area (HMA), which consists of municipalities of Helsinki, Espoo, Vantaa and Kauniainen, formed 21% of the total occupied housing units. The total housing stock in Finland is approximately valued at 320 billion euros, and subsequently, the housing stock of HMA forms a significant part of this valuation. HMA also constitutes over one fifth of the population of Finland and an even larger share of the national GDP.

According to a study regarding the Finnish housing price dynamics by Oikarinen (2007), it is usual for governments all over the world to intervene in housing markets, due to the significant value of housing stocks and the general importance of housing for the economic development. This is also the case in Finland, where public policies play a prominent role in the housing markets by promoting affordable housing. Due to the prevalence of government intervention, the housing markets in Finland can be divided into two main sectors; privately financed sector and subsidized sector. The privately financed sector operates as a free market with no restrictions, and housing can be bought and sold at market prices. In the subsidized sector, prices and rents are publicly regulated and controlled. Privately financed sector constitutes approximately 85% of housing in Finland, while the subsidized sector forms approximately 15%.[3] However, in the case of the HMA, the subsidized sector forms a much larger part of the housing stock. This is due to the fact that a large part of the subsidized housing stock is concentrated in the major metropolitan regions of Finland. According to a study by Kajosaari (2016), by the end of the year 2015 the subsidized sector constituted approximately one fourth of the housing units and also one fourth of the new housing production in the city of Helsinki.

Since this study focuses on predicting the housing price index for HMA secondary market apartments, the subsidized sector is ignored for its price regulations. Regardless, it is still important to recognize that the policy changes regarding the subsidized sector can affect the housing prices also in the privately financed sector, especially in the major metropolitan regions. The exact influences of subsidized housing on house prices are unclear to a great extent, but according to empirical evidence from the Norwegian housing markets by Nordvik (2007), the subsidized housing production also increases the overall production and thus it is also likely to lead to slightly lower prices of the privately financed housing. In the case of Finland, housing units from the subsidized sector are released from regulations after a variable time period of 10 to 45 years, and after that they are considered to be a part of the privately financed sector.

When it comes to the composition of the housing stock, in the end of the year 2017, 64% of the dwellings in Finland were owner-occupied while 33% were rental dwellings, according to data by Statistics Finland. Rest of the dwellings are composed from so-called right-of-occupancy apartments and various other mixed forms of

[3] Statistics Finland (www.stat.fi) reports statistics regarding the composition of the housing stock in Finland. As of spring 2020, the privately financed sector constitutes approximately 85% of the total housing stock. However, this figure can fluctuate significantly from year to year as older housing units from the subsidized sector are released from regulations and then considered as a part of the privately financed sector.

tenure. According to Kivistö (2012), the owner-occupation rate for Finland is close to the Western Europe average. In Finland, rented housing is more common in major cities than in the whole country on average. This is also the case in HMA, where 44% of the dwellings are rented.

Regarding the ownership structure of the rental apartments, data from KTI (2019) implies that professional non-subsidized investors own 22% of the rental dwellings in Finland, while subsidized investors and municipalities own 42%. Private investors own 36% of the rental dwellings. There is no specific data available about the ownership structure of the owner-occupied apartments, but it is evident that most of them are privately financed and free from price restrictions (Oikarinen, 2007). Multi-storey apartment buildings compose over 45% of the total occupied housing units in Finland, while detached housing and row housing compose 39% and 14% respectively. In the case of HMA, as much as three fourths of the occupied housing units are in multi-storey apartment buildings[4]. According to calculations by Oikarinen (2007), the share of privately financed apartment dwellings, which is also the main focus of this study, is around 40% of the total housing stock in HMA.

2.2 Historical development of Finnish apartment markets

The development of Finnish apartment markets over the past decades has featured some major price fluctuations. Kivistö (2012) discusses these price fluctuations in great detail and argues that housing market developments are often expressions of the concurrent conditions of the overall economy. Finnish capital markets have gone through numerous notable structural changes over the past decades, and the effects from these structural changes have also sparked the developments in the housing markets. Due to constraints in data availability, the empirical part of this study focuses only on the time period from 1999 onwards. Earlier developments are still discussed to provide sufficient background to understand the structural changes that affected the Finnish housing markets. Figure 1 illustrates the aforementioned price volatility in the

[4] Statistics Finland (www.stat.fi) reports statistics regarding the composition of the housing stock in Finland. As of spring 2020, multi-storey buildings are the most common and fastest-growing form of housing in both HMA and Finland as a whole.

appreciation of privately financed apartments sold in the secondary market in Finland, from Q1/1988 to Q4/2018.

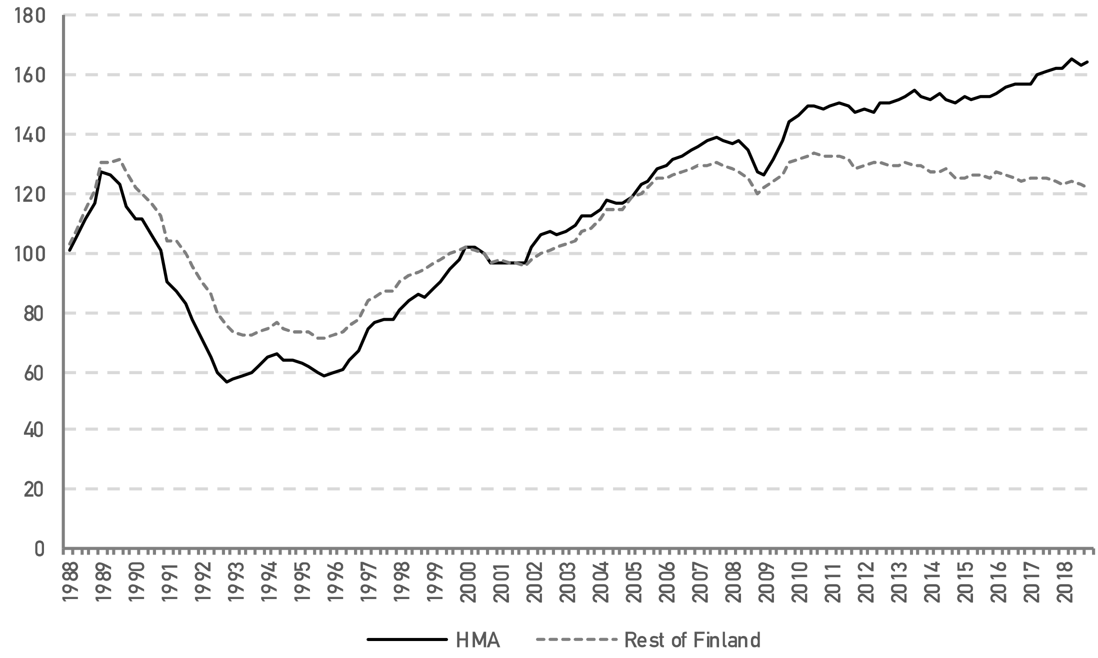

Figure 1. Apartment price indices, real, 2000 = 100, Q1/1988 – Q4/2018

In the 1970s, Finnish financial sectors, especially the real estate markets, were heavily regulated and hard to enter. Market entry and development activity were restricted among others by financial regulations and agreements between construction companies, financial institutions and government (Karakozova, 2004). According to Kivistö (2012), notable demographic changes in Finland led to increased demand for housing and into a construction boom in 1972-1974 of more than 70,000 housing units per annum at its peak. Subsequently, due to high demand, real housing prices peaked in 1973. After the oil crisis in 1973, a recession followed and it led to high inflation and long-lasting decline in real housing prices, although the nominal housing prices continued to rise.

In the 1980s, especially from 1987 onwards, housing prices in Finland increased dramatically and a housing bubble was formed. The bubble lasted from 1987 to 1989 and in this time period real housing prices increased approximately 60%. The bubble eventually burst, and that, together with other economic factors, lead to a depression period and a steep incline in real housing prices that lasted well into the early 1990s.

The main reason for the housing bubble is thought to be the gradual opening of capital markets and heavy deregulation that took place in Finland in the late 1980s (Koskela et al., 1997; Kivistö, 2012). In the mid 1980s, the Bank of Finland gradually liberalized the banking system, which improved the availability of mortgages for retail clients and also reduced the required down payments. The average lending rate constraints on mortgages were also discontinued and the amount of government rent control decreased. These structural changes led to a huge growth in household debt and eventually the housing bubble (see e.g. Huovari et al., 2005; Oikarinen, 2007). However, results from Oikarinen (2007) suggest that despite the deregulation that took place in the 1980s, there has not been a notable change in the relationship of real housing prices and fundamentals.

Following the housing bubble, it took several years until the real housing prices started to appreciate again. Eventually in 1996, real housing prices started to grow steadily. The growth trend was abrupted briefly by the deflation of the IT stock market bubble in early 2000s and the subprime mortgage crisis in 2008, but in both cases the downturn in prices lasted only for a year or so until the prices reached their previous highs. Moreover, it is evident that HMA prices have started to diverge from the prices of the rest of the Finland in the past decades. The increase in growth rate variation across regions has been apparent from the 1990s. Especially after the subprime mortgage crisis in 2008, HMA housing prices have appreciated rapidly while the housing price index for the rest of the Finland has entered a downward trend.

One explanation offered for the diverging prices is the main ongoing macro-trend that is affecting the dwelling preferences in Finland: urbanization. Finnish cities were originally built to the rural areas of Finland, to the proximity of local factories. However, according to a paper by Huovari, Pakkanen and Volk (2005), due to changes in the landscape of work, people in Finland tend to move to larger cities in search of jobs and education. This leads to rising demand for housing in areas like the HMA, while simultaneously the demand in rural areas of Finland decreases. Housing supply is constrained in the HMA due to the scarcity of land, and this, together with rapid population and income growth in the HMA region leads to the sharper price growth compared to the rest of the country (Oikarinen, 2007).

When coming back to the main topic of this study – namely the forecasting of the HMA housing market – one should keep these discussed characteristics and developments in mind, as they might affect the forecasting results discussed later.

3 THEORETICAL BACKGROUND

This chapter presents the key findings from relevant literature regarding the forecastability of real estate, and especially the forecastability of housing markets. Firstly, the concept of housing price forecasting is discussed and relevant literature regarding the forecastability of housing markets is presented. Secondly, different variables that are commonly utilized for forecasting purposes are presented. Depending on the data availability in the HMA region, these variables are later deployed in the empirical part of this study. Finally, the forecasting and modelling methods that are evaluated in the later parts of this study are presented. The extant empirical research, comparing the so-called simple forecasting methods to more complex ones, is also reviewed. The forecasting and modelling methods, which are presented in the final section, are chosen based on their prevalence in the earlier real estate forecasting literature.

3.1 Concept of house price forecasting

Housing, and more generally real estate, has many distinguishing features that differentiate it from other asset classes. These features include high transaction costs, relatively weak liquidity, large unit size, heterogeneity, lack of short-selling opportunities, informational problems to name a few. Due to these distinctive characteristics, the research on real estate has historically been much more limited in comparison to other asset classes, such as stocks and bonds. According to Karakozova (2004), it was only in the 1980s that investment analysis of real estate on a portfolio level became more widespread. That said, real estate has become a much more popular subject of studies in recent years.

Despite the apparent lack of research compared to other asset classes, real estate as an asset class and especially housing, has many features that make it desirable for forecasting. For example, as a result of the large unit size and high transaction costs, housing markets are often thin and have lower liquidity compared to markets of other assets. High transaction costs are also likely to delay the adjustment of housing prices to shocks in fundamentals, since due to the large costs, owner-occupants are not likely to quickly adjust their housing habits and real estate investors are not likely to quickly adjust their portfolios (see e.g. Edin & Englund, 1991). These factors, combined with sluggish housing supply, lead to increased inefficiencies in housing markets, in a sense that it often takes several quarters for prices to reflect new information (Oikarinen, 2007).

Fama (1991) defines economically sensible variation of the efficient market hypothesis as a situation where prices reflect all the available information to the point where the marginal benefits of acting on the information do not exceed the marginal costs. Even if housing markets were informationally efficient according to this definition, it seems evident that housing price movements are predictable considering that housing markets adjust to new information quite sluggishly (Oikarinen, 2007). Various studies, such as Linneman (1986); Devaney, Evans and Rayburn (1987); Case and Shiller (1989); Gyourko and Voith (1992); Gu (2002); and Schindler (2011), have since found evidence that housing prices (or returns) even exhibit positive autocorrelation. However, the extant literature is less focused on the fact whether this forecastability offers marginal benefits that exceed the marginal costs.

One of the earliest papers examining the efficiency of housing markets was published by Linneman (1986). In this paper Linneman studies the market efficiency of the Philadelphia residential market using hedonic risk-adjusted prices. Linneman finds evidence of serial correlation in the data but deduces that the predictability is inadequate for financial gain due to the high transaction costs associated with real estate. A study by Devaney et al. (1987) concerning the housing returns in the city of Memphis reaches similar conclusions using different methodologies.

However, in an influential study by Case and Shiller (1989) regarding the market efficiencies in four detached housing markets: Atlanta, Chicago, Dallas and San Francisco, they find evidence of substantial predictability which corresponds to trading profits. They build two corresponding weighted repeat sales indices for each city, and then regress the quarterly observations in one index on the one-year lagged data from the other index. They document substantial predictability with predictive R^2 varying between 0.11 and 0.48 and with the average trading profits varying between 1% and 3%. They also find that the forecasting results for individual homes are significantly less accurate when compared to the city-wide index.

Gyourko and Voith (1992), Gu (2002) and Schindler (2011) add to this research by Case and Shiller. Gyourko and Voith (1992) studied 56 different metropolitan statistical areas (MSA) and found consistent results with Case and Shiller (1989) and suggest that “prescient market timers might have been able to make money in selective markets”. Gu (2002) studied the autocorrelation of the Conventional Mortrage Home Price Index (CMHPI) and found that degree and the sign of autocorrelation varies significantly over time and location. Schindler (2011) studied 20 different national and metropolitan indices and found evidence of strong autocorrelation even at 24-month lags.

3.2 Variables used in house price forecasting

In addition to historical prices and returns, other variables, such as various macro-economic variables, have been found to be accurate predictors of housing price development (see e.g. Linneman, 1986; Case & Shiller,1989). A reason for this is the fact that housing market has wide and strong connections with the rest of the economy, and local housing market outcomes reflect local economic conditions. However, the relative strength of these connections may vary in different markets and change over time, which convolutes the forecasting process. Also, the use of different indices in different studies might yield large discrepancies in the findings regarding particular indicator values (see e,g, Gerardi et al., 2010).

The subsequent analysis of research on housing price determination suggests that several main factors, such as gross domestic product (GDP), interest rate, population growth, housing starts and completions, employment level, construction costs, vacancy rate and disposable income can be singled out as the most relevant and prevalently used variables in housing price modelling. However, when it comes to distinguishing the explanatory power or relative significance of parameters, there seems to be little agreement. Also, none of the reviewed studies contained all of the aforementioned variables.

a) GDP

A study by Case, Goetzmann and Rouwenhorst (2000) discusses cross country real estate cycles and concludes that there is evidence of international co-movements in real estate returns. Real estate is not portable and hence the competition across markets and countries should be low; therefore the co-movement across markets and countries should also be low. The study concludes that the co-movements are caused by the effects of GDP changes on real estate returns, since GDP changes are highly correlated between countries. Similarly, in a study regarding lending booms and real estate bubbles, Collyns and Senhadji (2002) find real GDP to be a significant fundamental in real estate modelling since it is an indicator for the aggregated level of income per capita and population size.

Jin and Zeng (2003) contribute to the literature by generating a general equilibrium model that illustrates the relationship between monetary business cycles and house prices in the US real estate market. In line with the findings from earlier literature, they are able to reproduce the fact that house prices and real GDP are positively correlated. Furthermore, they conclude that besides the strong correlation between GDP and house prices, monetary policy and nominal interest rates also play a special role in the determination of house prices.

In Finnish context, Karakozova (2004) examined the predictability of office returns in Helsinki area. Karakozova employed national GDP growth, since the provincial growth data for the Helsinki area was not available. She discovered that national GDP growth had a strong impact on property rents, and hence, national GDP was also useful when used in forecasting purposes.

b) Interest rate

Typically, interest rate variable tracks the movements in the mortgage rate over time. This is especially true in Finland, where most of the mortgages are tied to relatively frequently changing interest rates (Oikarinen, 2007). Rising interest rates might thus cause selling pressure, and also discourage potential buyers from acquisition, since the cost of having a mortgage and thus user costs rises concurrently (see e.g. Englund, 2011).

In addition to the previously mentioned study by Jin and Zeng (2003), number of other studies have found similar evidence that interest rates largely affect housing prices (see e.g. Hott and Monnin, 2008; Lind, 2009; McCord et al., 2011). For example, a study by Jacobsen and Naug (2004) examined the development of Norwegian housing market from year 1992 onwards using an econometric model that incorporated various explanatory variables, such as interest rate, housing construction, unemployment and household income. They found that housing prices react quickly and strongly to changes in interest rates, and that interest rate had the most explanatory power of all of the variables incorporated in their model. The authors conclude that the changes in interest rate directly affect the demand for credit, which instead directly affects the demand for housing.

c) Population growth

The use of population growth as an estimation variable for housing prices was first argued by Case and Shiller (1990). In their study, Case and Shiller follow up on their previous study from 1989 by performing strong-form efficiency tests exploring the forecastability of four different detached housing markets with a number of forecasting variables. They are able to show that economic predictors capture a significant part of real estate return fluctuations; their regressions have R^2 values ranging from 0.336 to 0.615. They use the change of adult population (ages from 22 to 44) as one of their forecasting variables and find that it is positively correlated with housing price changes. They also find similar evidence regarding the change in per capita real income.

Many studies have since found similar evidence (see e.g. Jud and Winkler, 2002). Borowiecki (2009) added to the literature by studying the housing markets in Switzerland over a 17-year long period using vector-autoregressive models and found that the Swiss housing prices were influenced most by population growth of the adult population (ages from 20 to 64). The study found that 1 percent change in adult population led to approximately 2% higher house price growth.

d) Housing starts and completions

Housing starts are argued to be a good indicator of housing price development, since increase in the number of available dwellings increases supply and puts downward pressure on housing prices. Rapidly increasing residential construction also often leads to an over-saturated housing market (Borowiecki, 2009).

A study by Rae and van den Noord (2006), regarding the forces driving the Irish housing market, found that the increased per capita housing stock had a significant negative impact on the price of secondary market housing. Furthermore, it is noteworthy to mention that the authors find that the negative impact is caused by the disequilibrium in the market for new houses spilling over to the secondary market. Moreover, similar evidence was found by Jacobsen and Naug (2004) as they argue that in the long-term an increase in housing construction will result in reduced housing prices, while a reduction in housing construction will result in increased prices. However, Jacobsen and Naug point out that in the short-term it is difficult to determine how housing starts affect house prices, since it takes a long time to build new dwellings, and even when considering housing completions the yearly stock increase is relatively small compared to the total housing stock.

e) Other variables: Employment level, disposable income, vacancy rate and construction costs

Housing prices are also generally bid up as a result of other variables, such as better local employment opportunities, higher disposable incomes enjoyed by the residents and higher occupancy rates. These variables indicate higher demand and higher construction costs, which in turn imply higher replacement costs. Adams and Füss (2010) conducted a cointegration analysis for macroeconomic determinants consisting of 15 countries using panel data for a time period of over 30 years. They examined the long-term equilibrium between housing prices and macroeconomic variables and found that variables such as employment level, money supply and industrial production had a significant elevating impact on local housing prices.

Furthermore, in his study regarding housing valuations and long-run equilibrium relations in the Helsinki metropolitan area, Oikarinen (2005) found that the three main factors determining the real housing price level in HMA in the long horizon are the level of disposable income, real lending rate and the construction cost index. The author argues that the higher construction costs lower the level of construction and thus the future supply. Interestingly, Oikarinen also finds that in addition to current disposable income, income expectations can likewise be used as an indicator for housing price developments.

The role of vacancy rate was studied by Laakso (2000) in the Finnish context. He used annual panel data of 85 Finnish regions covering a 15-year period and found that housing price developments are positively influenced by employment and income growth. Furthermore, Laakso discovered that higher vacancy rate negatively affected the house prices, as was expected based on the economic theories. In turn, Riddel (2004) long-term housing market equilibrium relationships in the US using data from 1967 to 1998 using a multiple error-correction model. The approach used by Riddel allowed her to separate supply-side disturbances from demand-side disturbances. She found that in the short-run construction costs were associated with higher housing prices. However, quite surprisingly, she did not find evidence of rising vacancy rates affecting housing prices negatively. Riddel remarks that the surprising result regarding the effect of vacancy rate might be a product of some misspecification in the model or simply a reflection of the complexity of housing price dynamics.

3.3 Forecasting methods

In this section, the literature regarding the classification of different forecasting methods is extensively reviewed to form an adequate context for the following comparisons. Later in this section, the exact forecasting methods that are compared in the empirical part of this study are presented and discussed in greater detail. Finally, the extant literature comparing simple and complex forecasting models is reviewed.

3.3.1 Classification of forecasting methods

As more and more economic data has become readily available, real estate forecasting has gained more traction as a noteworthy research subject. Subsequently, numerous mathematical models have been constructed as a way to understand the behavior and predict the developments in real estate markets.

Lizieri (2009) presented a commonly used classification of real estate forecasting models. According to Lizieri, forecasts can be either formal or intuitive. Formal forecasts fall in to two different categories: qualitative and quantitative. Qualitative methods apply qualitative data such as expert opinions and surveys. Quantitative techniques mainly rely on analyses derived from statistical data. As a result, quantitative methods tend to avoid personal biases and they are seen as more objective compared to qualitative methods, and hence, this study focuses solely on the quantitative methods. Quantitative methods can be classified in various ways, based on the object and approach being used, but according to studies by Makridakis, Wheelwrigth and Hyndman (1998, p. 8); Lizieri (2009); and Jadevicius (2014), the two main categories of quantitative methods are univariate and multivariate. Univariate models attempt to predict the future based solely upon the underlying patterns contained in the data, while multivariate models predict the future based on past and current values of other variables in the environment that are related to the variable being forecasted (Makridakis et al., 1998, p. 8). In the real estate forecasting literature univariate models are often referred to as time-series forecasting methods or extrapolative methods, while multivariate models are referred to as causal methods or explanatory methods. ARIMA and Exponential smoothing are considered as time-series methods, whereas Simple Regression, Multiple Regression, VAR, ARIMAX and Econometric Modelling are considered to be causal methods. The classification of these mentioned forecasting models is illustrated in Figure 2 below.

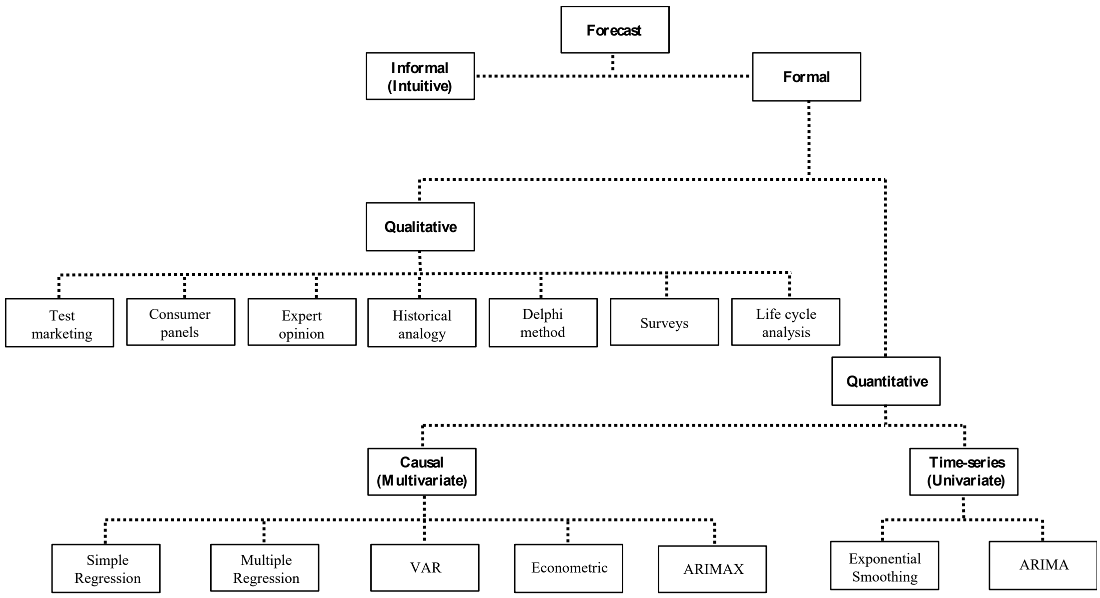

Figure 2. Classification of prominent real estate forecasting methods. Adapted from: Lizieri (2009) and Jadevicius (2014)

Researchers and analysts use different housing market modelling methods for different purposes. Generally, the choice of the forecasting model depends on the intended use of the model, i.e. whether the model is used for forecasting, or other applications such as testing of an economic theory or suitability analysis of theoretical frameworks. Shmueli (2010) defines predictive modelling, or in other words forecasting, as "the process of applying statistical model or data mining algorithm to data for the purpose of predicting new or future observations". Put another way, predictive models are often heavily relied on historical data to extrapolate the future and thus the causal relation between input and target variables is not often emphasized. In sum, predictive models sometimes sacrifice theoretical accuracy for empirical one. This does not mean that the interpretability of the model should be ignored, but it is less important when it comes to forecasting accuracy.

Explanatory models, in contrast, are defined as models that aim to explain the causality and relationship between the dependent variable and independent variables. Explanatory models also always require that the function is built supporting the estimated hypotheses, but the benefit gained from these models is that they allow for the opportunity to understand how different factors affect the housing market, and how markets respond to changes in key variables. Thus, explanatory models allow "storytelling" which rationalizes the forecasts and hence, explanatory models are often used for explaining the past rather than for predicting the future. (see e.g. Shmueli, 2010; Jadevicius, 2014)

This study focuses solely on predictive modelling and thus, even though causal methods are used, they are not evaluated for their theoretical accuracy but used strictly in a forecasting sense.

3.3.2 Forecasting methods evaluated in this study

This study compares Autoregressive Integrated Moving Average (ARIMA), Simple Regression (SR), Multiple Regression (MR), Vector Autoregression (VAR) and Autoregressive Integrated Moving Average with a vector of explanatory variables (ARIMAX) models. The use of these models is based on the classification of the most prevalent forecasting methods in real estate literature as seen on Figure 2. Exponential Smoothing and Econometric Modelling are omitted from the scope of this study.

a) Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA), also commonly referred to as Box and Jenkins model, is an atheoretical model, meaning that it is not based on any underlying economic theory. ARIMA models assume the future value of the forecasted variable to be a function of the last observations and white noise error term. In other words, ARIMA models simply seek to produce forecasts by capturing the empirically relevant features of observed data series. Hence, ARIMA models are of benefit when regression-based models are not available or difficult to use, e.g. in the case of data paucity (Brooks and Tsocalos, 2010, p. 225).

As Brooks and Tsocalos (2010, p. 241) present, ARIMA model is formed as a combination of autoregressive (AR) and moving average (MA) specifications. As such, ARIMA models are often denoted as ARIMA (p,d,q), where the autoregressive operator AR is of order p and the moving average MA operator is of order q with the data differenced d times. ARIMA models can only be used with stationary data, which means that with non-stationary data the data has to be differenced. The AR operator implies that the future values can be forecasted based on the past observations, while the MA operator involves random shocks and error terms into the model.

By changing the orders of the operators in ARIMA models, many different variations of the model can be created. However, this flexibility leads to the conundrum of choosing the most appropriate model specification. According to Makridakis et al. (1998, p. 16), researchers tackle this issue by using alternative “information criteria” techniques, in addition to subjective reasoning when selecting the most suitable specification. One of these techniques is the Akaike Information Criterion (AIC). AIC helps to select the most parsimonious model with the lowest AIC values indicating best model specification. Redundant variables in the model result in higher AIC value. One caveat regarding AIC is that AIC tends to select higher-order specifications compared to other information criteria (Brooks and Tsocalos, 2010, p. 249).

According to various studies (see e.g. Crawford & Fratantoni, 2003; Brooks and Tsocalos, 2010; Vishwakarma, 2013) ARIMA models, as well as other univariate time-series models, have been found to be useful for short-term forecasting scenarios. In line with this, a study by Stevenson (2007) investigating forecasting accuracy of ARIMA models finds that ARIMA models tend to over- or under-estimate crucial turning points in the longer-term. Nonetheless, ARIMA models are highly prevalent in real estate forecasting literature, partly due to their low demands for data. A study by Crawford and Fratantoni (2003) uses univariate models to forecast aggregate home price changes in separate parts of the US and found that despite the lower model fit of the ARIMA models, they still performed well in out-of-sample forecasting and in point forecasts. In addition to this, Stevenson (2007) finds that ARIMA models are extremely useful in forecasting in housing market booms when the market does not necessarily follow fundamentals, since ARIMA models capture the broad market trends. However, Stevenson warns that the forecasts obtained with ARIMA specifications might differ considerably from forecasts obtained with other models, which might convolute the forecasting process. Furthermore, studies by Clements and Hendry (1996) and Tse (1997) denote that forecasts from ARIMA models are adaptive enough to bear structural breaks.

b) Simple Regression (SR) and Multiple Regression (MR)

Simple Regression (SR) and Multiple Regression (MR) are multivariate models, which, in contrast to ARIMA and other univariate models, are often based on some relevant economic theory. Brooks and Tsocalos (2010, p. 73) define regression models, such as SR and MR, as models that try to explain how the changes in other variables affect the forecasted variable. By understanding these relationships between the variables, one can presumably forecast the future development of the chosen variable, *ceteris paribus.* Brooks and Tsocalos mention that regression models have been proven to be reliable methods of forecasting real estate developments over the medium- and long-term horizons, due to their capability to exploit causal relationships between various fundamental variables.

Brooks and Tsocalos (2010, p. 74) demonstrate SR as a regression of a dependant variable Y on a single explanatory variable X. Increase or decline in the explanatory variable will lead to an increase or decline in the dependant variable. However, as discussed previously in this study, housing markets are more often than not thought to be influenced by more than one exact variable. This is the main reasoning behind the MR method, where the regression results depicting the dynamics of the dependant variable Y are gained utilizing a set of interdependent explanatory variables (Brooks and Tsolacos, 2010, p. 108). This should, in theory, allow the obtainment of more accurate modelling results (see e.g. Jadevicius, 2014).

As Brooks and Tsocalos (2010, p. 194) and Al Marwani (2014) suggest, SR and MR are broadly applied in real estate literature to assess price changes. A study by Jadevicius (2014) even mentions that SR and MR specifications are the most used modelling approaches in real estate forecasting. Jadevicius also argues that regressions are especially common in medium-term forecasting, where, according to theoretical reasoning, the importance of understanding the economic connections between the variables becomes increasingly important. The high prevalence of regressions in empirical modelling can also be explained by their ease of use and uncomplicated interpreting process.

The difficulty with regression-based models, such as SR and MR, is the possibility of autocorrelation and heteroscedasticity. According to Al-Marwani (2014), one should always assess whether these issues provide disturbances to the modelling results. Put simply, the explanatory variables should not be correlated with each other, and the variance of errors should be constant across the period. However, these issues can often be detected and assessed in various ways when building the model, for instance with the Durbin-Watson test and White´s test (see e.g. Brooks and Tsocalos, 2010, p. 149).

c) Vector Autoregression (VAR)

Vector Autoregression (VAR) is a modelling approach which captures linear interdependencies among multiple variables. VAR models are often categorized as hybrids between univariate time-series models and economic models, in the sense that they do often incorporate economic data from multiple variables, but similarly to univariate time-series models they do not require much theoretical knowledge about the relationships between the utilized variables (Koop, 2006; Brooks and Tsocalos, 2010, p. 337). Brooks and Tsocalos (2010) even mention that VAR models are atheoretical by nature.

Furthermore, according to Brooks and Tsocalos (2010, p. 352), one advantage of VAR modelling is that all variables are treated as endogenous. This means that in contrast to previously discussed models, VAR models do not only exploit the effect of the forecasted variable on itself and the effects of the utilized variables on the forecasted variable, but also the effect of the forecasted variable on the utilized variables. Hence, this should allow the model to capture more features of the data rather than just past observations and errors of the series (Brooks and Tsocalos, 2010, p. 352).

VAR models are based on the same fundamental idea as univariate autoregressive (AR) models, but VAR models allow more than one evolving variable. The conventional VAR specification is a system where each of the variables in the system depend on lagged and current values of the other variables and error terms. Moreover, this leads to the issue of deciding the length of the lags in the specification. According to Brooks and Tsocalos (2010, p. 340), this issue is often decided using the Akaike information criterion (AIC) or the Akaike multivariate information criterion (MAIC). Other common issues related with VAR modelling include stationarity and choosing the level of parameterization (Brooks and Tsocalos, 2010). Also, in line with other regression-based models such as Simple Regression and Multiple Regression, VAR is also susceptible to autocorrelation and heteroscedasticity.

Brooks and Tsocalos (2010, p. 362) discover that VAR models are often used in forecasting scenarios where the underlying theory suggests that causal relationships are bi-directional or multi-directional. This has led to wide adaption of VAR models among economists (see e.g. Cogley & Sargent, 2005). When describing the benefits of VAR modelling, Brooks and Tsocalos (2010) refer to studies by Sims (1972) and McNees (1986). These studies conclude that VAR models provide a flexible and powerful theory-free method for data description and forecasting, which is often seen as superior to traditional structural models.

d) ARIMAX

Autoregressive Integrated Moving Average model with Exogenous Explanatory Variable(s), often referred to as ARIMAX, is a widely used variation of the ARIMA specification. ARIMAX is built in a similar way to its ARIMA counterpart, but in addition to Autoregressive (AR) and Moving Average (MA) components ARIMAX incorporates a Vector of Explanatory Variable(s) (X). In an equivalent manner to ARIMA modelling, the application of ARIMAX requires that the utilized data series is stationary (Makridakis et al., 1998, p. 9).

In a study regarding the comparison of different real estate forecasting methods, Karakozova (2004) mentions that the added component of relevant explanatory variables enables the creation of forecasts with a greater accuracy. Additionally, she discovers that when compared to other forecasting methods, ARIMAX model is superior in predicting turning points in Helsinki property market, due to its capability to pick up the scale of shocks and resultant persistence effects present in the data. Karakozova concludes, that since ARIMAX does not incorporate much of long-run information, it is able to pick up the scale of shocks and resultant persistence effects and hence performs better in the detection of market irregularities.

3.3.3 Comparisons between simple and complex models

Since real estate researchers and analysts have a plethora of different forecasting techniques at their disposal, they are often faced with the dilemma of choosing the right one. This issue is exacerbated in the almost inevitable case where different techniques give entirely different results about the object being modelled. Hence, the general difficulty of choosing and committing to a certain forecasting method has sparked research comparing forecasting methods between each other. However, as Tonelli (2004) puts it in his study regarding the forecasting of the Brisbane commercial real estate market: "limited success has been achieved in finding a reliable and consistent model".

One of the more fundamental aspects that divides opinions between researchers is whether adding variables and statistical complexity to the model actually improves the forecasting accuracy. Case in point, some studies have proposed the question of whether simple models perform equally or better than the complex ones. To discuss this topic at hand, we must first determine how simple and complex models are defined and how they differ from each other.

Unsurprisingly, there are no exact definitions or methods to distinguish simple and complex models from each other (see e.g. Buede, 2009). What is more, simple and complex models have very rarely been directly compared in various fields of science, let alone the field of real estate forecasting (Jadevicius, 2014). However, there are a few broad definitions that illustrate typical characteristics of more complex specifications. In a study regarding the predictability of UK office rents, Chaplin (1999) defines complex models as structures that require a large amount estimations and numerous variables. Furthermore, in their study regarding the ways of validating large scale models of urban development, Batty and Torrens (2001) define complex models as entities that are coherent in a recognizable way but whose elements, interactions and dynamics generate structures and relationships that cannot be defined *a priori*. To synthesize these definitions, complex models are characterized by a higher number of explanatory variables, a greater amount of required data and a more complex set of relationships. On the contrary, simpler models are characterized by uncomplicated structures, fewer variables and smaller amounts of required data.

Simple models are considered as a more traditional approach to modelling, while complex models have become a more common feat as the modelling technology has progressed over the past decades (Jadevicius, 2014).

In a study by Jadevicius, Sloan and Brown (2013), regarding the forecasting performance of various models in an UK property market context, the authors use similar definitions to categorize simple and complex models. Using their classification criteria on the models utilized in this study, Simple Regression, Multiple Regression and ARIMA models fall into the category of simple forecasting techniques, while VAR is categorized as a more complex structure. According to the authors, the categorization of the ARIMAX model can be ambiguous as it is often categorized as a simple forecasting technique, even though some papers (see e.g. Durka & Pastorekova, 2012) describe it as a complex specification.

When it comes to the empirical evidence conducted on the relative performance of simple and complex models, the literature seems to be somewhat equivocal. Although the performance of a specific forecasting technique always depends on the occasion and the selected samples and horizons, a vast amount of the available literature advocates for simpler modelling techniques. The alleged advantages of simpler models were first presented outside of the real estate literature in the broad field of econometrics. As an example, a publication by Makridakis et al. (1998) present the history of developments in modern forecasting and discusses ways of improving the usefulness and accuracy of the commonly used forecasting methods. The authors mention that due to the inherent flaws and errors that exist in statistical forecasting, there is always a sense of uncertainty and that forecasting errors cannot be entirely eliminated by more complex models or more gifted forecasters. Moreover, Kennedy (2002) discusses the same topic within econometrics and remarks that one should begin the analysis with simpler models as they are much less prone to errors and inconsistencies since sources of model failures are easier to detect. Furthermore, Kennedy adds that simpler models also require less know-how and that their results are easier to interpret, which leads to lesser amount of significant oversights in the analysis phase. The author concludes that one should opt for simpler methods over more complex ones if the assumptions and strengths of a simpler method are reasonable for the presented research problem. A more recent study by Orrell and

McSharry (2009) regarding the so-called “pitfalls” of forecasting models finds similar evidence. The authors explain that as models become more complex the number of elements they contain increases exponentially. A direct result of this is the flexibility observed in the more complex models, which also coevally leads to higher historical fits. The authors do not however find any implications that higher historical fits can lead to actual improvements in the forecasting accuracy, and deem that “while increasing the complexity of a model naturally gives more freedom to provide a better fit to the data, a model with too many parameters will not distinguish between the generative dynamics that we wish to extract”.

To continue, studies in the real estate discipline have likewise reached similar conclusions. Series of academic papers that were first to create a coherent argument for less complicated forecasting methods in the real estate discipline were published by Chaplin (1998; 1999). In his first study, Chaplin (1998) studied the predictability of multiple UK office rent indices and found that naïve forecasting methods outperformed other models with higher historical fits. This led him to conclude that choosing the forecasting method based on historical fit might hinder the predictive ability. In his second study, Chaplin (1999) examined the forecastability of the nation-wide Hillier Parker real office rent index in the UK for one-year periods using data from 1985 to 1994. Chaplin found that at least in the UK, property market researchers and analysts more often than not base their selection of the appropriate modelling method on the historical explanatory power of the model. This is despite the fact that in addition to his earlier study, Chaplin (1999) finds further evidence that high explanatory power bears no relationship to the actual forecasting performance. Chaplin concludes his second study by mentioning that although there is a temptation to search for more accurate and sophisticated modelling methods, complex methods are often inclined to mistake noise for information. These findings therefore indicate that complex models tend to fit historic data with greater accuracy as they contain more variables and are able to extract more information from the data, but the higher historical fit does not always translate into superior forecasting results.

Furthermore, in the context of real estate markets, numerous other studies have since found supporting empirical evidence (see e.g. Patrick et al., 2000; Vishwakarma., 2013; Jadevicius and Huston; 2014). However, to the best of my knowledge, only a handful of direct comparisons between forecasting methods have focused on housing market forecasts. In one of the rare papers explicitly comparing forecasting methods in the housing market context, Crawford and Fratantoni (2003) compare regime-switching models with a simpler ARIMA model using state-level repeat transactions data for California, Florida, Massachusetts, Ohio and Texas. The authors find that while the more complicated regime-switching models are very useful for characterizing historical house price patterns due to their superior in-sample fit, simpler models might be more useful for out-of-sample point forecasts. The authors point out that more complicated models run the risk of overfitting the data especially when the sample is relatively small.

However, some studies have reached conflicting results. In the real estate discipline, opposing evidence is presented by Stevenson and McGrath (2003), who compare four alternative forecasting models in the London office market setting. They forecast the CB Hillier Parker London Office index for a 3-year out-of-sample period with semi-annual data from 1977 to 1996, using an ARIMA model, a Bayesian VAR (BVAR) model, an OLS based single equation model and a simultaneous equation model. The results reveal that the ARIMA model was by far the worst performing model, while the Bayesian VAR model was by far the best performing model. The authors accredit the dissatisfactory performance of ARIMA model to the large coincidental upswing in the index. The forecasts generated with ARIMA models are entirely based on historic data, and thus ARIMA model is not able to pick up this upswing. Contrary to this, BVAR model is capable of adjusting to this turning point the market and thus provides adequate forecasting performance.

To summarize, the literature is still unclear on which modelling methods are generally advisable for forecasting purposes. Some researchers advocate for more complex methods as they are able to pick up more information from the data, while others advocate for simpler methods as they are less likely to mistake noise for data. However, the literature is quite unanimous on the fact that the models of higher complexity run a higher risk of inexpert use and are a lot less user-friendly.

4 DATA DESCRIPTION & METHODOLOGY

This chapter presents the methodological techniques and datasets used in the empirical part of this study. Firstly, the main data sources used to gather the data for this study are discussed, as well as the limitations in the data availability and their ramifications on the data selection. Secondly, the statistical tests used to avert possible biases in the samples are presented. Finally, the chapter is concluded by discussing the specific methodological aspects which are used to form and evaluate different forecasts.

4.1 Data and its acquisition

As it often is the case with housing, the available data is far from being optimal for empirical analysis. Nonetheless, a considerable effort has been put into the data acquisition to assemble as comprehensive and reliable dataset as possible. All of the data utilized in this study is gathered from public sources so that the results can be beneficially applied in industry or future research. In the following, the data acquisition is explained in greater detail.

The data samples used in this study are comprised of real values rather than nominal values. Moreover, all of the samples are either acquired as or transformed to semi-annual frequency. The use of less frequent data could lead to data paucity, and the use of more frequent data could lead to various statistical inaccuracies, especially in the HMA setting where the housing market could be described as “thin” (Oikarinen, 2007). Further, all of the data samples are transformed into half-over-half growth rates to make the samples approximately stationary, as this allows the accommodation of the various forecasting methods.

4.1.1 Dependent variable

This study is conducted using the price index growth for privately financed multi-storey building apartments sold in the secondary market in the HMA area as the dependent variable. The index is provided by Statistics Finland from 1988 onwards as a quality-adjusted index, meaning that it is standardized for composition changes regarding the location, the type of building and the number of rooms. Although the index is not quality-adjusted in regard to micro-location, floor area, year of completion and so on, it should still considerably reduce the heterogeneity in the data.

To further reduce the heterogeneity in the dependent variable, this study uses semi-annual data from the relatively homogenous area of HMA. In addition, the HMA has a relatively high amount of semi-annual transactions which in turn should result in a smoother index. The utilized index also disregards new construction and more heterogenous forms of housing (such as terraced housing and detached housing) outright and instead solely focuses on secondary market apartments multi-storey buildings, which should again reduce the heterogeneity in the data. The increased heterogeneity should result in more reliable and accurate empirical results. Due to constraints in the explanatory variables, a sample from 1999 H2 to 2018 H2 is utilized in the empirical part of this study. As mentioned in Section 2.2, there has not been any notable changes in the relationship of real housing prices and fundamentals in the HMA over the past decades, and as such, this sample should be free of notable structural breaks. Figure 3 showcases the graphical illustration of the dependent variable. The figure illustrates how the sample from 1999 H2 to 2018 H2 is highly volatile.

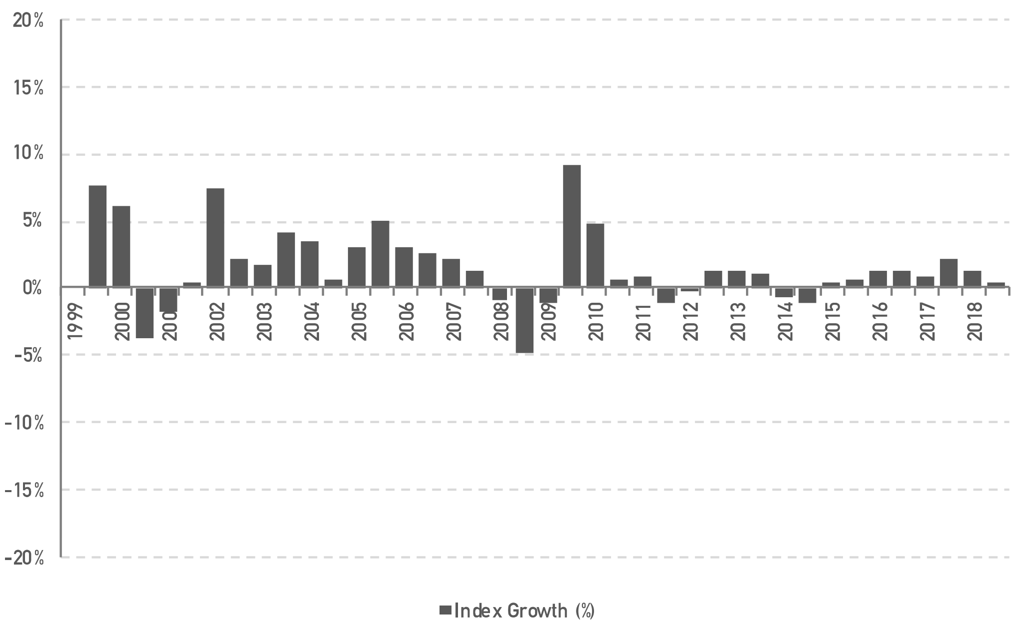

Figure 3. Price index for apartments in the HMA area growth (%, semi-annual), H2/1999 – H2/2018

4.1.2 Explanatory variables

As the literature review of this study highlights, GDP, interest rate, population growth, housing starts, housing completions, employment level, disposable income, vacancy rate and construction costs are all indicators frequently used in housing price forecasting. Unfortunately, data on all of these variables was not sufficiently available for the designated purposes, and as such, this study employs GDP, Interest rate, New contracts, Population growth, Housing starts, Housing completions and Disposable income as explanatory variables. However, of these variables, only Population growth has regional data available for the HMA. Due to this, national data is used for the rest of the variables.

Ideally, we would like to use local data that accurately portrays the specific characteristics of the particular defined area. However, it is not unusual that regional economic data is not obtainable, and as a result, it is not unusual to use nationwide variables instead. In fact, the use national data in the case of regional data paucity is advocated by several studies. As mentioned in the literature review, Karakozova (2004) discovered that national data was useful when forecasting the office returns in the Helsinki area. In addition to this, Hekman (1985) argues that national data can be used to model regional economic forces, as long as the national and the regional data are moving into the same direction. It seems sensible to assume that this is the case with the HMA, as the HMA is the main economic center of Finland (Oikarinen, 2007). Regardless, it seems reasonable to keep in mind the lack of local data when evaluating the empirical results.

The data for the national gross domestic product of Finland is obtained from Statistics Finland, who provide seasonally and per working day adjusted semi-annual data for the GDP starting from 1990. Similarly, the seasonally and per working day adjusted data for the national disposable income of households in Finland is also obtained from Statistics Finland. Semi-annual data for the disposable income is available from year 1999 onwards. When it comes to the Interest rate variable, this study utilizes the average interest rate on new mortgages, as it is comprehensively available in Finland. The average interest rate on new mortgages was selected over the 12-month Euribor rate, as the average mortgage rate also encompasses the profit margins that lenders impose on top of the reference rates. Furthermore, data on the volume of new mortgage contracts is also used as an explanatory variable to gain an extensive view of the Finnish mortgage market. The semi-annual data on the average rate on new mortgages, as well as the volume of new mortgage contracts, was acquired from Bank of Finland, starting from year 1990. What is more, this study utilizes the semi-annual data for the population growth of the HMA as one of the explanatory variables. The data on the population growth is provided by the Helsinki Region Trends starting from 1991. Finally, the dataset is concluded with Housing starts and Housing completions variables. Housing starts variable is gathered as the number of apartments contained in the granted building permits for multi-storey residential buildings in Finland. Similarly, Housing completions variable is gathered as the number of completed dwellings in multi-storey buildings in a set time span. Statistics Finland provides monthly data of granted building permits and completed dwellings starting from 1995, which is then transformed to semi-annual frequency. On the whole, semi-annual growth data for the explanatory variables is collectively available from 1999 H2 onwards. The graphical illustrations of the data are presented in Figures 4, 5 and 6. The figures exhibit the developments in the explanatory variables from 1999 H2 to 2018 H2.

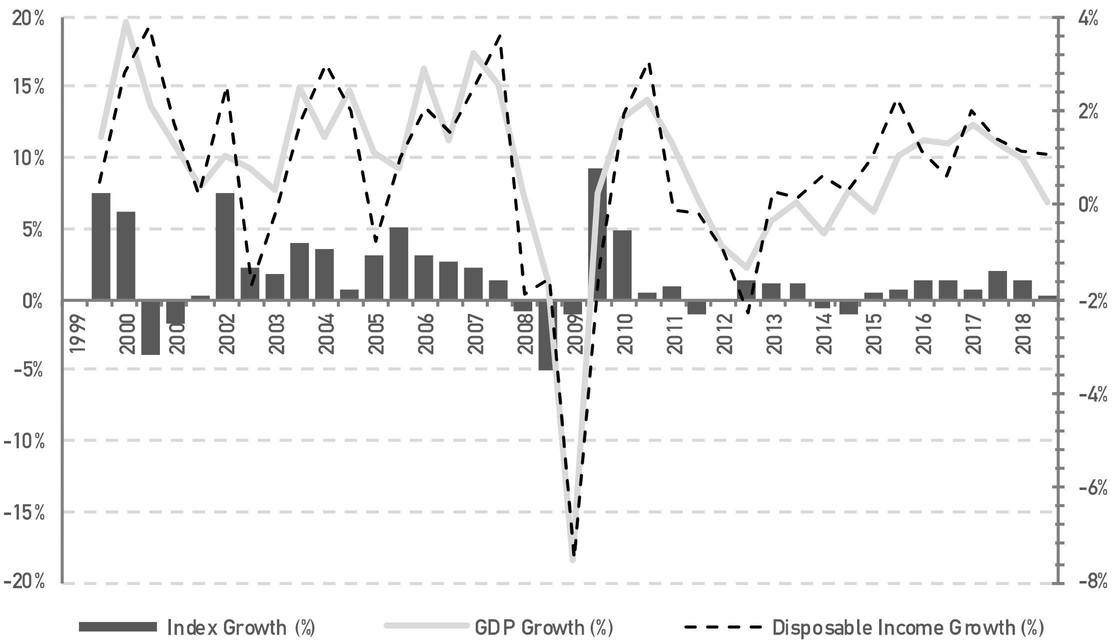

Figure 4. GDP and disposable income growth (%, semi-annual), H2/1999 – H2/2018

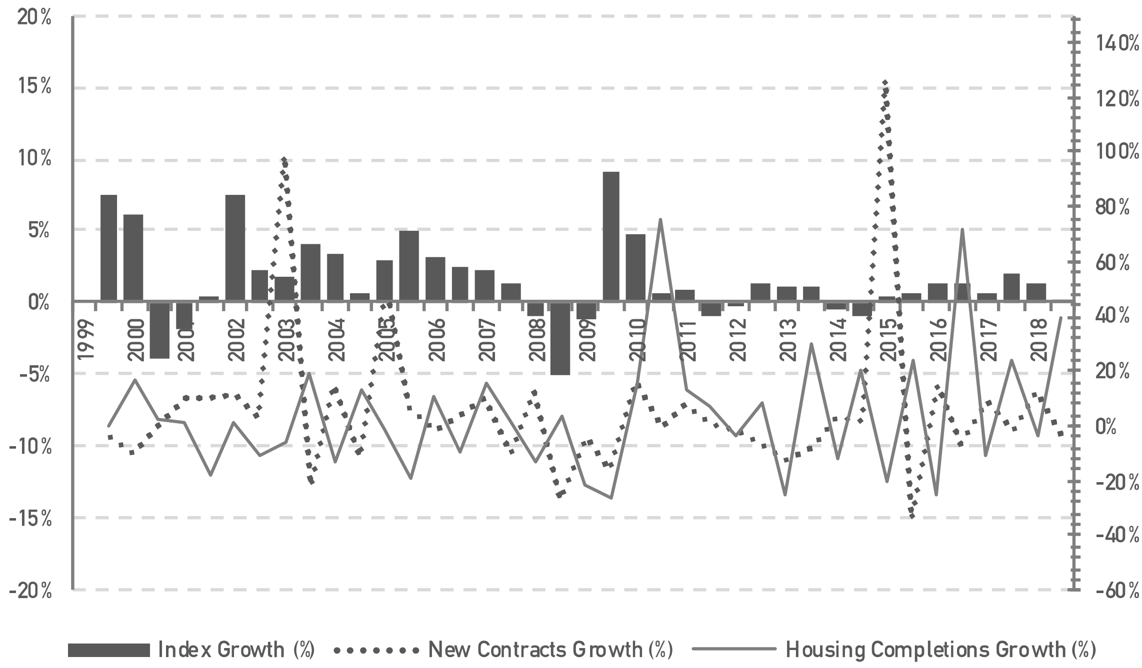

Figure 5. New contracts and housing completions growth, (%, semi-annual), H2/1999 – H2/2018

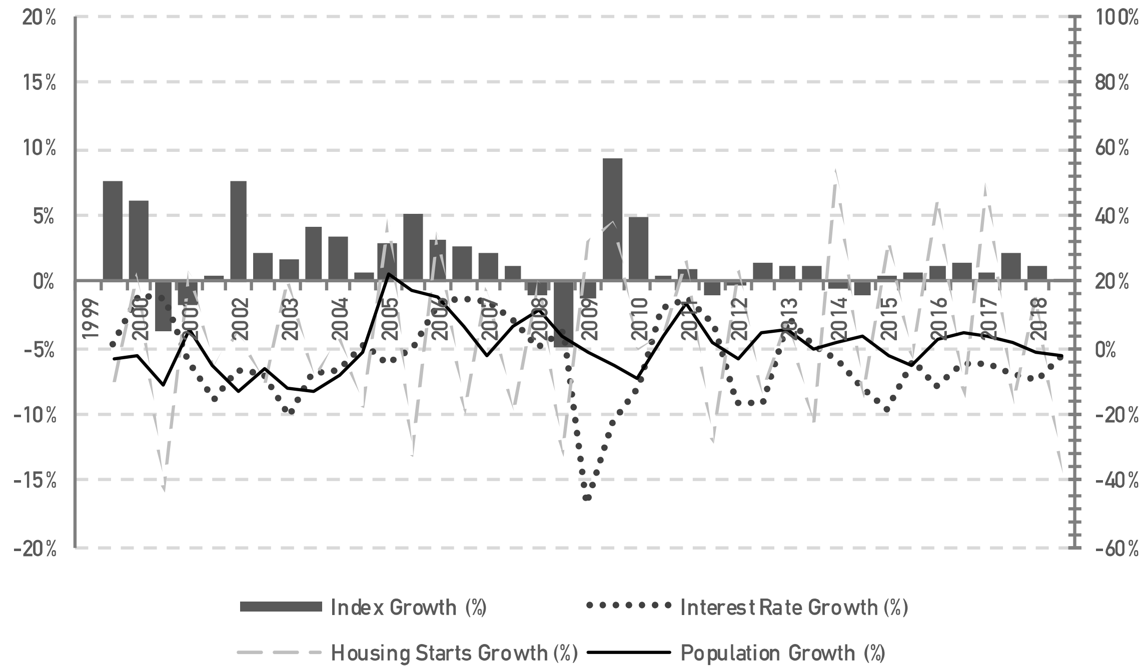

Figure 6. Interest rate, housing starts and population growth (%, semi-annual), H2/1999 – Q2/2018

4.1.3 Further steps in the data selection

Appropriately, the data used in this study ranges from 1999 H2 to 2018 H2 and covers a period of almost 20 years, which in turn translates to 39 observations in each of the samples. Later in the forecasting process, the samples are divided into *ex-ante* and *ex-post* periods. The *ex-ante* period works as an initialization period where the forecasting models are formed and configurated, and the *ex-post* period works as a hold-out period against which the forecasting models are tested. This study considers an *ex-ante* period from 1999 H2 to 2016 H1 and an *ex-post* period from 2016 H2 to 2018 H2. The *ex-ante* period in this study covers 34 observations, which is considered to be substantial for both univariate and multivariate time-series modelling, as 20 observations is often considered as a minimum amount (see e.g. Mouzakis and Richards, 2007; Jadevicius, 2014). The summary statistics for the variables utilized in this study and their data sources are presented in table 1. All of the samples are of same length, continuous and of same frequency.

Table 1. Summary of selected variables

		Descriptive statistics			
Variable name	Source	Start	Mean	SD	*n*
Price index (%, HOH)	Statistics Finland	1999 H2	1.615	2.879	39
GDP (%, HOH)	Statistics Finland	1999 H2	0.808	1.818	39
IR (%, HOH)	Bank of Finland	1999 H2	-3.274	12.971	39
NC (%, HOH)	Bank of Finland	1999 H2	6.332	28.942	39
PG (%, HOH)	Helsinki Region Trends	1999 H2	0.744	8.191	39
HS (%, HOH)	Statistics Finland	1999 H2	3.998	26.250	39
HC (%, HOH)	Statistics Finland	1999 H2	4.612	23.013	39
DI (%, HOH)	Statistics Finland	1999 H2	0.708	2.043	39

GDP = Gross domestic product, IR = Interest rate, NC = New contracts, PG = Population growth, HS = Housing starts, HC = Housing completions, DI = Disposable income, *n* = number of observations, HOH = half-over-half growth, SD = standard deviation.

The data samples are then tested for unit-root and stationarity using three statistical tests; Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) test for unit-root and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for stationarity. The use of multiple assessment methods is advised by Brooks and Tsocalos (2010, p. 382), as using only a single test might lead to significant oversights.

For unit-root tests, such as ADF and PP, the null hypothesis is that the series possesses a unit-root and hence is not stationary. On the contrary, for stationarity tests, such as KPSS, the null hypothesis is that series is stationary. Table 2 reports results from unit-root and stationarity tests for all of the variables. All of the data samples are confirmed to be stationary at a 5 percent significance level. Hence, the particular data samples are later used in the model parameterization.

Table 2. Unit-root and stationarity test results for selected variables

	Test results		
Variable name	ADF	PP	KPSS
Price index (%, HOH)	-6.293	-4.651	0.047
GDP (%, HOH)	-3.780	-3.801	0.086
IR (%, HOH)	-4.047	-3.774	0.041
NC (%, HOH)	-4.348	-8.102	0.077
PG (%, HOH)	-3.427	-3.248	0.104
HS (%, HOH)	-3.984	-12.393	0.075
HC (%, HOH)	-3.636	-7.727	0.058
DI (%, HOH)	-3.823	-3.988	0.112

This table showcases the results for the Augmented Dickey-Fuller test (ADF), the Phillips-Perron (PP) test and the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS). The null hypothesis of the ADF test and the PP is that unit root is present, while the null hypothesis of KPSS test is that series is stationary. The critical values at a 5 percent significance level for the applied sample sizes are as follows: Augmented Dickey-Fuller: -2.89; Phillips-Perron: -2.89; Kwiatkowski-Phillips-Schmidt-Shin: 0.146.

Since ADF, PP and KPSS tests are tailored for detecting nonstationarity in the form of a unit root in the process, they do not necessarily detect nonstationarity of the seasonal kind. Thus, appropriate steps are taken to ensure that the utilized data is also non-seasonal. The data for GDP and DI variables is obtained as seasonally and per working day adjusted. IR, NC and PG variables, as well as the dependent variable, do not show any evidence of seasonality. However, HS and HC variables exhibit seasonal variation, and thus seasonal dummy variables are accordingly used to eliminate potential complications.

4.2 Methodology

This section starts with describing in detail the five models used to compute the forecasts and then continues on to explain the exact steps used in the forecasting performance evaluation. The exact formulations introduced in this section are used to carry out the empirical assessment presented in the following chapter.

4.2.1 Model formulation and implementation

As it was noted earlier in section 3.2.2, this study utilizes and compares Autoregressive Integrated Moving Average (ARIMA), Simple Regression (SR), Multiple Regression (MR), Vector Autoregression (VAR) and Autoregressive Integrated Moving Average with a vector of explanatory variables (ARIMAX) models. This section presents in detail the exact implementation of these models in this study.

a) ARIMA model

ARIMA model is formed as a combination of autoregressive (AR) and moving average (MA) specifications. The combined model specification is expressed as follows:

$$Y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + b_1 u_{t-1} + b_2 u_{t-2} + \cdots + b_q u_{t-q} + u_t, \tag{1}$$

where Y_t is the dependent variable, μ is constant term, ϕ_p is pth order autoregressive parameter, Y_{t-p}, is past values of the dependent variable, b_q is qth order moving average parameter and u_t is the error term at time *t* (Brooks and Tsocalos, 2010, p. 241).

b) SR model

As Brooks and Tsocalos (2010, p. 74) explain, SR is a regression of a dependant variable Y on a single explanatory variable X. The exact form of the SR specification used in this research is:

$$Y_t = \alpha + \beta x_t + e_t, \tag{2}$$

where Y_t is the dependent variable, x_t is an explanatory variable at time *t*, and e_t is an error term at time *t*.

c) MR model

MR model is similar to SR model, but with *n* number of regressors:

$$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{nt} + e_t, \tag{3}$$

where Y_t is the dependent variable, $x_{1t}, x_{2t}, \dots, x_{kt}$ are the explanatory variables at time t, and e_t is an error term at time *t* (Brooks and Tsocalos, 2010, p. 108).

d) VAR model

The formulation of VAR model in this study can be illustrated with a simplified case of two interdependent variables x_t and y_t:

$$y_t = \delta_1 t + \phi_{11} y_{t-1} + \cdots + \phi_{1p} y_{t-p} + \beta_{11} x_{t-1} + \cdots + \beta_{1p} x_{t-p} + e_{1t} \tag{4}$$

$$x_t = \delta_1 t + \phi_{11} x_{t-1} + \cdots + \phi_{1p} x_{t-p} + \beta_{11} y_{t-1} + \cdots + \beta_{1p} y_{t-p} + e_{1t} \tag{5}$$

These simultaneous equations form a model VAR (p), which has two variables and p lags of each of the variables (Koop, 2006; Jadevicius, 2014). This study utilizes an equivalent specification, only with more variables.

e) ARIMAX model

The ARIMAX model is formulated similarly to ARIMA model, but with an additional vector of explanatory variables:

$$\begin{aligned} Y_t = {} & \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + b_1 u_{t-1} + b_2 u_{t-2} + \cdots \\ & + b_q u_{t-q} + \sum_{i=0}^{n} y_i X_{t-1} + u_t, \end{aligned} \tag{6}$$

where Y_t is the dependent variable, μ is constant term, ϕ_p is pth order autoregressive parameter, Y_{t-p}, is past values of the dependent variable, b_q is qth order moving average parameter, X_t is a vector of explanatory variables and u_t is the error term at time t (Jadevicius, 2014).

4.2.2 Performance evaluation

The utilized forecasting performance evaluation process has three major steps. Firstly, the forecasting models are parameterized within the *ex-ante* period. The *ex-ante* period spans from 1999 H2 to 2016 H1. The in-sample accuracy and the forecasting prowess of the assorted models is determined within this time period using traditional model accuracy methods, such as R^2 and Akaike information criterion. Secondly, the forecasting models are assessed for possible biases, such as autocorrelation and heteroscedasticity. Possible autocorrelation is examined with Durbin-Watson test and possible heteroscedasticity with White´s test. Thirdly, forecasts are created for the *ex-post* period for one, two, three, four and five steps ahead, i.e. from 2016 H2 to 2018 H2. This allows the assessment of forecasting accuracy in both short- and long-run scenarios. The *ex-post* accuracy assessment is done by calculating Theil´s U and root-mean-square error (RMSE) values for each of the forecasts. The details of the selected statistical tests are presented below:

a) Akaike information criterion (AIC)

Akaike information criterion (AIC) is a commonly used tool that estimates out-of-sample forecasting error and thus provides means for the selection of the most suitable model (Makridakis et al., 1998, p. 16). As Stevenson and McGarth (2003) and Karakozova (2004) explain, AIC is also useful in identifying the correct order of ARIMA(X) models, as a visual analysis of ACF and PACF is unlikely to produce as accurate results. As Akaike (1974) presents, AIC is calculated as follows:

$$AIC = n * \ln(\hat{\theta}) + 2 * k, \tag{7}$$

where n is the length of the time-series, $\hat{\theta}$ is the maximum likelihood estimate and k is the number of independently adjusted parameters within the model.

Sugiura (1978) and Hurvich and Tsai (1989) both argue that there is a substantial probability that AIC overfits small samples. Thus, they conceived AICc, which is more a suitable alternative when the sample sizes are small. This alternative criterion is also used in this study:

$$AIC_C = AIC + (2k^2 + 2k)/(n - k - 1), \tag{8}$$

where, correspondingly to AIC, n is the length of the time-series, k is the number of independently adjusted parameters within the model.

However, since both AIC and AICc are based on the maximum likelihood estimate $\hat{\theta}$, the criteria will decrease with smaller sample sizes, as the likelihood estimate subsequently increases. This means, that in this study the AICc results are not entirely comparable between different sample sizes, i.e. between lagged and non-lagged models, as the criteria will favor lagged models as they are based on smaller sample sizes. This same bias would persist with any information criteria, and thus, AICc is still used in this study. The bias could possibly lead to large inconsistencies in the model selection, and consequently, it has to be considered when evaluating the empirical results.

b) Durbin-Watson test

As it was mentioned in Section 3.3.2, autocorrelation and heteroscedasticity might provide disturbances to regression-based models. However, these disturbances can be detected and assessed with various statistical tests. To assess autocorrelation, this study utilizes Durbin-Watson test:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} {e_t}^2}, \tag{9}$$

where T is the number of observations and e_t is the error at time t.

The value of d always varies from 0 to 4. Hatekar (2010) suggests that values under 1 or more than 3 might imply autocorrelation and are thus a definite cause for concern. However, the exact threshold value for the Durbin-Watson test depends on the number explanatory variables and observations.

c) White´s test

The presence of heteroscedasticity is tested for with White´s test. In White´s test, the squared residuals from the original regression model are regressed onto a set of regressors comprised of the original explanatory variables with their cross-products and squares. Then, the R^2 obtained from the latter regression is multiplied with the number of observations n, and the result compared to χ^2 -distribution:

$$\chi^2 > nR^2_{\hat{\delta}^2}. \tag{10}$$

If χ^2 is not greater than $nR^2_{\hat{\delta}^2}$, the test implies that there is evidence of heteroscedasticity.

d) Theil´s U

After parameterizing the models using in-sample data, the models are used to generate *ex-post* forecasts. The precision of each *ex-post* forecast is assessed with two different measurements, one of which is Theil´s U. Theil´s U is a relative accuracy measurement, that compares the forecasted result with the result of a naïve forecast with minimal historical data. Theil´s U can help to eradicate models with considerable inaccuracies, since it squares the deviations and thus exaggerates errors. Theil´s U takes values higher than 0, with a value close to 0 indicating an accurate forecast. Values higher than 1 indicate that a naïve forecasting model would out-perform the proposed forecasting method. The formulation for the statistic is presented below:

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{\hat{Y}_{t+1} - Y_{t+1}}{Y_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{Y_{t+1} - Y_t}{Y_t}\right)^2}}, \tag{11}$$

where $\hat{Y}$ is the forecasted value at time *t*, Y_t is the actual value at time *t* and *n* is the number of observations (Theil, 1989).

e) Root-mean-square error (RMSE)

A second method used in this study to assess the *ex-post* forecasting accuracy is root-mean-squared error (RMSE). RMSE represents the standard deviation of the prediction errors, i.e. how spread out the prediction errors are. RMSE value of 0 would indicate a perfect forecasting performance. RMSE values cannot be compared across different datasets, as the measure is scale-dependent. However, RMSE has been found to be a useful tool in comparing forecasting model performance (see e.g. Karakozova, 2004; Brooks and Tsocalos, 2010, p. 271). RMSE is formulated as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(\hat{Y}_t - Y_t\right)^2}{n}}, \tag{12}$$

where $\hat{Y}$ is the forecasted value at time *t*, Y_t is the actual value at time *t* and *n* is the number of observations (Brooks and Tsocalos, 2010, p. 271).

5 RESULTS

In this chapter, the empirical findings of this study are presented. Firstly, the in-sample estimates obtained from ARIMA, SR, MR, VAR and ARIMAX models are assessed. Secondly, after establishing the forecasting models using in-sample data, the out-of-sample forecasting performance is examined for each of the selected models. The chapter is then concluded by interpreting the empirical results and discussing the robustness of these results.

5.1 In-sample estimates

5.1.1 ARIMA model estimates

This section showcases the in-sample estimates gained from the ARIMA model. The formulation of the utilized model is specified in Section 4.2.1 as Equation 1. As ARIMA models can have any AR and MA orders, the correct order selection is based on Akaike Information Criteria with the maximum order set at 4. Table 3 presents the generated information criteria values.

Table 3. Akaike Information Criteria for the ARIMA models

	AICc values				
	AR (0)	AR (1)	AR (2)	AR (3)	AR (4)
MA (0)	—	174.66	168.65	170.62	172.62
MA (1)	170.84	172.73	170.62	172.46	174.42
MA (2)	172.17	172.86	172.62	174.40	176.40
MA (3)	170.93	172.85	174.51	175.44	176.41
MA (4)	172.84	173.29	176.03	175.55	176.18

This table showcases the results for Akaike Information Criteria with a correction for a small sample (AICc). A lower AICc value indicates a better parameterized specification.

As can be seen from the table, Akaike Information Criteria values indicate that ARIMA (2,0,0) is the best parameterized specification among the competing ARIMA models. Thus, ARIMA (2,0,0) is the ARIMA specification selected for the forecasting process. Figure 7 illustrates the in-sample fit of ARIMA (2,0,0).

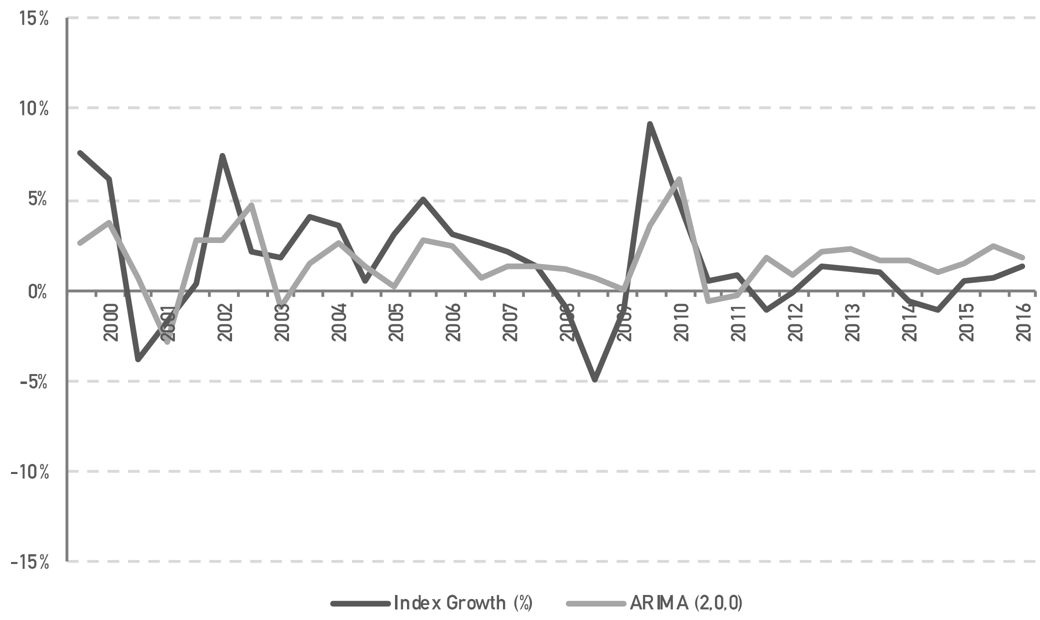

Figure 7. ARIMA (2,0,0) in-sample fit

As it is seen from Figure 7, ARIMA (2,0,0) is able to capture some of the deviation in the Index Growth (%) series. Nonetheless, the model seems to under-estimate the substantial deviations in the series, such as the downturn of 2008 and the subsequent rise. ARIMA (2,0,0) obtains an adjusted R^2 of 0,29.

5.1.2 SR model estimates

SR model is estimated for each of the seven explanatory variables. In addition, as it might take several quarters for housing prices to reflect new information (See Section 3.1 for more information), each explanatory variable is estimated for three different lags: t, t-1 and t-2. In the regressions regarding HS and HC, the regressions were first run with seasonal dummy variables. However, as the dummy variables did not improve the models and were statistically insignificant, they were omitted from the final regressions. The formulation of the utilized SR model is specified in Section 4.2.1 as Equation 2. Table 4 presents the statistical analysis for the regressions.

Table 4. Akaike Information Criteria and R^2 for the SR models

Lag	t		t-1		t-2	
Explanatory variable	R^2	AICc	R^2	AICc	R^2	AICc
GDP (%, HOH)	0.1195	173.57	0.1352	165.02	0.1355	159.92
IR (%, HOH)	0.0016	178.79	0.3264	159.77	0.1052	159.94
NC (%, HOH)	0.0000	177.84	0.0075	169.53	0.0082	162.79
PG (%, HOH)	0.0248	176.99	0.0164	169.23	0.0003	163.05
HS (%, HOH)	0.0344	176.65	0.0148	169.29	0.0000	163.06
HC (%, HOH)	0.0080	177.57	0.0746	167.21	0.0464	161.54
DI (%, HOH)	0.0353	176.62	0.1461	164.56	0.1068	159.44

This table showcases the results for Akaike Information Criteria with a correction for a small sample (AICc) and R^2 for three different lags: t, t-1 and t-2. A lower AICc value indicates a better parameterized specification and higher R^2 indicates a better in-sample fit. GDP = Gross domestic product, IR = Interest rate, NC = New contracts, PG = Population growth, HS = Housing starts, HC = Housing completions, DI = Disposable income, HOH = half-over-half growth.

As can be seen from the table, while IR_{t-1} has the highest R^2, AICc suggests that DI_{t-2} is the best suited explanatory variable for modelling the Index Growth (%) series. Thus, SR model with DI_{t-2} as an explanatory variable is then selected for further examination. The model obtains Durbin-Watson test result of 1.361, which suggests that there is no great concern for autocorrelation. Similarly, a p-value of 0.905 obtained from the White´s test suggests that there is no evidence of heteroscedasticity. Table 5 presents the details of the SR model with DI_{t-2} as an explanatory variable and Figure 8 illustrates the in-sample fit.

Table 5. Details of the SR model

Explanatory variable(s)	Estimate	t-value	p-value
Intercept	0.012	2.724	0.011
DI_{t-2}	-0.120	-3.876	0.001
Diagnostic tests			
DW	1.361	—	0.312
White´s test	—	—	0.905
R^2	0.106	—	—
Adjusted R^2	0.076	—	—

This table showcases the results of the Simple Regression model, as well as the results of the Durbin-Watson test (DW) and the White´s test. The null hypothesis of the Durbin-Watson test is that the residuals are not autocorrelated, and the null hypothesis of the White´s test is that the variance of errors is homoscedastic.

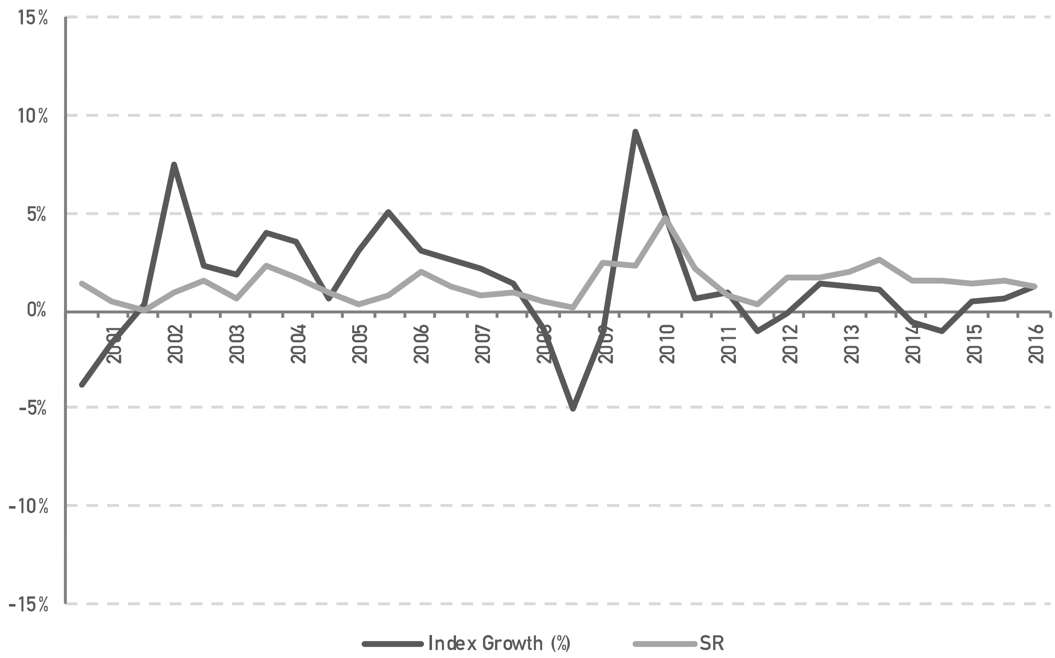

Figure 8. SR model in-sample fit

As illustrated in Figure 8, the SR model is able to obtain some of the deviations in the Index Growth (%) series, but overall the explanatory power is low. What is more, the DI variable curiously changes its sign to negative when lagged for two periods. As with the ARIMA model, the SR model is not able to capture the downturn in 2008 and generally underestimates the changes in the dependant variable. The low explanatory power is also evident from the obtained adjusted R^2 of 0.08, which is considerably lower than the one obtained from the ARIMA model.

5.1.3 MR model estimates

MR model (specified in Section 4.2.1 as Equation 3) is built following a procedure often referred to as “stepwise selection”, where variables are added to the regression in order of their statistical significance. Variables are only added to the regression if they are significant on a prespecified tolerance level, and similarly, variables are removed from the regression if their significance falls below this prespecified level. In this case, a significance level of 10% was used. Correspondingly with the SR model, each of the seven explanatory variables were examined in three different lags. As with the SR model, the seasonal dummy variable was omitted from the final regression as it did not seem to improve the specification. Table 6 presents the details of the MR model obtained with "stepwise selection" and Figure 9 illustrates the in-sample fit.

Table 6. Details of the MR model

Explanatory variable(s)	Estimate	t-value	p-value
Intercept	0.006	1.406	0.171
GDP_t	0.638	3.064	0.005
IR_{t-1}	-0.133	-5.028	0.000
PG_{t-2}	0.077	1.808	0.082
HC_t	-0.034	-1.774	0.087
Diagnostic tests			
DW	1.722	—	0.260
White´s test	—	—	0.809
R^2	0.564	—	—
Adjusted R^2	0.499	—	—

This table showcases the results of the Multiple Regression model, as well as the results of the Durbin-Watson test (DW) and the White´s test. The null hypothesis of the Durbin-Watson test is that the residuals are not autocorrelated, and the null hypothesis of the White´s test is that the variance of errors is homoscedastic.

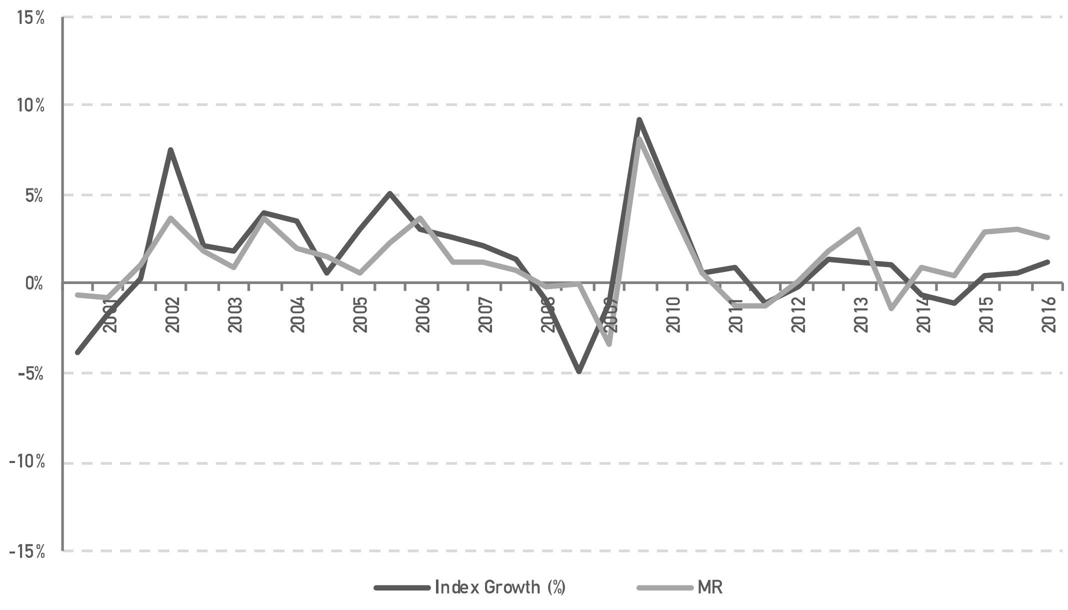

Figure 9. MR model in-sample fit

As the figure illustrates, the MR model mostly tracks changes in the dependant variable. The obtained adjusted R^2 of 0.499 implies that changes in GDP, Interest Rate, Population Growth and Housing Completions are collectively able to explain approximately half of the deviations in the Index Growth (%). Yet, the MR model seems to overestimate the dynamics of the dependent variable, mainly from 2012 onwards.

Despite the overestimation, the model seems to be well parameterized. All of the explanatory variables are correctly signed and significant at a 10 % significance level. What is more, Durbin-Watson test result of 1.722 suggest that there are no difficulties with autocorrelation. Similarly, White´s test p-value of 0.809 indicates that we do not reject the null hypothesis of homoskedasticity.

5.1.4 VAR model estimates

VAR model (specified in Section 4.2.1 as Equation 4) is then estimated using the same variables as in the MR equation, i.e. Gross Domestic Product, Interest Rate, Population Growth and Housing Completions. The VAR model treats all variables as endogenous and thus also examines the effect of the forecasted variable on itself. The selection of lag length of the VAR model is based on Akaike Information Criteria, and as with the ARIMA model, the maximum lag length is set at 4. Table 7 presents the lag length criteria.

Table 7. Lag length criteria of the VAR model

VAR model	AICc	R^2
VAR (1)	16.866	0.510
VAR (2)	17.339	0.574
VAR (3)	17.424	0.706
VAR (4)	14.648	0.799

This table showcases the results for Akaike Information Criteria with a correction for a small sample (AICc) and R^2 for four different VAR (p) models: VAR (1), VAR (2), VAR (3) and VAR (4). A lower AICc value indicates a better parameterized specification and higher R^2 indicates a better in-sample fit.

As can be seen from the table, Akaike Information Criteria values indicate that VAR (4) is the best parameterized specification among the competing VAR (p) models. Accordingly, VAR (4) is selected for further examination. Table 8 presents the details of the VAR (4) model and Figure 10 illustrates the in-sample fit.

Table 8. Details of the VAR (4) model

Variable	Estimate	t-value	p-value
$Index_{t-1}$	0.252	0.748	0.471
GDP_{t-1}	0.105	0.139	0.892
IR_{t-1}	-0.107	-1.283	0.229
PG_{t-1}	-0.026	-0.329	0.749
HC_{t-1}	0.034	0.752	0.469
$Index_{t-2}$	-0.374	-1.004	0.339
GDP_{t-2}	-0.042	-0.056	0.956
IR_{t-2}	-0.003	-0.030	0.977
PG_{t-2}	0.108	1.041	0.322
HC_{t-2}	-0.036	-0.792	0.447
$Index_{t-3}$	-0.154	-0.498	0.630
GDP_{t-3}	0.479	0.725	0.485
IR_{t-3}	-0.085	-0.889	0.395
PG_{t-3}	-0.071	-0.613	0.554
HC_{t-3}	-0.051	-1.490	0.166
$Index_{t-4}$	0.052	0.162	0.875
GDP_{t-4}	0.922	1.302	0.222
IR_{t-4}	-0.139	-1.523	0.159
PG_{t-4}	0.098	1.030	0.327
HC_{t-4}	0.018	0.473	0.646
Diagnostic tests			
DW	1.883	—	0.374
White´s test	—	—	1.000
R^2	0.799	—	—
Adjusted R^2	0.396	—	—

This table showcases the results of the VAR (4) model, as well as the results of the Durbin-Watson test (DW) and the White´s test. The null hypothesis of the Durbin-Watson test is that the residuals are not autocorrelated, and the null hypothesis of the White´s test is that the variance of errors is homoscedastic.

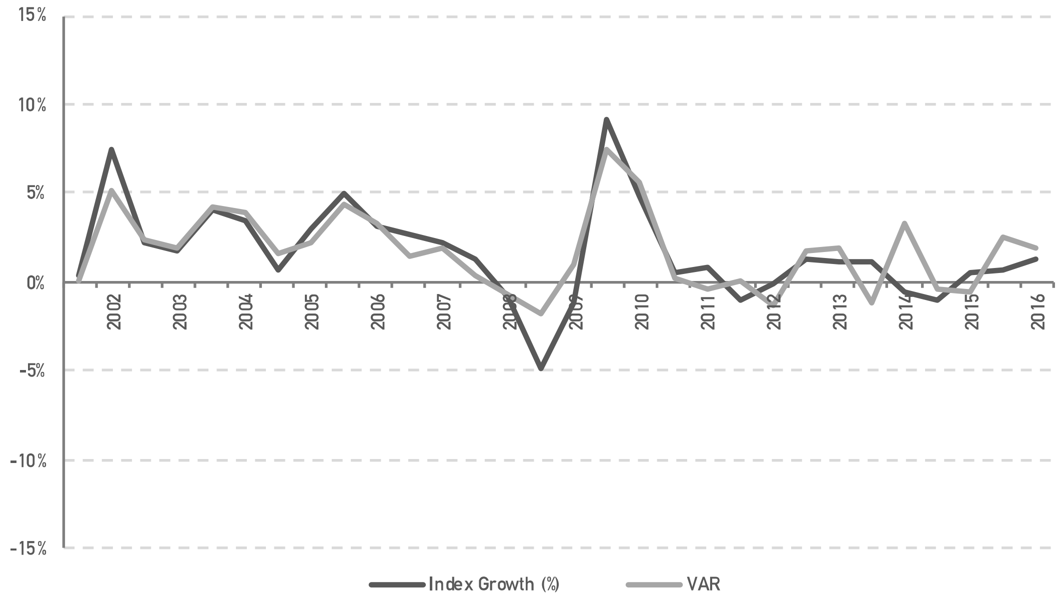

Figure 10. VAR (4) model in-sample fit

As it is seen from Figure 10, VAR (4) explains a remarkably large part of the deviations in Index Growth (%) series, apart from few inaccuracies in mid-2010s. VAR (4) obtains a very high R^2 of 0.799. However, the empirical explanatory power of VAR (4) model is of low accuracy, as indicated by low p-values and low adjusted R^2. In other words, VAR (4) struggles to capture the conceivably causal relationship between the variables. Nevertheless, Durbin-Watson test result of 1.883 and White´s test p-value of 1.000 indicate that autocorrelation and heteroskedasticity are not interfering with the modeling results.

5.1.5 ARIMAX model estimates

Finally, ARIMAX model (specified in Section 4.2.1 as Equation 5) is estimated. ARIMAX is estimated for each of the seven explanatory variables in three different lags: t, t-1 and t-2. As is the case with the ARIMA model, ARIMAX is run with the maximum AR and MA orders set at 4. The most suitable model for forecasting purposes is then decided with Akaike Information Criteria. As with the previous specifications, seasonal dummies did not seem to improve the models utilizing HS and HC variables and were thus omitted. Table 9 presents the generated information criteria values.

Table 9. Akaike Information Criteria for the ARIMAX models

	AICc values																			
ARIMAX order	1,0,0	1,0,1	1,0,2	1,0,3	1,0,4	2,0,0	2,0,1	2,0,2	2,0,3	2,0,4	3,0,0	3,0,1	3,0,2	3,0,3	3,0,4	4,0,0	4,0,1	4,0,2	4,0,3	4,0,4
GDP_t	172.77	170.76	166.44	165.57	167.52	165.08	162.79	164.72	164.74	166.81	166.32	168.38	166.71	168.3	168.14	167.74	166.58	168.39	168.42	170.29
GDP_{t-1}	162.53	163.44	160.53	165.82	162.97	162.91	164.81	163.41	163.58	165.09	164.67	165.92	163.88	165.09	165.83	165.43	166.28	164.83	166.76	169.33
GDP_{t-2}	157.80	158.10	158.38	160.10	161.25	157.69	159.61	159.73	158.88	160.88	159.67	161.50	161.88	160.88	162.46	160.61	162.55	162.89	164.70	164.59
IR_t	176.59	174.73	174.81	174.46	175.12	170.50	172.47	174.47	176.35	177.91	172.47	174.25	176.18	177.18	177.14	174.47	176.21	178.22	178.31	178.05
IR_{t-1}	157.00	158.79	158.73	160.66	160.22	158.46	159.90	156.36	160.02	161.74	160.44	160.52	158.33	160.04	162.63	160.71	161.64	163.64	165.58	162.71
IR_{t-2}	159.80	158.42	158.34	159.59	161.46	157.41	159.34	161.03	160.13	162.01	159.38	160.68	163.04	161.97	164.02	160.85	162.85	162.94	165.17	164.53
NC_t	176.66	174.53	174.47	174.82	174.94	170.36	172.35	174.35	176.27	177.82	172.35	174.31	176.29	176.83	176.60	174.35	176.30	178.26	178.20	177.86
NC_{t-1}	169.91	169.86	170.09	165.97	165.95	166.44	168.35	170.35	166.70	168.82	168.34	169.30	167.24	167.91	168.81	170.32	171.20	173.30	169.91	170.00
NC_{t-2}	160.68	158.43	158.15	160.00	162.62	157.56	159.52	160.01	159.26	161.14	159.55	160.57	162.84	164.12	163.14	160.67	162.67	164.67	165.02	164.41
PG_t	176.26	174.72	174.82	174.85	175.06	170.54	172.52	174.52	176.40	176.83	172.52	174.38	176.31	177.10	177.53	174.51	176.33	178.26	178.36	178.18
PG_{t-1}	169.56	168.83	168.62	168.30	170.11	165.18	167.17	169.13	169.46	171.43	167.17	169.17	170.52	171.39	171.85	169.11	170.90	173.07	172.89	173.20
PG_{t-2}	160.57	158.01	158.06	159.89	161.32	157.69	159.47	159.90	160.19	161.88	159.64	161.01	161.89	163.43	163.69	160.82	162.82	162.94	163.63	164.65
HS_t	175.03	172.93	173.50	172.93	172.95	168.83	170.83	172.77	174.59	175.13	170.83	172.68	174.39	176.23	175.24	172.78	174.48	176.34	176.78	176.14
HS_{t-1}	169.90	169.77	169.88	168.40	171.14	166.20	168.20	170.19	170.51	171.75	168.20	170.00	171.61	171.74	173.36	170.20	171.87	171.49	173.82	173.83
HS_{t-2}	160.79	158.43	158.23	160.04	161.33	157.65	159.50	160.07	160.34	162.00	159.61	160.73	161.71	161.90	163.76	163.76	162.71	162.84	165.09	164.33
HC_t	176.20	174.67	174.83	174.00	175.22	170.62	172.57	174.57	176.42	178.02	172.57	174.35	176.32	176.98	176.53	174.56	176.33	178.32	178.32	177.63
HC_{t-1}	168.38	168.66	168.80	164.74	166.43	165.96	167.96	169.95	167.22	168.00	167.96	169.65	166.80	168.45	170.49	169.95	171.47	168.71	169.70	171.64
HC_{t-2}	160.57	158.63	158.37	160.14	163.19	157.68	159.61	160.17	160.00	161.80	159.65	161.29	163.14	161.75	165.07	160.85	162.85	162.91	164.53	163.92
DI_t	175.63	172.73	171.22	171.34	173.29	168.58	170.45	172.00	171.41	174.15	170.52	172.44	173.97	174.96	174.62	171.73	173.70	171.84	173.70	172.93
DI_{t-1}	163.67	164.25	166.05	166.55	168.70	163.56	165.47	167.31	168.53	170.98	165.42	167.27	166.76	168.34	171.89	167.20	168.86	170.51	172.51	173.16
DI_{t-2}	158.81	158.61	158.07	159.39	161.38	157.69	159.62	159.35	160.18	162.04	159.67	161.45	161.35	161.98	163.87	160.34	162.33	164.30	163.80	164.68

This table showcases the results for Akaike Information Criteria with a correction for a small sample (AICc). A lower AICc value indicates a better parameterized specification.

As can be seen from the table, Akaike Information Criteria values indicate ARIMAX (2,0,2) IR_{t-1} to be the best parameterized specification. Thus, ARIMAX (2,0,2) IR_{t-1} is then selected for further examination. Figure 11 illustrates the in-sample fit of the model.

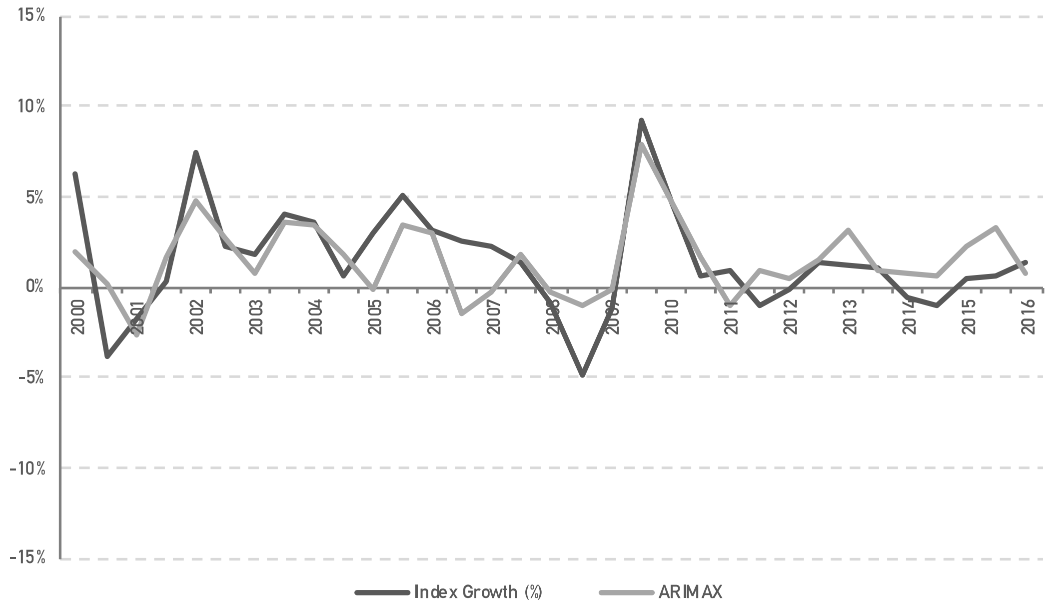

Figure 11. ARIMAX (2,0,2) IR_{t-1} model in-sample fit

As can be observed from Figure 11, ARIMAX (2,0,2) IR_{t-1} tracks the Index Growth (%) series well. The model achieves a relatively high adjusted R^2 of 0.54, which is notably higher than that of ARIMA, as the added component of relevant explanatory variable seemingly enhances the explanatory power of the model.

5.2 Out-of-sample forecast estimates

This section is structured in a following way. Firstly, the descriptive statistics are presented for all of the models selected in the previous section, expressing the predictive power of each model. The predictive power of the specified models is examined in the ex-post period spanning from 2016 H2 to 2018 H2. Secondly, the findings that can be derived from the results are discussed, as well as whether there is a clear distinction in the forecasting power between the models.

The precision of each *ex-post* forecast is assessed with Theil´s U and RMSE (specified in Section 4.2.2 as Equation 10 and Equation 11 respectively). Table 10 below summarizes the results.

Table 10. Out-of-sample forecasting accuracy

Period	2016 H2	2017 H1		2017 H2		2018 H1		2018 H2	
Model	RMSE	U	RMSE	U	RMSE	U	RMSE	U	RMSE
ARIMA	0.569	1.879	0.921	0.592	0.802	0.580	0.702	0.736	0.845
SR	0.751	0.541	0.582	0.571	0.645	0.573	0.613	0.595	0.628
MR	1.277	3.338	1.726	0.905	1.434	0.945	1.393	0.880	1.253
VAR	2.854	8.006	169.23	2.021	3.320	1.990	2.916	2.608	3.334
ARIMAX	1.027	2.952	1.489	0.778	1.230	0.761	1.068	1.093	1.340

This table showcases the accuracy measurements for each of the examined forecasting models for the *ex-post* period spanning from 2016 H2 to 2018 H2. The forecasting accuracy is measured with Theil´s U (U) and Root-mean-square error (RMSE). Both Theil´s U and RMSE take values higher than 0, with a value close to 0 indicating an accurate forecast. In the case of Theil´s U, values below 1 indicate that the proposed forecasting method would out-perform a naïve forecasting model.

As the results from Table 10 indicate, there are large variations in the *ex-post* forecasting accuracy of the selected models. On the whole, the forecasting accuracy is not desirable as the error measurement statistics (Theil´s U and RMSE) obtain quite high values. However, most of the forecasts reach a Theil´s U value below 1, which indicates that they would indeed out-perform a naïve forecasting model.

Evidently, the SR model is able to reach lower error measurement statistics as its counterparts. The SR model is in fact only outperformed on one occasion, as the ARIMA model produces a better forecast for the one-step-ahead forecast for 2016 H2. More complicated models, such as VAR and ARIMAX, perform poorly on average, yielding far higher error measurement statistics than the other competing models. Figures 12, 13 and 14 illustrate the out-of-sample forecasting ability of each model.

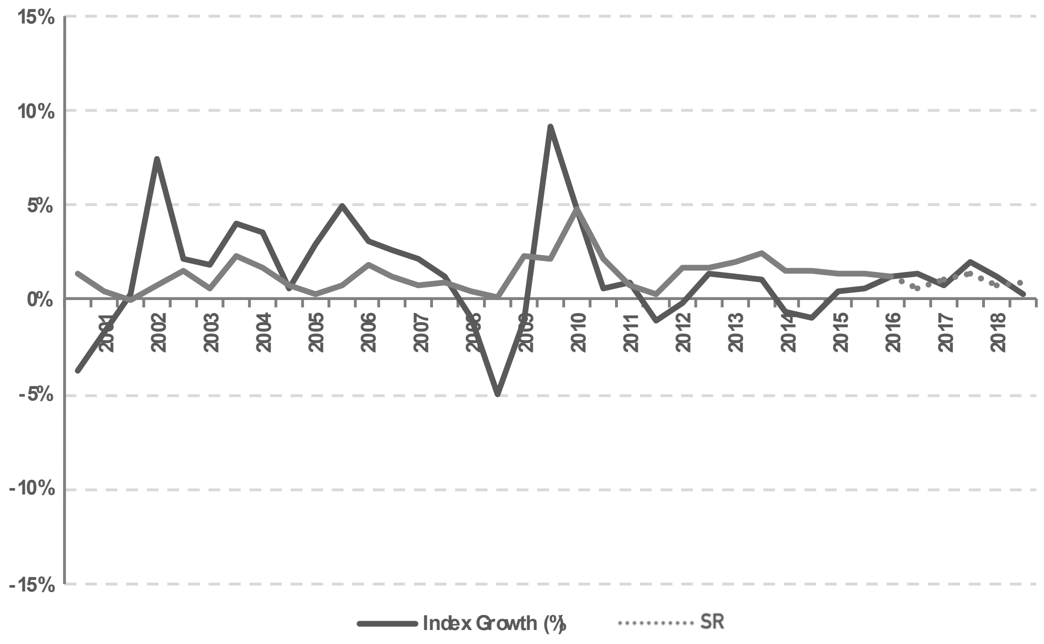

Figure 12. Out-of-sample accuracy of the SR model

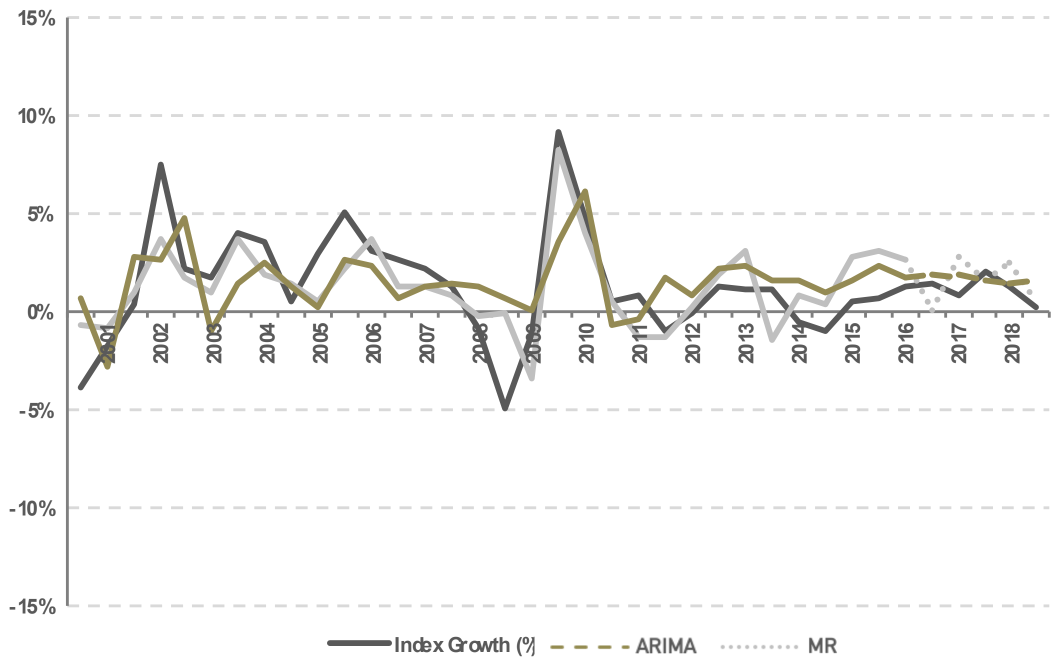

Figure 13. Out-of-sample accuracy of ARIMA and MR models

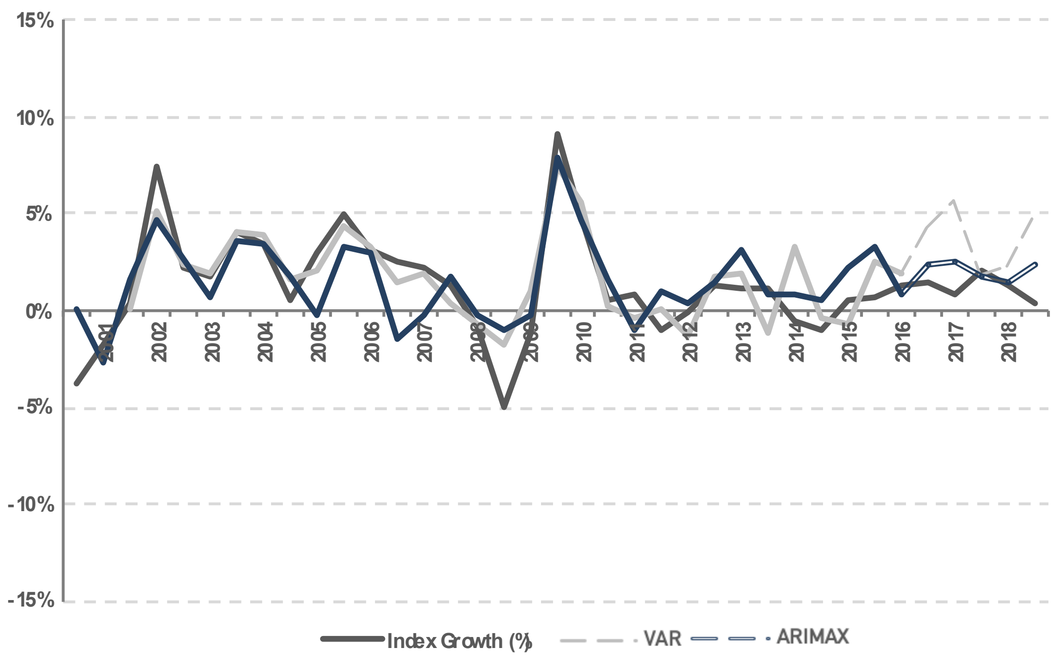

Figure 14. Out-of-sample accuracy of VAR and ARIMAX models

The consistent forecasting accuracy achieved by SR model is evident from Figure 12, as the out-of-sample forecasts from the SR model mimic the movements in the Index Growth (%) series quite accurately. What is more, the figure also illustrates the relatively low in-sample R^2 of the SR model, which makes it far superior out-of-sample performance quite peculiar. Furthermore, Figures 13 and 14 illustrate that while the other competing models, such as VAR and ARIMAX, track the movements of the Index Growth (%) series considerably well in-sample, they seem to exaggerate the magnitude of the volatility out-of-sample.

5.3 Interpretation and robustness of the findings

This section discusses the interpretation of the main findings, as well as the robustness of these findings through the potential limitations and biases that the utilized dataset and methodology poses. The empirical evidence suggests that the SR model, which is the simplest of the evaluated models, outperforms the other more complicated models in predicting the evolution of the housing price growth in the HMA. That said, there are still multiple factors that might affect the empirical results in one way or another.

Although the study was executed with utmost meticulousness, there were some inevitable limitations and biases in the dataset and methodology. First of all, the study considers only a single time period from a single housing market. As such, the results from this study might not be applicable or reproductible in other timespans or in other markets. Secondly, the study utilizes semi-annual time-series data from 1999 H2 to 2018 H2, which contains only 34 observations for the initialization period. Although 34 observations were established to be sufficient for the means of the study, the study would benefit from a longer time-series. As the dataset contains such a low number of in-sample observations, it could be speculated that the complex models considered in this study inevitably overfit the data with too many parameters. Thirdly, in addition to the low number of in-sample observations, there are also other limitations with the utilized data. Local data for the HMA was hard to come by and as such, nation-wide variables were used. It is also evident that it is very challenging to make accurate predictions in the HMA housing market with the selected explanatory variables, which is apparent through the low overall accuracy of the forecasts. Better accuracy might be achievable with the utilization of other explanatory variables. Finally, the choice of AICc as the model selection criterion could be another potential point of critique, as it is not considered accurate with altering time-series lengths.

However, a considerable effort has been put into the data acquisition and the methodology selection to produce as thorough an empirical analysis as possible. The aforementioned limitations and biases might somewhat distort the results, but it seems reasonable to believe that as long as they are considered, the empirical results of this study are reliable enough to make well-grounded conclusions. As such, the main finding of this study can be derived. Simple models can outperform their more complex counterparts – at least in certain markets and in certain conditions. In the case of the HMA, the SR model seems to flourish in data scarcity, as the more complex models, on the contrary, run the risk of overfitting the relatively small data sample. However, it seems probable that utilizing a lengthier and more descriptive dataset could significantly tip the scales in favor of more complex models, as it would allow them to extract significantly more information from the data. Further, the results from this study also allude that in-sample R^2 of the forecasting model is not correlated with the accuracy of its out-of-sample forecasts.

6 CONCLUSIONS

The object of this study was to investigate whether it is possible to gain similar forecasting performance from simple forecasting models compared to more complex specifications in housing market context. Although there are numerous studies trying to predict price patterns of various housing markets, only a handful of studies have compared the varying performance of different forecasting methods. In Finnish housing market context, the issue has been largely unexplored, as previous academic research has primarily focused on understanding the market dynamics (see e.g. Kuismanen et al., 1999; Oikarinen, 2007). This study was intended to fill this void.

This study contributes to the existing literature in two main ways. Firstly, to the best of my knowledge, this is the first study comparing various forecasting methods in the Finnish housing market context. As housing market participants are often faced with a plethora of available forecasting methods, choosing the right one could have large-scale implications for the achieved accuracy. The results from this study highlight that added model complexity does not necessarily yield better results. Therefore, market participants should embrace simplicity and user-friendliness in their forecasts. However, it seems probable that the shortcomings of the more complex models in this study are aggravated by the very specific features of the utilized dataset. Hence, above all, market participants should acknowledge that the obtained forecasting results are always not only largely dependent on the chosen methodology, but also on the utilized dataset. Secondly, this study adds to the research on the housing market of Helsinki metropolitan area by broadly examining its forecastability with publicly available data. Majority of the previous literature in the HMA context has concentrated on the forecastability of commercial and industrial real estate markets.

The empirical analysis of this study compares the predictive power of five distinct forecasting models out-of-sample: Autoregressive Integrated Moving Average (ARIMA), Simple Regression (SR), Multiple Regression (MR), Vector Autoregression (VAR) and Autoregressive Integrated Moving Average with a vector of explanatory variables (ARIMAX). The model selection follows previous literature, as it is based on the classification of prominent real estate forecasting models by Lizieri

(2009). Overall, 470 distinct specifications are calculated, of which the best parameterized specifications are selected for each of the forecasting methods. Then, 25 forecasts in total are generated. Due to constraints in data availability, the empirical analysis considerers a semi-annual data sample from 1999 H2 to 2018 H2.

The obtained results clearly indicate that the Simple Regression (SR) model outperforms the other considered models, as it achieves the most consistent and accurate forecasts for the growth-rate of the HMA housing index out-of-sample. Since the SR model also has the lowest in-sample R^2 of the compared models, the results indicate that the in-sample R^2 is not correlated with the actual forecasting accuracy. Thus, it seems that simple forecasting models can outperform their more complex counterparts – at least in data scarcity, as the more complex run the risk of overfitting small data samples. These findings are in line with previous literature, as similar evidence of the low correlation with the in-sample R^2 and the obtained forecasting accuracy, and the tendency for the more complex models to mistake noise for information are broadly observed in the field of commercial property forecasting (see e.g. Chaplin, 1999; Patrick et al., 2000; Karakozova, 2004). However, it seems probable that utilizing a lengthier and more descriptive dataset could significantly tip the scales in favor of more complex models, and thus the results from this study might not be applicable or reproductible in other timespans or in other markets.

As such, there are still multiple areas for further research. Firstly, as more economic data becomes readily available for the HMA, it would be of great interest to examine whether the preeminence of simpler models persists over multiple time periods and with lengthier in-sample datasets. Secondly, it would be relevant to examine how state-of-the-art forecasting methods based on artificial intelligence would perform against the more traditional methods evaluated in this study. Artificial intelligence methods, such as artificial neural networks, pattern recognition and data mining, are quickly gaining traction in the field of real estate forecasting and have shown great promise. What is more, artificial intelligence methods often utilize regularization technique to address overfitting, which was the main concern with the complex models utilized in this study. Finally, the effect of structural breaks on the obtained forecasting accuracy warrants further research. Structural breaks have been pinpointed as one of the main culprits of forecast failures, and hence, examining which of the forecasting techniques are best able to bear structural breaks, e.g. policy changes concerning the housing sector, is crucial.

REFERENCES

Adams, Z., & Füss, R., (2010). Macroeconomic determinants of international housing markets. *Journal of Housing Economics, 19,* 38-50. doi:10.1016/j.jhe.2009.10.005.

Al-Marwani, H., A. (2014). *An approach to modelling and forecasting real estate residential property market.* Brunei University Press.

Batty, M., & Torrens, P. (2001). Modeling complexity: The limits to prediction. *Cybergeo, 2001(36)*. doi:10.4000/cybergeo.1035

Belsky, E. & Prakken, J. (2004). *Housing wealth effects: Housing's impact on wealth accumulation, wealth distribution and consumer spending.* Cambridge: Joint Center for Housing Studies, Harvard University.

Bork, L., & Møller, S. 2015. Forecasting house prices in the 50 states using Dynamic Model Averaging and Dynamic Model Selection. *International Journal of Forecasting, 31(1),* 63-78. doi:10.1016/j.ijforecast.2014.05.005

Borowiecki, K., (2009). The determinants of house prices and construction: An empirical investigation of the Swiss housing economy. *International Real Estate Review, 12(3),* 193–220. Retrieved from: https://econpapers.repec.org/article/ireissued/v_3a12_3an_3a03_3a2009_3ap_3a193-220.html

Brooks, C. (2008). *Univariatetime series modelling and forecasting*. Cambridge University Press.

Brooks, C., & Tsolacos, S. (2010*). Real estate modelling and forecasting.* Cambridge University Press.

Brown, J., Song, H., & McGillivray, A. (1997). Forecasting UK house prices: A time varying coefficient approach, *Economic Modelling, 14(4),* 529-548. doi:10.1016/S0264-9993(97)00006-0

Buede, D. (2009). Errors associated with simple versus realistic models. *Computational & Mathematical Organization Theory, 15(1),* 11-18. doi:10.1007/s10588-008-9047-x

Case, B., Goetzmann, W., & Rouwenhorst, K. (2000). *Global real estate markets - Cycles and fundamentals.* NBER.

Case, K., & Quigley, J., & Shiller. R. (2005). Comparing wealth effects: The stock market versus the housing market. *Advances in Macroeconomics, 5(1),* 1235-1235. doi:10.3386/w8606

Case, K., & Shiller, R,. (1989). The efficiency of the market for single-family homes. *American Economic Review, 79,* 125-137. doi:10.2139/ssrn.1491725

Case, K., & Shiller, R,. (1990). Forecasting prices and excess returns in the housing market. *Real Estate Economics, 18(3),* 257-273. doi:10.1111/1540-6229.00521

Chaplin, R. (1999). The predictability of real office rents. *Journal of Property Research, 16(1),* 21-49. doi:10.1080/095999199368247

Clements, M.P. and Hendry, D.F. (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics, 58,* 657-684. doi:10.1111/j.1468-0084.1996.mp58004005.x

Collyns, C., & Senhadji, A. (2002). *Lending booms, real estate bubbles, and the Asian crisis.* IMF.

Crawford, G., & Fratantoni, M. (2003). Assessing the forecasting performance of Regime-Switching, ARIMA and GARCH models of house prices. *Real Estate Economics, 31,* 223-243. doi:10.1111/1540-6229.00064

Durka, P., & Pastorekova, S. *(2012). ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series?* International Conference Mathematical Methods in Economics. Retrieved from: http://mme2012.opf.slu.cz/proceedings/pdf/024_Durka.pdf

Edin, P. A., Englund, P. (1991). Moving costs and housing demand: Are recent movers really in equilibrium. *Journal of Public Economics, 44 (3),* 299-320. doi:10.1016/0047-2727(91)90017-V

Englund, P. (2011). *Swedish house prices in an international perspective.* Riksbank.

Fama, E.F. (1991). Efficient capital markets: II. *The Journal of Finance, 46,* 1575-1617. doi:10.1111/j.1540-6261.1991.tb04636.x

"Gerardi, K. S., Foote, C. L., & Willen, P. S. (2010). *Reasonable people did disagree: Optimism and pessimism about the U.S. housing market before the crash.* Federal Reserve Bank of Boston.

Gu, A., (2002). The predictability of house prices. *Journal of Real Estate Research, 24,* 213-234. Retrieved from: https://ssrn.com/abstract=986619

Gyourko, J. & Voith, R., (1992). Local market and national components in house price appreciation. *Journal of Urban Economics, Elsevier, 32(1),* 52-69. doi:10.1016/0094-1190(92)90014-C

Hatekar, N. (2010). *Principles of econometrics: An introduction (using R).* SAGE Publications

Hekman, J.S. (1985). Rental price adjustment and investment in the office market. *Real Estate Economics, 13, 32-47.* doi:10.1111/1540-6229.00339

Hott, C., & Monnin, P. (2008). Fundamental real estate prices: An empirical estimation with international data. *The Journal of Real Estate Finance and Economics, 36,* 427-450. doi:10.1007/s11146-007-9097-8

Huovari,. J, Pakkanen, M., Volk, R. (2005). *Development of regional housing market till 2007.* Pellervo Economic Research Institute.

Hurvich, C., Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika, 76(2*), 297-307. doi:10.1093/biomet/76.2.297

Jacobsen, D., & Naug, B. (2004). What influences the growth of household debt? *Economic Bulletin, 3(4),* 103–111. Retrieved from: https://www.norges-bank.no/globalassets/upload/publikasjoner/economic_bulletin/2004-03/jacobsen.pdf

Jadevicius, A. (2014). *An evaluation of the use of combination techniques in improving forecasting accuracy for commercial property cycles in the UK.* Edinburgh Napier University.

Jadevicius, A., & Huston, S. (2015). Property market modelling and forecasting: simple vs complex models. *Journal of Property Investment & Finance, 33(4),* 337-361. doi:10.1108/JPIF-08-2014-0053

Jadevicius, A., Sloan, B., & Brown, A. (2013). *Property market modelling and forecasting: A case for simplicity.* European Real Estate Society (ERES).

Jin, Y., & Zeng, Z. (2004). Residential investment and house prices in a multi-sector monetary business cycle model. *Journal of Housing Economics, 13,* 268-286. doi:10.1016/j.jhe.2004.08.001

Jud, G. & Winkler, D., (2002). The dynamics of metropolitan housing prices. *Journal of Real Estate Research, 23,* 29-46. doi:10.5555/rees.23.1-2.363x90jp6p70pp26

Kajosaari, A. (2016). *Regional living habits in Helsinki.* Retrieved from: https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/17_01_02_Tilastoja_41_2016_Kajosaari.pdf

Karakozova, O. (2004). Modelling and forecasting office returns in the Helsinki area. *Journal of Property Research, 21(1),* 51-73. doi:10.1080/0959991042000254579

Kennedy, P.E. (2002). Sinning in the basement: What are the rules? The ten commandments of applied econometrics. *Journal of Economic Surveys, 16,* 569-589. doi:10.1111/1467-6419.00179

Kivistö, J. (2012). *An assessment of housing price developments against various measures.* Bank of Finland

Koop, G. (2006). *Analysis of financial data*. John Wiley & Sons, Ltd.

Koskela, E., Loikkanen, H., A. Virén, M. (1992). House prices, household savings and financial market liberalization in Finland. *European Economic Review, 36(2-3),* 549-558. doi:10.1016/0014-2921(92)90112-A

KTI Kiinteistötieto Oy (2019). *The Finnish Property market*. Retrieved from: https://kti.fi/wp-content/uploads/The-Finnish-Property-Market-2019.pdf

Kuismanen, M., Laakso, S. & Loikkanen, H. (1999). Demographic factors and the demand for housing in the Helsinki Metropolitan Area. *Government Institute for Economic Research, Discussion Papers, 191*. Helsinki: VATT.

Laakso, S. (2000). *Regional housing markets in boom and bust: The experience from Finland.* Pellervo Economic Research Institute.

Lind, H., (2009). Price bubbles in housing markets: Concept, theory and indicators. *International Journal of Housing Markets and Analysis, 2,* 78-90. doi:10.1108/17538270910939574

Linneman, P. (1986). An empirical test of the efficiency of the housing market. *Journal of Urban Economics, 20(2),* 140-154. doi:10.1016/0094-1190(86)90003-3

Lizieri, C.M. (2009). *Forecasting and modelling real estate.* Retrieved from: https://www.yumpu.com/en/document/read/40485543/forecasting-and-modelling-real-estate-colin-lizieri-henley

Makridakis, S., Wheelwright, S.C., and Hyndman, R.J. (1998). *Manual of forecasting: Methods and applications.* John Wiley & Sons, Inc.

McCord, M., McGreal, S., Berry, J., Haran, M., & Davis, P. (2011). The implications of mortgage finance on housing market affordability. *International Journal of Housing Markets and Analysis, 4,* 394-417. doi:10.1108/17538271111172175

Mouzakis, F., & Richards, D. (2007). Panel data modelling of prime office rents: A study of 12 major European markets. *Journal of Property Research, 24(1),* 31-53. doi:10.1080/09599910701297713

Muellbauer, J., & Murphy, A. (1997). Booms and busts in the UK housing market. *Economic Journal, 107(445),* 1701-1727. doi:10.1111/j.1468-0297.1997.tb00076.x.

Nordvik, V. (2007). *Boligpriser og forventningsdannelse: sammenhengen mellom forventet og faktisk boligpris*. Norges byggeforsikringsinstitutt.

Oikarinen, E. (2005). *Is housing overvalued in the Helsinki Metropolitan Area?* ETLA.

Oikarinen, E. (2007). *Studies on house pricing dynamics.* Turku School of Economics Press.

Orrell, D. and McSharry, P. (2009). System economics - Overcoming the pitfalls of forecasting models via a multidisciplinary approach. *International Journal of Forecasting, 25(4),* 734-743. doi:10.1016/j.ijforecast.2009.05.002

Patrick, J., Okunev, J., Ellis, C. & David, M. (2000). Comparing univariate forecasting techniques in property markets. *Journal of Real Estate Portfolio Management, 6(3),* 283-306. doi:10.5555/repm.6.3.467l255618870230

Rae, D., & van den Noord, P., (2006). *Ireland's housing boom: What has driven it and have prices overshot?* OECD Economics Department Working Papers.

Rayburn, W., Devaney, M. and Evans, R. (1987). A fest of weak-form efficiency in residential real estate returns. *Real Estate Economics, 15,* 220-233. doi:10.1111/1540-6229.00429

Riddel, M. (2004) Housing-market disequilibrium: An examination of housing-market price and stock dynamics 1967-1998. *Journal of Housing Economics, 13(2),* 120-135. doi:10.1016/j.jhe.2004.04.002

RIL (2019). *State of the built environment in Finland 2019.* Retrieved from: https://www.ril.fi/media/2019/roti/roti_2019_raportti.pdf

Schindler, F. (2011). Market efficiency and return predictability in the emerging securitized real estate markets. *Journal of Real Estate Literature, 19,* 111-150. doi:10.2139/ssrn.1622720

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25(3),* 289–310. doi:10.1214/10-STS330

Stevenson, S. (2007). A comparison of the forecasting ability of ARIMA models. *Journal of Property Investment & Finance, 25(3),* 223- 240. doi:10.1108/14635780710746902

Stevenson, S., & McGarth, O. (2003). A comparison of alternative rental forecasting models: Empirical tests on the London office market. *Journal of Property Research, 20,* 235-260. doi:10.1080/0959991032000162338

Sugiura, (1978). Further analysts of the data by akaike' s information criterion and the finite corrections. *Communications in Statistics - Theory and Methods, 7(1),* 13-26. doi: 10.1080/03610927808827599

Theil, H. (1989). The development of international inequality, 1960–1985. *Journal of Econometrics, 42*, 145-155. doi:10.1016/0304-4076(89)90082-1

Timothy, C.,Sargent T. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics, 8(2),* 262-302. doi:10.1016/j.red.2004.10.009

Tonelli, M., Cowley, M. and Boyd, T. (2004). Forecasting office building rental growth using a dynamic approach. *Pacific Rim Property Research Journal, 10(3),* 283-304. doi:10.1080/14445921.2004.11104164

Tse, Y.R. (1997). An application of the ARIMA model to real-estate prices in Hong Kong. *Journal of Property Finance, 8(2),* 152-163. doi:10.1108/09588689710167843/

Vishwakarma, V., K. (2013). Forecasting real estate business: Empirical evidence from the Canadian market. *Global Journal of Business Research, 7(3),* 1-14. doi:10.1108/IJHMA-03-2019-0036