AN EMPIRICAL ANALYSIS OF NORWEGIAN HOUSE PRICES

by

Kjetil Oeding Amundstad

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in Economics

Charlotte

2017

Approved by:

____

Dr. Craig A. Depken, II

____

Dr. Azhar Iqbal

____

Dr. Thomas Mayock

ABSTRACT

KJETIL OEDING AMUNDSTAD. An Empirical Analysis of Norwegian House Prices. (Under the direction of CRAIG A. DEPKEN II)

This study estimates an econometric model to explain changes in Norwegian house prices. This is useful as it expands the understanding of how and why Norwegian house prices move as they do. The study utilizes Statistics Norway's house prices Index as the dependent variable. The study regards an array of potential determinants of house prices and employs statistical tests to determine the variables' quality. Finally, a vector autoregressive model is estimated. The final VAR model specification is based on a selection procedure made up of SBC, pseudo out-of-sample RMSE, and impulse response functions. The best model specification consists of the variables house price, interest rate, unemployment rate, and two different measures of housing starts. This study uses 60 quarterly observations covering 15 years of data.

ACKNOWLEDGEMENTS

I wish to express my genuine gratitude to Dr. Craig A. Depken II for his investment during the course of this study. I would also like to acknowledge Statistics Norway for providing me with almost all data material used in this thesis. I should like to thank Dr. Azhar Iqbal for providing me with the theoretical understanding necessary to complete this thesis, and for serving as a member of my committee. I should also like to thank Dr. Thomas Mayock for serving as a member of my committee. I would like to thank the Belk College of Business at the University of North Carolina at Charlotte and specifically the organizers of the dual degree program with Copenhagen Business School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Changes in house prices can have far-reaching implications for a country's economy. Housing often represents the largest investment in an individual household's portfolio, as well as the largest share of wealth in most nations' balance sheets [Eurostat, 2013a]. As a consequence, understanding why house prices change over time is crucial for any economy's stability.

Figure 1.1 illustrates the almost continuous increase in Norwegian house prices for the past 15 years. The most significant deviation from this trend happened after the financial crisis in the period 2007 - 2008, where house prices fell 18 percent adjusted for inflation [NRK, 2012]. Economists were quick to interpret this fall as a clear signal of an overpriced housing market. However, by the third quarter of 2009, Norwegian housing prices had already risen beyond their pre-crisis values. The quick comeback and continuing growth the following years has been seen as a relatively unique case in international context. This has led to a discussion of whether Norwegian house prices really can be explained by fundamental factors or if there are bubble tendencies present in the Norwegian housing market.

Multiple renowned economists, including Jeff Gundlach, Robert Shiller, and Nouriel Roubini supported this view as early as 2013 [Shiller, 2008]. Thenceforth, the price of brent crude oil, Norway's biggest export, has dropped below 50 percent of its average 2013 value. This drop has affected almost all aspects of Norway's economy. The unemployment rate is rising and wages are stagnant.

FIGURE 1.1: SSB NORWEGIAN HOUSE PRICE INDEX (2005=100)

The principal objective of this study is to estimate a useful econometric model of Norwegian house prices in order to predict future movements in the variable. The purpose of Chapter 2 is to present relevant economic theory of value to this study. Chapter 3 will discuss the variety of measurement standards that exist for house price indices, before introducing the house price index chosen in this study to represent house prices. A literature review will be carried out in Chapter 4 to uncover variables that have been suggested to determine Norwegian house prices in past studies. Chapter 5 will finally present the proposed variables that ought to be included in the model.

The selection of best model specification will follow a step-wise procedure inspired by Vitner and Iqbal (2010) and will be the topic of chapters 6 - 8. First, the variables are tested to determine if they are (or can be made) stationary processes, a requirement for further modelling. Second, one-by-one Granger-causality tests are run for the purpose of selecting the best variables based on the Chi-squared test. As the three first steps reveal a number of multivariate model specifications, the fourth step is to select the final model specification based on the model that provides the best compromise of lowest Bayesian information criterion and lowest out-of-sample root-mean-square error measure. Finally, impulse response functions are used to further understand the interaction between the variables in the system. A univariate model of house prices is also be developed for comparative purposes.

This study is based on quarterly data for the period 2002 – 2016. Although it is a relatively short time span, it was necessary to keep it to this period as some data series did not have a longer history. Vitner and Iqbal (2010) use 22 years years worth of data in their study of the American housing market. On the other hand, Jacobsen and Naug (2004) estimate their model of Norwegian house prices based on 14 years worth of quarterly data, which leads me to the conclusion that the sample period selected is sufficient.

CHAPTER 2: SUPPLY & DEMAND

The objective of this chapter is to explain the supply and demand concept in context of the housing market. As this is the primary model used to explain price determination in economic theory, it is an important concept to consider. All fundamental determinants of housing prices considered in this study will have an effect on either the demand side or the supply side of the housing market (or both).

Demand is defined as the quantity of a commodity consumers are willing to buy at different prices, while supply is defined as the quantity producers wish to offer at different prices [Whelan and Msefer, 1996]. Producers and consumers are assumed to respond in opposite ways to changes in price. This theory requires some alterations for our purpose, as it assumes the goods in a market to be homogeneous in nature, where one item is the perfect substitute of another, which is not the case for the housing market [Tirole, 1994]. The housing market is heterogeneous in many dimensions, such as in regards to size, type of dwelling and location. Additionally, the infrequent rate of transactions in the housing market makes comparing prices highly problematic [Silver, 2012]. An aggregate model of the housing market thus require standardization in regards to a number of factors to make the products comparable. How to handle these problems will be elaborated on in Chapter 3. For now, it is only important to recognize their existence.

What follows is a description of both the demand side and the supply side of the housing market, primarily focused on discussing matters of significance in the housing market. Finally, this chapter will look at how demand and supply of housing determines the market clearing price in both the short and the long run.

2.1 Demand

Housing is considered a "normal good" in economic terms, meaning that demand for housing generally increases with income. Indirectly, this means that demand increases as price decreases. This price-quantity relationship is illustrated by the downward sloping demand curve as described in Figure 2.1. However, the heterogeneous characteristic of the housing market makes measuring demand for housing somewhat problematic. Ideally, all sub-markets for housing, in regards to both location and type, should be analyzed individually before being aggregated. This would ensure that useful information from sub-markets did not become lost. However, as this task would be both complex and time-consuming, it is not a feasible approach for this study. Additionally, the varying quality of regional data sets would also make it problematic.

2.2 Supply

Supply is generally expected to be upward sloping and convex for changes in price. Higher prices motivates the development of more housing, and vice versa. Additionally, as prices rise more people will be inclined to put their existing homes on the market. However, unlike many markets where an increase in demand can be matched with an almost instant increase in supply, this is not so for the housing market. The production cycle, measured from the decision to build until house completion, takes a considerable amount of time, which makes housing supply inflexible to sudden movements in demand, particularly in the short run. This type of highly inelastic supply means that the shape of the supply curve in the short run is close to vertical [Boug and Dyvi, 2009], as depicted in Figure 2.1. This finding has been essential to the modelling of housing prices by other researchers, which will be further elaborated on in Chapter 4.

2.3 Market Clearing Prices in the Short Run

Figure 2.1 illustrates the theoretical price-quantity relationship in the housing market. At time 0, the market clearing price will be P_0^*, which is at the intersection between the demand curve (D_0) and the supply curve (S_0). A positive shock to Norwegian housing demand, caused by factors such as an increase in income levels or lower interest rates, will lead the demand curve to shift to the right. As supply is assumed to be fixed in the short run, the shift will not lead to an immediate increase in quantity from the supply side of the economy. The result is a short term market clearing quantity that will remain at roughly the same level, while the new short term market clearing price, P_1^*, will be higher.

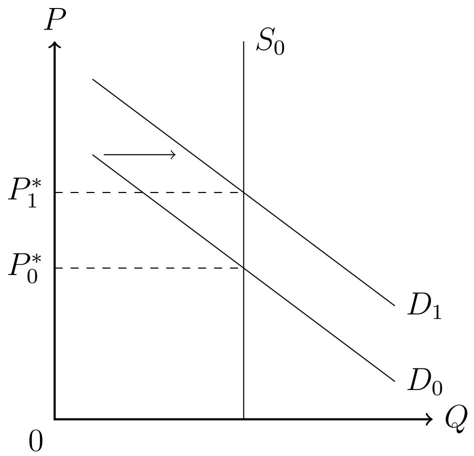

Figure 2.1: SHORT RUN SUPPLY AND DEMAND

2.4 Market Clearing Prices in the Long Run

In the long run, higher house prices will increase the profitability in the building industry. This effect will lead to an increased supply of houses. Consequently, as we can see from Figure 2.2, the supply curve is more elastic in the long run than in the short run. Therefore, supply is expected to increase in the long run to S_1 as a result of an increase in demand, and house prices will move from P_1^* to P_2^* as new houses are supplied to the market. Note that the shift in demand thus has a stronger effect on house prices in the short term than in the long term.

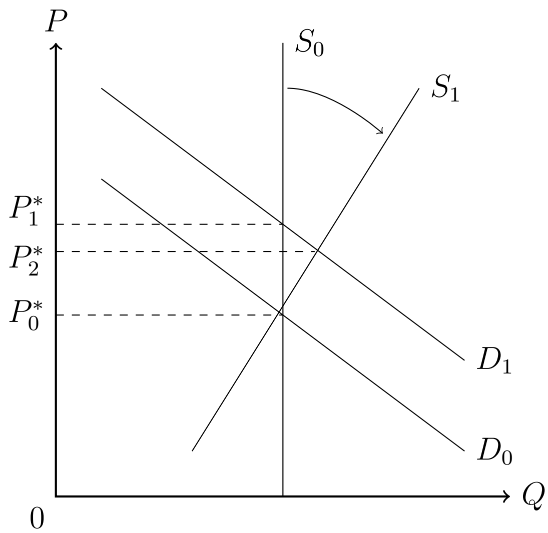

FIGURE 2.2: LONG RUN SUPPLY AND DEMAND

Chapter 4 will further explain how other researchers have interpreted this price relation in the housing market and how it has assisted them in building models on housing prices.

CHAPTER 3: MEASUREMENT STANDARDS

A significant element to consider in this study is the choice of independent variable to use as the proxy for house prices. Therefore, the first objective of this chapter is to provide a thorough investigation into how house price indices are measured and how they differ amongst each other. The second purpose of this chapter is to introduce the independent variable selected to represent house prices in this study.

There are many uses of house price indices (HPI). Policy-makers rely on house price indices in developing monetary and fiscal policies, as well as macro policies directed at the financial and banking sectors [Eurostat, 2013a]. They are also used as a macroeconomic indicator of economic growth, in inflation targeting, as well as for international comparisons. The methodology behind a HPI often reflects the designer's objectives. As a result, one should be careful comparing house price indices across countries and organizations, because the way they are modelled lacks consistency. Eurostat is currently developing a handbook on residential property prices indices, with the hope that future HPIs will have a more consistent framework that allows for easier cross-country analysis [Eurostat, 2013b]. However, to date there is no commonly agreed upon best practice.

Creating a house price index is a challenge, primarily because of housing's inherent heterogeneity. Each dwelling has a unique set of characteristics, including location, age of construction, type of structure, and materials used [Eurostat, 2013a]. Additionally, the infrequency of transactions adds to the difficulty of developing a good index. A variety of different methodologies have been developed to cope with these challenges. What follows is a description of the three prevailing methodologies for constructing house price indices: the hedonic regression method, the repeated sales method, and the appraisal-based method.

3.1 Hedonic Regression Methods

Hedonic regression methods are based on the assumption that there is a correlation between the market value of a property and some specific utility-bearing attributes of that property. In other words, the purpose of the hedonic regression method is to explain the observed market price by the sum of the structural and locational elements that can be implicitly priced [Jiang et al., 2014]. This type of method uses regression techniques to make a "quality adjusted price" by controlling for characteristics that create heterogeneity [Eurostat, 2013a].

Eurostat states that this model is the most efficient method for making use of all available data. However, the methodology is not straight forward, and requires a high degree of subjectivity, which can lead to model specification bias [Jiang et al., 2014]. Economist Robert Shiller argues that the hedonic approach also can lead to spurious regression effects [Shiller, 2008]. It is also a fairly data-intensive method and thus relatively expensive to implement. Additionally, if price trends differ significantly across regions, locational characteristics can be especially challenging to model.

3.2 Repeat Sales Methods

The Repeat Sales method was first developed by Bailey et al. (1963). It uses dwellings that have been sold at least two times during a particular period, and compares prices paid for the same property over time. This matched-property procedure ensures that the dwellings compared are similar (assuming only minor improvements and deteriorations). This methodology is easy to use, inexpensive, and relatively objective. The only characteristics required are the price, address, and purchase date.

However, there are some significant drawbacks to this type of methodology. First, improvements and deteriorations of the dwelling in-between transactions are not taken into consideration. Next, the methodology can only use dwellings that have been sold at least two times within the sample period. Thus, a significant portion of the housing market is excluded from these types of models. This problem may lead to sample selection bias, as some dwellings may be traded more frequently than others, leading to some types being over-represented in the sample [Eurostat, 2013a]. In particular, houses in the lower price range tend to be involved in a higher number of transactions. Thus, an inexpensive house is expected to be traded on the market more frequently than an expensive house. Despite this drawback to this methodology, some of the most popular house price indices are repeat sales indices, including the S&P/Case-Shiller home price index.

3.3 Appraisal-Based Methods (SPAR)

Appraisal-based methods use assessed property values. Appraisals are used in some countries for tax purposes, and can often work as a relatively good market price proxy [Eurostat, 2013a]. This method uses the assessed value to compare with a single sale price. Therefore, compared to repeat-sales models, an appraised-based methodology can include all appraised dwellings, and is not limited to only dwellings that have been sold multiple times. In other words, appraisal-based models are not subject to sample selection bias, as the repeat-sales models. Similar to repeat-sales models, no dwelling-specific characteristics are necessary except for sale prices and appraisals.

However, the methodology does have some drawbacks. Most notably, appraisal-based methods depend on the quality of the appraisals, a factor that varies significantly among countries. Additionally, this methodology can only be used in certain countries; not all countries have official government appraisals. Although significantly less widespread than the two aforementioned methodologies, the appraisal-based method is important to mention due to its popularity in Scandinavia. Both Sweden and Denmark utilize this methodology for their main house price indices. This methodology is simplistic and does not require econometric modeling.

3.4 The Norwegian House Price Index

The Norwegian house price index is a quarterly published house price index produced by Statistics Norway (SSB) [StatisticsNorway, 2016]. The index is primarily used by the Ministry of Finance and the Central Bank of Norway, as well as by researchers. The index consists of three sub-indices for different housing classes as well as 11 sub-indices for different regions, totaling 33 sub-indices. These indices are then aggregated to a total House Price Index. All sub-indices are created using the hedonic methodology. Researchers on the Norwegian housing market generally seem to favor this index over other alternatives. The other index that could be considered is the house price index created by the Norwegian Association of Real estate agents (NEF). However, besides NEF's index having a monthly frequency, the two indices are based on the same assumptions and methodologies. The main difference in price development is assumed to come from a slightly different weighing procedure among the sub-indices. Even though a monthly frequency could have advantages as it involves more information, it is not likely that explanatory variables utilized in this thesis will be available at a similar frequency. This paper utilizes SSB's quarterly house price index as the dependent variable.

CHAPTER 4: LITERATURE REVIEW

Research on house prices has understandably been a relatively hot topic for Norwegian economists in recent years. Two projects stand out, namely Statistics Norway's MODAG (KVARTS) models and Jacobsen and Naug's house price models (2004). Both of these projects will be discussed in turn, as their approaches and findings will be of assistance in the further development of this thesis.

4.1 MODAG and KVARTS

MODAG and KVARTS are the core macroeconomic models utilized by the Norwegian Ministry of Finance for matters requiring analyses and forecasts for the Norwegian economy [StatisticsNorway, 2017a]. The models are comprised of a number of complex and specific sub-models, including models dealing with the development of real house prices. MODAG and KVARTS are developed by Statistics Norway, which is Norway's central bureau of official statistics [StatisticsNorway, 2017b]. The main difference between MODAG and KVARTS is that the former relies on yearly data, while the latter relies on quarterly data [StatisticsNorway, 2017a]. As this study uses quarterly data in its analysis, KVARTS is the primary focus. However, as both models are built on the same fundamentals, references to KVARTS and MONDAG will be used interchangeably. Figure 4.1 is a reproduction of variable relationships according to the KVARTS model [Boug and Dyvi, 2009]. In KVARTS, the three main determinants of real house prices are household's disposable real income, the real interest rate after tax, and building costs. The effect from building costs can be viewed as an indirect effect, as it first is seen affecting housing starts, which affects the aggregate housing stock. The housing stock, which is the maximum supply of houses in the marketplace, helps determine the market clearing house price, as explained in Chapter 2.

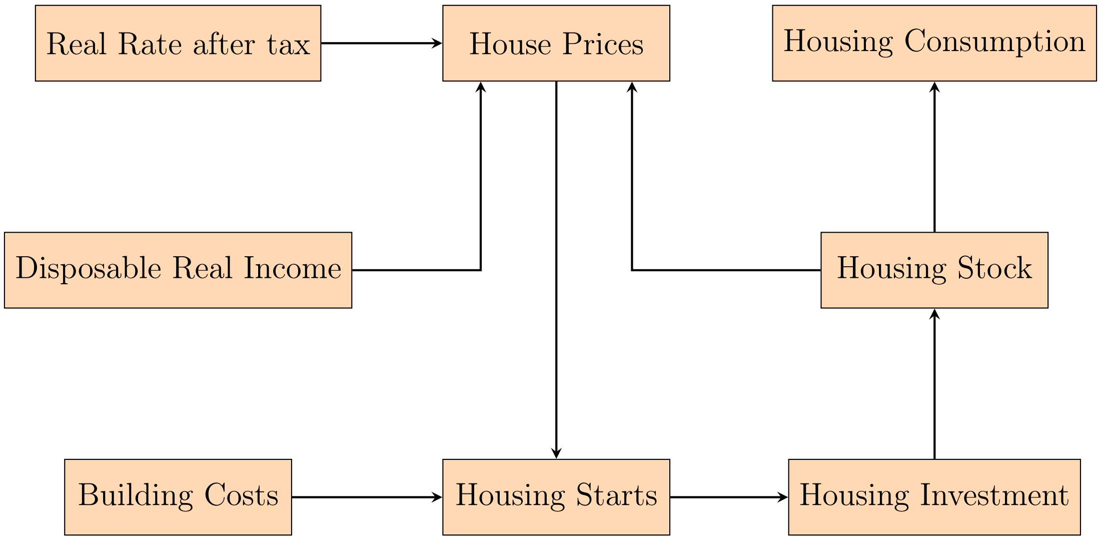

FIGURE 4.1: MODAG/KVARTS HOUSING MODEL

Demand for dwellings depends on the real house price, the household's disposable real income, and the real interest rate after tax. Supply of dwellings is given by the existing stock, which is assumed to be fixed in the short term, but depends on new investments and depreciation in the long term. Investments depend on the relationship between real house prices and the building costs.

In KVARTS, house prices are modelled from the demand side, whereas changes in the housing stock comes from the supply side. Therefore, a given house stock is used when determining the market clearing house price. KVARTS explains demand for housing by the following expression:

$$K = K^E(\underset{-}{P_k}, \underset{+}{Y}, \underset{-}{r}). \tag{4.1}$$

Equation 4.1 models demand for housing as dependent on the real disposable income for households (Y), and the user-price for a dwelling. The user-price, which is how much it actually costs to use a dwelling for a given amount of time, is determined by the house price (P_k), the real interest rate after tax (r), and depreciation of the dwelling. Statistics Norway excludes depreciation of the dwelling in Equation 4.1 to simplify the expression. A positive shock to either P_k or r is expected to have a negative impact on demand, whereas a positive shock to Y is expected to have a positive impact on demand.

In the short run, the housing stock is assumed to be fixed, and Equation 4.1 can be inverted to explain the market clearing house price:

$$P_k = P_k(\underset{-}{K}, \underset{+}{Y}, \underset{-}{r}). \tag{4.2}$$

As we assume that the housing stock is given in the short run, house prices will mainly be affected by movements in real disposable income and the real interest rate after tax:

$$P_k = P_k(\underset{+}{Y}, \underset{-}{r}; \underset{-}{K}). \tag{4.3}$$

A positive shock to real disposable income for households is expected to have a positive impact on house prices in Equation 4.2 and 4.3, while a positive shock to the real interest rate after tax is expected to have a negative impact on house prices. Finally, a positive shock to the given housing stock is expected to have a negative impact on house prices.

4.2 Jacobsen and Naug

Another important study is Jacobsen and Naug (2004). In their article, the authors derived an econometric regresson model with the objective of explaining Norwegian house prices. The aim of the project was to understand whether the tripling in house prices that had occurred in the period 1992-2004 could be explained by fundamental factors.

Jacobsen and Naug's work follows many of the same fundamental assumptions as the MONDAG project. House prices are determined by demand and supply. Supply, measured by the housing stock, is relatively stable in the short run, so that house prices are determined by demand in the short term. However, in the long run, a measure of the housing stock should be included in the model. Jacobsen and Naug further explain there are two different demands for dwellings: one for living purposes and one for investment purposes. However, they clearly state that the former is far larger in the Norwegian housing market, and is thus what should primarily be considered. Jacobsen and Naug utilize the house price index developed by Norwegian Association of Real estate agents (NEF), and test the following variables in their analysis:

- Household's Nominal Salary
- Index for Paid Rent and Collected Rent in CPI

- Different measures of the real interest rate after taxes
- Housing Stock (as it is measured in the national accounts)
- Unemployment Rate
- Backdated development in House Prices
- Household debt
- Different population measures
- Measures of movement and centralizing (population)
- Consumer confidence index (households' expectations to the country's economy)

The proxies used for house rents were all insignificant. The authors explain these results as an effect of the large amount of housing cooperatives in Norway, where rent is not often adjusted. The authors also did not find significant effects of household debt on house prices. Finally, relocation and other demographic relations had no significant effect on house prices in the sample period. Nevertheless, it is argued that demographic changes have an indirect effect through its effects on salaries.

Possibly Jacobsen and Naug's most notable finding is that models were significantly better when using measures of nominal interest rates as an explanatory variable rather than the real interest rate. Following this discovery, they tried including nominal interest rates and inflation as two separate variables. This resulted in inflation becoming insignificant, and with a poor directional accuracy. As a result of these findings, Jacobsen and Naug's model is based on nominal house prices and nominal interest rates. After testing a variety of different interest rates in their models, they find the banks' lending rate the most appropriate measure. This rate is significant in all models estimated.

Aggregate salaries in the economy had a significant effect on house prices. Statistics Norway utilizes household income as their proxy of income in their model MONDAG, whereas Jacobsen and Naug apply labor income, or salaries, as the explanatory variable. Jacobsen and Naug justify this choice by arguing that tax-related volatility in stock dividends have had a significant effect on the development of household disposable income in the later years of their study period, and that these fluctuations did not have a strong effect on household demand for housing. The housing stock and the unemployment rate both had significant negative effects on housing prices.

Finally, TNS Gallup's indicator of household expectations about the country's economy is considered. The purpose of this measure is to predict Norwegian households' economic behaviour [Stäubert, 2017]. This indicator proved to be highly significant. However, Jacobsen and Naug found that the indicator was correlated with both the interest rate and the unemployment rate. Thus, the utilized consumer confidence indicator has controlled for these effects.

Jacobsen and Naug's final model included the following explanatory variables:

House Price: Price index for used dwellings (Source: FINN.no)

Interest Rate: Bank average lending rate (Source: Central Bank of Norway)

Unemployment Rate: Norwegian unemployment rate (Source: ATEA)

Salary: Aggregate nominal salary-income in the economy (Source: SSB)

Housing Stock: Housing stock measured in fixed prices (Source: SSB)

Consumer Confidence: Indicator of households' expectations to the country's economy (Source: TNS Gallup)

CHAPTER 5: VARIABLE SELECTION

The purpose of this chapter is to present all variables that will be considered as determinants of Norwegian house prices. The choices of variables are primarily based on the literature review performed in the preceding chapter. Arguments will also be presented to defend the choices of proxies for the selected determinants. Although the sources of most data used is expressed for the individual variables when discussed, Appendix 3 provides a full list of data-sources, as well as links for obtaining the data.

5.1 Interest rate

A few considerations had to be made in order to find the most appropriate proxy of interest rates. The first question is whether to use nominal or real rates. Jacobsen and Naug (2004) consider this particular problem, and find that models that use nominal rates usually perform better. This predicament leads to whether inflation then should be included as an independent variable. Jacobsen and Naug find that inflation (CPI) is not significantly helpful as a predictive variable of house prices. Nevertheless, CPI will be included as a variable for further testing. On the basis of these findings, nominal values of both interest rates and income will be used in this study.

There are a few available data-sets that can be used as a proxy for the interest rate. The following two different interest rates are further tested in section

6: mortgage companies' average lending rate to households and banks' average lending rate to all. Both variables are found in Statistics Norway's StatBank. It is expected that the interest rate will have a negative effect on house prices, as higher interest rates mean more expensive borrowing.

5.2 Unemployment Rate

Data on the unemployment rate is another contribution from Statistics Norway. The data-set is reported to follow the standards set by the International Labor Organization (ILO). Additionally, labor force statistics are required to follow the standards set by the European Union's Statistical Cooperation agreement (Under the EEA agreement) [Andersen, 2015]. It is expected that the unemployment rate will have a negative effect on house prices, as an increase in the unemployment rate would lead to expectations of lower growth in salaries. A higher unemployment rate is expected to increase household uncertainty about the future of their own and others solvency, which in turn is expected to decrease household willingness to pay for housing [Jacobsen and Naug, 2004].

5.3 Income and Salary

Both household disposable income and wage will be evaluated as potential variables to include in the final model. Wage is an aggregated index of wage indices from all industries in the economy, whereas income is a seasonally adjusted income . The data-sets can both be found in Statistics Norway's StatBank. The MONDAG project uses disposable real income as an explanatory variable while Jacobsen and Naug uses nominal labor income. Jacobsen and Naug report that in their sample period, tax-motivated fluctuations in dividends had important consequences for the measured trends in household disposable income. However, these fluctuations were not assumed by the authors to have much effect on household housing demand. Thus, they decided to utilize labor income as their explanatory variable instead. Which one will be utilized in this paper depends on the test results in Section 6. Both data-sets are stated in nominal terms. Household income is assumed to have a positive effect on house prices.

5.4 Expectations Barometer

An indicator of households' expectations about the country's economy will also be considered. A quarterly survey developed by Kantar TNS on the behalf of Finance Norway measures Norwegian Households' expectations to their own and their country's economy [Stäubert, 2017]. Although Jacobsen and Naug find this indicator to be highly significant, they also reported that the indicator was highly correlated with both the interest rate and the unemployment rate. Thus, a multivariate Granger-causality test should be performed to check whether this variable has any additional predictable power in a model with an interest rate measure and the unemployment rate already included. Although it would be possible to develop an adjusted measure controlling for the effects of interest rates and the unemployment rate, this has not been prioritized.

5.5 Population Measures and Net Migration

The two variables used to measure population effects on house prices are Statistics Norway's variables on population and net migration. It is assumed that growth in population measures will have positive effects on house prices. The MONDAG model does not report any use of such a variable. Jacobsen and Naug note that they test a number of demographic variables in their models, but that none had direct effects on house prices.

5.6 Housing Starts

Housing starts are considered the only variable in the model that should affect house prices from the supply-side of the economy. This will be done through an increase in the housing stock. A number of other variables of the housing supply were considered, such as the housing stock, housing investment, and building costs. However, these variables were not available in the quarterly frequency required. In MONDAG, it is explained that they use housing starts instead of other variables such as the housing stock and investments, as housing starts are closer in time to the actual decision-making point. Figure 4.1 shows the relationship postulated by the MONDAG model. The total effect of housing starts on house prices is uncertain because of a bidirectional relationship between house prices and housing starts. Higher house prices means increased profitability in the industry, thus more housing starts. However, more housing starts also mean an increase in supply, thus lower prices.

Although housing starts is the only variable tested from the supply-side, a decision was made on the basis of a pure data-mining procedure (no economic theory) to also split up housing starts in two new variables: maximum two living units and minimum three living units. All three variables will be individually and collectively tested to see which one does a better job at explaining prices. Nevertheless, as Jacobsen and Naug (2004) explain, there are two different demands for dwellings: one for living purposes and one for investment purpose. It can be argued that buildings with minimum three living units correlates stronger with investment purposes, whereas buildings with maximum two living units are more affected by demand fueled by living purposes. The two variables are illustrated below

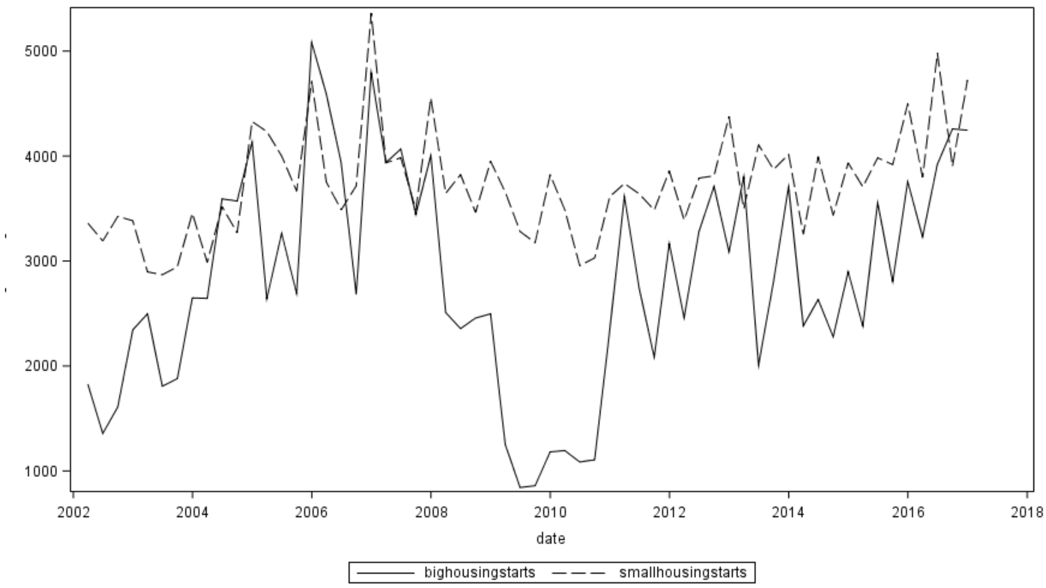

Figure 5.1: HOUSING STARTS

5.7 Debt

There are varying results whether household debt levels Granger-cause house prices. Oikarinen (2009), Fitzoatrick and Mcquinn (2007) and Berlinghieri (2010) all suggest that debt levels Granger-cause house prices. Anundsen and Jansen (2013) contributes to this research by suggesting that there are self-reinforcing effects between housing prices and household borrowing in Norway. On the other hand, Jacobsen and Naug (2004) suggests that debt does not Granger-cause house prices in their sample. Here, the credit indicator K2 developed by Statistics Norway will be tested for effects on house prices. K2 can further be split to either measure household gross domestic debt or general public gross domestic debt. Both variables will be researched.

5.8 Variable Summary Statistics

Table 5.1 provides the summary statistics for all variables discussed in this section. As can be seen, there are 60 quarterly observations covering 15 years of data. The summary statistics are given in level form.

Table 5.1: SUMMARY STATISTICS (2002:1 - 2016:4)

Variable	N	MEAN	STD	MIN	MAX
House Price Index	60	135.022	35.364	81.4	199.3
Banks' Average Lending Rate to All	60	5.033	1.386	3.46	8.7
Mortgage Comp. Lending Rate (House)	60	4.480	1.444	2.42	7.83
Unemployment Rate	60	3.692	0.737	2.1	4.9
CPI	60	1.873	1.068	-1.4	4.71
Expectation Barometer	60	116.132	13.104	84.6	131.167
Income	60	238,759.6	53,605.98	155,607	333,463
Salary	60	119.449	19.438	87.706	149.138
Credit Indicator (General Public)	60	3,294,318	1,093,236	1,632,488	5,141,122
Credit Indicator (Household)	60	1,932,349.9	633,697.090	921,832	3,066,407
Net Migration	60	7932.517	3480.142	1967	14185
Housing starts	60	6561.333	1418.137	4033	10158
Housing starts (>2u*)	60	2825.567	1032.229	842	5083
Housing starts ($\leqslant 2u*$)	60	3735.77	508.601	2869	5363

u*=living units

CHAPTER 6: STATISTICAL TESTING OF VARIABLES

This chapter will first elaborate on the concept of stationary processes, before all data series can be used for modelling purposes each will undergo the necessary transformations and tests in order to secure that they are in fact stationary. Next, the idea of Granger-causality will be elaborated and considered for the variables of interest. This will determine which variables should be further considered as determinants of housing prices in the final model.

6.1 Stationary Processes

One of the most important considerations in any time series modelling is to properly determine whether a data series has a unit root. If a data series does not have a unit root, the time series is said to be stationary, and appropriate for modelling purposes. If the data series does have a unit root, this needs to be adressed before utilizing the time series, as utilization of non-stationary data with OLS will lead to the problem of spurious regression [Silvia et al., 2014].

"A stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed" [Gujarati, 2004]. It should be noted that this definition only holds true for weakly stationary processes. However, due to the practical approach of this thesis, it is regarded as sufficient. Mathematically, the definition provided can be expressed by the three following properties:

$$Mean : E(Y_t) = \mu \tag{6.1}$$

$$Variance : var(Y_t) = \sigma^2 \tag{6.2}$$

$$Covariance : \gamma_k = E(Y_t - \mu)(Y_{t+k} - \mu)) \tag{6.3}$$

where Y_t is a stochastic time series and k is the number of periods after period t.

If these statistical properties vary over time, the time series is said to be non-stationary. If this is so, it means that the behavior of the time series in the sample period is not generalizable for time periods outside the sample [Gujarati, 2004]. In other words, the time series will not be useful for econometric time series modelling. To solve the problem of non-stationary processes, it is possible to transform the data to obtain stationary processes.

6.1.1 Data Transformation

Transformation of data is often necessary as many economic time series in level form are non-stationary. In fact, not a single variable tested in this paper was stationary in its level form. The chosen data transformation for all variables utilized in this study is the first difference of logarithms. There are mainly two reasons why a logarithmic data transformation is preferable [Stock and Watson, 2011]. First, many of the data series utilized have an exponential increasing trend, which is very common in economic time series. However, the percentage growth is often more stable over time, which makes the logarithmic growth nearly linear.

Secondly, Stock and Watson state that "the standard deviation of many series is usually approximately proportional to its level." This would indicate that a variable in its level form will have an increasing variance as it grows over time. In its logarithmic form however, the variance should be close to constant. Finally, the reason why this paper utilizes the first difference of the logarithmic transformation is to remove the trend component from the data sets, thus obtaining a constant mean.

6.1.2 Unit Root tests

The classic example of a non-stationary time series is a random walk. This theory suggests that past values cannot be used to predict future values. Imagine a time series as the pathway of a wild chicken. You can trace its steps, but it is highly unlikely that you will be able to predict where it will go next. A random walk process may come in three distinct forms: a random walk (RW), a random walk with drift (RWD), and a random walk with drift around a stochastic trend (RWDT) [Gujarati, 2004]. The difference between a RW and a RWD is that the former does not have a constant term, which means that the mean is not significantly different from zero, whereas the latter has a mean that is significantly different from zero. A RWDT process means that the process additionally has a trend component that changes with time. These three processes can mathematically be expressed as:

$$RW : \Delta Y_t = \delta Y_{t-1} + u_t \tag{6.4}$$

$$RWD : \Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t \tag{6.5}$$

$$RWDT : \Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t, \tag{6.6}$$

where t equals time and u_t is a white noise error term.

All these forms need to be tested for unit roots, thus, the original unit root test developed by Dickey and Fuller in 1979 is in reality three separate tests. The null hypothesis (H_0) for each of these tests is that $\delta = 0$, meaning that the time series is non-stationary. The alternative hypothesis (H_1) is that $\delta < 0$, meaning that the time series is stationary. It can be seen that if $\delta = 0$, Equation 6.4 will simply become $\Delta Y_t = u_t$, which means that the difference between Y_t and Y_{t+1} is determined solely by the error term (thus a random walk). For Equation 6.5, the only difference is that the mean is nonzero. Equation 6.6 will consist of a nonzero mean as well as a time-dependent trend variable.

Since the original DF test was developed in 1979, a number of more advanced tests have been developed to cope with different limitations of the original test. As the original test assumed u_t to be uncorrelated, and this may not always be the case, the Augmented Dickey-Fuller test (ADF) was developed to handle this issue [Gujarati, 2004]. In 1996, Elliott, Rothenberg, and Stock developed the Dickey-Fuller Generalized Least Squares test (DF-GLS), a modified version of the ADF test that tests for more than one unit root [Elliott et al., 1996]. Both the ADF and the DF-GLS tests are performed on the variables described above. However, where there are contradicting results, the DF-GLS test are preferred, as this test is assumed to have significantly better statistical properties. According to one source, the probability of avoiding the situation of rejecting a true (H_0) is 0.75 for the DF-GLS test compared to only 0.31 for the ADF test [Iqbal, 2016]. The reason why both tests are performed is partly to illustrate the difference in modelling choices that could have occurred if the study only relied on the ADF test, as some other researchers on similar topics have done. The critical values of the tau statistics are different for the three processes. Table 6.1 therefore provides a * for each statistic that allows us to reject the null hypothesis with 95 percent confidence.

According to the ADF test the variables income, housing starts (all) and housing starts (max two living units) are stationary in their logarithmic form. This is not the case when testing the variables with the DF-GLS test. Relying only on the ADF test would thus introduce spurious regression. However, the DF-GLS test indicates that both the unemployment rate and the expectations barometer are stationary in their logarithmic form. When applying the first difference of logarithms, all but one variable are stationary. Although the ADF test finds that both credit indicators are stationary in the first difference form, the DF-GLS test does not. Therefore, the second difference form is utilized for these variables. The DF-GLS test for the second difference form suggests that only the household credit indicator is stationary. Thus, this is the only indicator of debt this study will proceed with.

Table 6.1: UNIT ROOT TESTS

Variable (All variables are in natural logarithms)	ADF Test (Tau statistic)			DF-GLS Test	
	Zero Mean	Single Mean	Trend	Single Mean	Trend
House Prices	-	-	-2.131	-	-2.642
Interest Rate (Average lending rate from banks)	-	-1.812	-	-1.253	-
Interest Rate (Average mortgage rate for households)	-	-0.904	-	-0.931	-
Inflation (Consumer Price Index)	-	-	-2.759	-	-2.913
Unemployment Rate	-	-1.901	-	-2.554*	-
Salary	-	-	-1.845	-	0.324
Income	-	-	-6.541*	-	-1.290
Expectations Barometer	-	-1.321	-	-2.644*	-
Net Migration	-	-2.380	-2.395	-1.404	-1.590
Housing Starts (All)	-	-3.368*	-	-1.661	-
Housing Starts (max two living units)	-	-5.161*	-	-1.865	-
Housing Starts (min three living units)	-	-2.816	-	-1.303	-
Credit Indicator (General Public)	-	-	0.51	-	-1.512
Credit Indicator (Household)	-	-	0.887		-1.154
ΔHouse Prices	-	-6.433*	-	-3.146*	-
ΔInterest Rate (Average lending rate from banks)	-3.363*	-	-	-3.407*	-
ΔInterest Rate (Average mortgage rate for households)	-3.809*	-	-	-3.114*	-
ΔInflation (Consumer Price Index)	-	-8.983*	-	-4.1078*	-
ΔUnemployment Rate	-8.112*	-	-	-1.978*	-
ΔSalary	-	-8.906*	-	-2.943*	-
ΔIncome	-	-15.037*	-	-3.580*	-
ΔExpectations Barometer	-2.925*	-	-	-3.432*	-
ΔNet Migration	-13.821*	-	-	-2.011*	-
ΔHousing Starts (All)	-12.037*	-	-	-2.022*	-
ΔHousing Starts (max two living units)	-14.615*	-	-	-3.295*	-
ΔHousing Starts (min three living units)	-9.704*	-	-	-2.151*	-
ΔCredit Indicator (General Public)	-	-3.4679*	-	-1.2132	-
ΔCredit Indicator (Household)	-	-4.2462*	-	-0.4629	-
2ΔCredit Indicator (General Public)	-15.533*	-	-	-1.7180	-
2ΔCredit Indicator (Household)	-15.853*	-	-	-3.5296*	-

* = Significant at the 0.05 level

2Δ = second difference

6.2 Granger Causality

6.2.1 Causation vs Correlation

Causation and correlation tend to be used interchangeably in daily conversation. However, when dealing with variable relationships in econometrics, the two terms differ significantly, and it is paramount to distinguish between them. To exemplify the difference, consider the correlation found between smoking and stress levels [Srivastava, 2015]. A careless researcher may interpret this finding as "smoking cigarettes leads to an increased level of mental stress" although it can be regarded just as likely that higher stress levels can influence a person to smoke, or even that there exists a third factor that causes both smoking and stress levels.

Finally, it may actually be that the two variables are coincidentally correlated. For example, a study reported that the variables "swimming pool drownings" and "amount of movies Nicolas Cage appeared in" has a correlation coefficient of 0.66 [Goldman, 2014]. Even though it might be tempting to blame Nicolas Cage for the 102 swimming pool drownings that occurred during 2009, it would clearly be a mistake. Thus, even though there is a proven correlation between the two variables, the existence and direction of causality is uncertain and requires further analysis.

6.2.2 Granger Causality Test

The article "Investigating Causal Relations by Econometric Models and Cross-spectral Methods" by Clive Granger (1969) first proposed a statistical method for handling such problems [Granger, 1969]. The Granger Causality Test dealing with bilateral causality for variables A and B involves estimating the following two regressions:

$$A_t = \sum_{i=1}^{n} \alpha_i B_{t-i} + \sum_{j=1}^{n} \beta_j A_{t-j} + u_{1t} \tag{6.7}$$

$$B_t = \sum_{i=1}^{n} \gamma_i B_{t-i} + \sum_{j=1}^{n} \delta_j A_{t-j} + u_{2t}. \tag{6.8}$$

Equation 6.7 assumes that A_t depends on past values of itself as well as past values of B, while equation 6.8 assumes that B_t depends on past values of itself as well as past values of A [Gujarati, 2004]. The procedure of the test goes as follows: First, regression 6.7 is carried out only with its own lagged variables (without lags of B_t), where the residual sum of squares (RSS_R) is obtained. Next, the regression is carried out with the lagged values of variable B included, and RSS_{UR} is obtained. Finally, an F-test is carried out based on RSS_{UR} and RSS_R. H_0 is that the lagged terms of variable B do not belong in the regression. Thus, if the F-statistic exceeds the critical value, H_0 can be rejected and we can say that B Granger-causes A.

Fortunately, powerful statistical tools such as SAS exists today that can easily obtain information on Granger causality without having to manually perform the steps described above. Nevertheless, it is important to understand the fundamental assumptions of the Granger causality test to properly analyze the results generated.

Table 6.2 presents the one-by-one Granger-causality test between the dependent variable and each of the the potential determinants. Each variable is tested to see if the variable individually Granger cause house prices. Thus, the table only provides the one-way Granger causality test results. The appropriate number of lags for each Granger-causality test is chosen by the lowest value of the Bayesian information criterion (SBC), following standard practise [Gujarati, 2004]. SBC was tested for each variable with up to 5 lags. The lags required here should correlate with the lags required in the final Vector Autoregressive model in section 8.2.

Table 6.2: GRANGER CAUSALITY

Variable	Lags/DF**	Chi-Square	Pr >Chi-Sq
Interest rate (Average lending rate from banks)	2	3.47	0.1764
Interest rate (Average mortgage rate for households)	1	3.94	0.0473*
Unemployment rate	2	20.26	<.0001*
Salary	1	5.82	0.0158*
Income	2	2.8	0.2469
Expectations barometer	2	9.74	0.0077*
Net Migration	2	11.73	0.0028*
Housing starts (All)	2	21.78	<.0001*
Housing starts (max two living units)	3	25.52	<.0001*
Housing starts (min three living units)	2	11.38	0.0034*
Credit Indicator (Household)	2	1.71	0.4263

*= Variable Granger-cause house prices at the 95 percent confidence level

** Number of lags are chosen based on lowest SBC

CHAPTER 7: APPROACHES TO MODELING

According to Gujarati (2004) there are five approaches to econometric modelling based on time series data: Exponential Smoothing methods, single-equation regression models, simultaneous-equation regression models, autoregressive integrated moving average models (ARIMA), and vector autoregression models (VARs). First, the alternative methods not utilized in this thesis will briefly be elaborated on for the purpose of further comparisons. Next, the Box-Jenkins (ARIMA) approach will be explained, as this framework will be utilized as a benchmark to evaluate the standard of the final model. Finally, the main modeling framework of interest, Vector Autoregressions (VAR), will be explained in detail.

7.1 Alternatives

Exponential smoothing methods consist of fitting curves through historical data. This is assumed to be the most simplistic approach with the least predictive power. Today, this methodology is primarily used as a supplementary technique to other more advanced forecasting approaches. Single-equation regression models are arguably the most common for research purposes similar to this one, and is the methodology utilized by Jacobsen and Naug in their study of the Norwegian housing market. Simultaneous-equation models are probably the most complicated framework for time series modeling. They were widely popular amongst forecasters and policy makers during the 1970s, but they often prove to be unreliable.

7.2 Autoregressive Integrated Moving Average (ARIMA)

When the ARIMA methodology was introduced by Box and Jenkins in 1970, it represented a significant shift from the complicated single-equation and simultaneous-equation frameworks popular at the time. ARIMA models rely solely on the probabilistic properties of the data rather than on economic theory, and is thus said to let the data speak for themselves [Gujarati, 2004]. Instead of explanatory variables as regressors, the ARIMA methodology uses past values of the dependent variable in addition to stochastic error terms.

An ARIMA model is in theory a combination of two types of statistical models: an autoregressive model and a moving average model [Gujarati, 2004]. An autoregressive model assumes that the value of a time series at time t solely depends on a portion of the value of the same time series at time $t-1$, plus an error term. Thus, a p-order autoregressive model, or an AR(p) model, is a model where the observation at time t depends on the values of the data in the periods p steps back, as well as a random error term [Gujarati, 2004]. A moving average model on the other hand assumes that the value of a time series at time t depends on a constant, as well as a moving average of the current and past error terms. Thus, in a MA(q) process, the value at time t depends on the error terms in the periods up to p lags back. A time series may have both autoregressive and moving average characteristics, which then requires an ARMA(p,q) model. Equation 7.1 illustrates a simple ARMA(1,1) model:

$$Y_t = \underbrace{\overbrace{\mu}^{\text{Constant}} + \overbrace{\alpha_1 Y_{t-1}}^{\text{AR(1)}} + \overbrace{\beta_0 \gamma_t + \beta_1 \gamma_{t-1}}^{\text{MA(1)}}}_{ARMA(1,1)} \tag{7.1}$$

The term integrated simply means how many times it is required to difference a time series before it turns stationary. As explained in section 6.1.1, the HPI utilized in this thesis is in the first-difference form. Thus, it can be said that the original time series will be an ARIMA(p,1,q) model.

7.3 Vector Autoregressions (VARs)

In 1980, Christopher Sims introduced Vector Autoregressions (VARs). The name can be decomposed to explain the methodology's origins, where the term vector indicate that the model consists of a vector with two or more variables, and autoregression to indicate that the methodology utilizes lagged values of the dependent variable [Gujarati, 2004]. Thus, VAR incorporates parts of both the ARIMA methodology and the simultaneous-equation methodology. However, where simultaneous-equation models require an analysis to determine which variables are endogenous and which are exogenous, VARs treat all variables as endogenous. Put differently, each variable in a VAR is explained by its own lagged values as well as lagged values of all other variables in the model.

This reasoning is very similar to the explanation of bilateral causality in section 6.2. In fact, there is only minor differences between how a simple two-variable VAR equation-set is expressed mathematically and and how the Granger-causality test was expressed in Equation 6.7 and 6.8. Equations 7.2 and 7.3, which demonstrate this point, contains k lags of A and B. The $U's$ are often called variable impulses or variable error terms.

$$A_t = \lambda + \sum_{i=1}^{n} \alpha_i B_{t-i} + \sum_{j=1}^{n} \beta_j A_{t-j} + u_{1t} \tag{7.2}$$

$$B_t = \lambda^{'} + \sum_{i=1}^{n} \gamma_i B_{t-i} + \sum_{j=1}^{n} \delta_j A_{t-j} + u_{2t} \tag{7.3}$$

7.3.1 Benefits and drawbacks of the VAR methodology

The primary reason why the VAR framework has become so popular in later years is due to the methodology's simplicity compared to simultaneous-equation models [Gujarati, 2004]. The fact that there is no need to distinguish between endogenous and exogenous variables adds to this. Although relatively simplistic, forecasts generally prove to be of better quality than those stemming from complicated simultaneous equation models.

The most notable drawback of the VAR methodology is the curse of dimensionality [Silvia et al., 2014]. A five variable VAR-model with four lags would mean having 20 parameters in each equation in addition to the constant term. This will require a high number of degrees of freedom, which could cause problems depending on the sample size.

CHAPTER 8: MODEL ESTIMATION

8.1 Considerations For Modeling

The number of variables and lags to include in the model will determine the quality of the model. If too many variables and/or lags are included, the researcher exposes himself to the problem of multicollinearity. If too few are included, the model may experience specification errors [Gujarati, 2004]. Consequently, the main question becomes how to decide the appropriate number of variables and lags to include. In this chapter, a number of different measures will be presented to assist the final choice of model specification, before the final model is estimated.

8.1.1 T-statistic vs F-statistic

To start answering these questions, it is tempting to follow the popular t-statistic approach often utilized when building single-equation regression models. The general idea is that variables (and lags of variables) can be included and excluded from the model on the basis of the variables' t-statistics. The t-statistic determines whether the specific variable of interest is a meaningful addition to the model or not. However, as VAR models usually include multiple lags of the same variables, many of these will be correlated with each other, consequently affecting the individual t-statistics. As a result, most economists suggests completely disregarding this statistical measure in VAR modeling and rather focus on collective

F-statistics, as it is done in the Granger-causality test [Gujarati, 2004]. Put differently, the predictive power of all lags included in the model of a single variable will be the main focus rather than the predictive power of each lag individually.

8.1.2 Bayesian information criterion (SBC)

Numerous model selection criteria that have been developed to assist the selection process among competing models, with the most prominent being the R^2, adjusted R^2, Akaike's Inforamtion Criterion (AIC) and Schwartz's Information Criterion (SIC/SBC). What all these criteria have in common is that they aim at minimizing the residual sum of squares (RSS). Stock and Watsson (2001) suggest avoiding R^2 completely due to the complicated dynamics in the VAR. The difference between AIC and SBC is that SBC penalizes harder for including additional variables [Gujarati, 2004]. SBC is therefore the preferred model selection criteria in this study. The SBC is defined as

$$SBC = n^{k/n} \frac{\sum \hat{u}^2}{n} = n^{k/n} \frac{RSS}{n}, \tag{8.1}$$

where n is the number of observations and k is the number of regressors including the intercept. The lower the value of SBC, the better the model.

8.1.3 Root-mean-square-error (RMSE)

Next, the models with the lowest SBC values are further considered by the RMSE criteria. Specifically, the performance is measured by the simulated real time out-of-sample root mean square error (RMSE) criteria, following the procedure of Vitner and Iqbal (2010). First, all models considered utilize the data set from 2002:Q1 to 2012:Q4 and forecast the next 8 quarters (2013:Q1 - 2014:Q4). Then, all models will include one more data point and redo the procedure. This recursive series of steps is performed until 2015:Q1 to 2016:Q4 is forecasted based on the data set from 2002:Q1 to 2014:Q4, at which time there are no more real values to compare against. For each step, the out-of-sample RMSE is calculated using the equation

$$RMSE = \sqrt{\frac{1}{8}\sum(Y_{t+h} - \hat{Y}_{t+h})^2}, \tag{8.2}$$

where $\hat{Y}_{t+h}$ is the h-step ahead forecasted value of Y_t based on the models estimated using data up until period t. Y_{t+h} is the real value of Y_t h steps ahead. Table 8.1 reports the overall average for each model based on this procedure. It is important to note that the RMSE measures are based on level data. In other words, the forecasted values, which are in first difference of logarithms, are transformed back into level values before used for RMSE calculations. The calculations can be seen in Appendix 3. Similar to Vitner and Iqbal (2010), an ARIMA(1,1,1) model is also estimated so that RMSE performance can be measured against a univariate model alternative.

8.2 Results and Final Model Specification

Table 8.1 presents the SBC and RMSE for eight different model specifications. It is important to consider both these criteria, as one represents out-of-sample fit and one in-sample fit. Moreover, SBC measures the overall goodness of fit for the entire equation system. As a result, a more favorable SBC from one model specification to the next does not necessarily indicate that one model is better at predicting house prices, as the reason also can be that the other variables included in the equation set does a better job at predicting each other.

Table 8.1: MODEL SPECIFICATIONS

Model	Model identifier	RMSE	SBC	Variables
VAR(2)	1	4.678	-1392.04	IR2, UR, Salary, HS
VAR(2)	2	4.638	-1453.18	IR2, UR, Salary, SHS, BHS
VAR(2)	3	4.736	-1728.2	IR2, UR, Salary, SHS, BHS, EXP
VAR(2)	4	5.135	-1470.22	IR2, UR, Salary, SHS, BHS, POP
VAR(3)	5	6.063	-1370.51	IR2, UR, Salary, SHS, BHS
VAR(2)	6	5.085	-1467.68	IR1, UR, Salary, SHS, BHS
VAR(1)	7	4.890	-1500.53	IR2, UR, Salary, SHS, BHS
ARIMA(1,1,1)	8	6.185	-256.183	-

*IR1 = Average lending rates from banks // IR2 = Average mortgage rate for households
**HS=Housing starts//BHS=min three living units//SHS=max two living units

Models 1 and 2 are estimated primarily to test for differences in using the combined and and broken up variables of housing starts, as discussed in chapter 5. Model 1 uses the aggregate housing starts in the economy while model 2 uses the variables maximum two living units and minimum three living units. Both the SBC and the RMSE indicate that model 2 is better. Consequently, models 2-6 is estimated using the two broken up measures of housing starts.

All models except for model 6 uses the household average mortgage rate as the proxy for interest rates. Initially, this choice was based on the Granger-causality test results from chapter 6 (See Table 6.2). Furthermore, models 2 and 6 are estimated to investigate if this choice was correct. Although model 6 has a marginally better SBC, RMSE is significantly better for model 2. Therefore, model 2 is favoured over model 6.

Model 3 is estimated using the same assumptions as model 2, but with the the expectations barometer included. This model provide a better SBC criterion compared to model 2. However, model 3 performs worse on the basis of RMSE. Jacobsen and Naug (2004) found that this variable was strongly correlated with both the unemployment rate and the interest rate. Therefore, a multivariate Granger-causality test was run to investigate whether the expectations barometer added predictive power to a model that already encompassed the variables unemployment rate and interest rate. Table 8.2 demonstrate that adding the expectations barometer to a multivariate model where the unemployment rate and interest rate already is incorporated will not add value, and actually cause the three variables to not collectively Granger-cause house prices at the 95 percent confidence level.

Table 8.2: MULTIVARIATE GRANGER CAUSALITY TEST

Test	DF	Chi-Square	Pr >ChiSq
UR, IR	2	6.29	0.043
UR, IR, EXP	3	7.1	0.069

Model 4 also follows the same assumptions as model 2, but includes net migration to the equation set. SBC becomes marginally better. However, the RMSE criterion tells us that model 2 is favourable.

Table 6.2 exhibit that most variables proved to individually Granger-cause house prices at the 95 percent confidence level with two lags included. This points to a VAR(2) model as the best specification. However, both a VAR(1) specification (model 7) and a VAR(3) specification (model 5) version of Model 2 are estimated to see the difference. Model 5 performed worse in every aspect compared to Model 2. However, model 7 produced a significantly lower SBC score. This is likely to be caused by the fact that the interest rate had a significantly lower SBC criterion in the Granger-causality test using one lag compared to two lags (see Table 6.2). On the other hand, model 2 does outperform model 7 on the basis of the RMSE criterion. Finally, Model 8 provides the results of the univariate model specification included in this study, an ARIMA(1,1,1). Clearly, the univariate model cannot compete with the multivariate specifications on the basis of SBC, although the RMSE criterion is surprisingly good. The only stationary variable that did not meet the criteria set out for inclusion in a model was the credit indicator, which proved not to Granger-cause house prices.

Model 7 in Table 8.1 is chosen as the best specification based on the aforementioned arguments. Model 2 was also considered, but eliminated due to model 7's superior SBC criterion and more simplistic specification. Figure 8.2 at the end of the chapter presents the final model in its entirety. The part of the equation-set that determines house prices is illustrated in Equation 8.3

$$\Delta HPI = 0.00377 + 0.23084\Delta HPI_{t-1} - 0.12396\Delta IR_{t-1} - 0.13185\Delta UR_{t-1} \\ + 0.61931\Delta SAL_{t-1} + 0.06340\Delta SHS_{t-1} + 0.01967\Delta BHS_{t-1} \quad (8.3)$$

*all variables in natural logarithms.

8.3 Model Analysis

This study finds that the best model for forecasting Norwegian house prices is a VAR(1) model with the average mortgage rate for households, the unemployment rate, salary, and finally two separate measures of housing starts (≤ 2 units and >2 units) as the included variables. Figure 8.1 presents the impulse response functions. Specifically, the figure shows impulses from the variables in the system and the individual responses they have on the house price. A positive impulse from the interest rate leads to a statistically significant negative effect on the house price. This effect is thus as suspected based on previous literature. Next, a positive shock to salary leads to a statistically significant positive effect on the house price, also corresponding well with previous research. A positive shock to the unemployment rate leads to a statistically significant negative effect on house prices, in line with previous studies. Finally, both variables representing housing starts have statistically significant positive effects on the house price. Whether the sum of effects from housing starts was positive or negative was previously unknown.

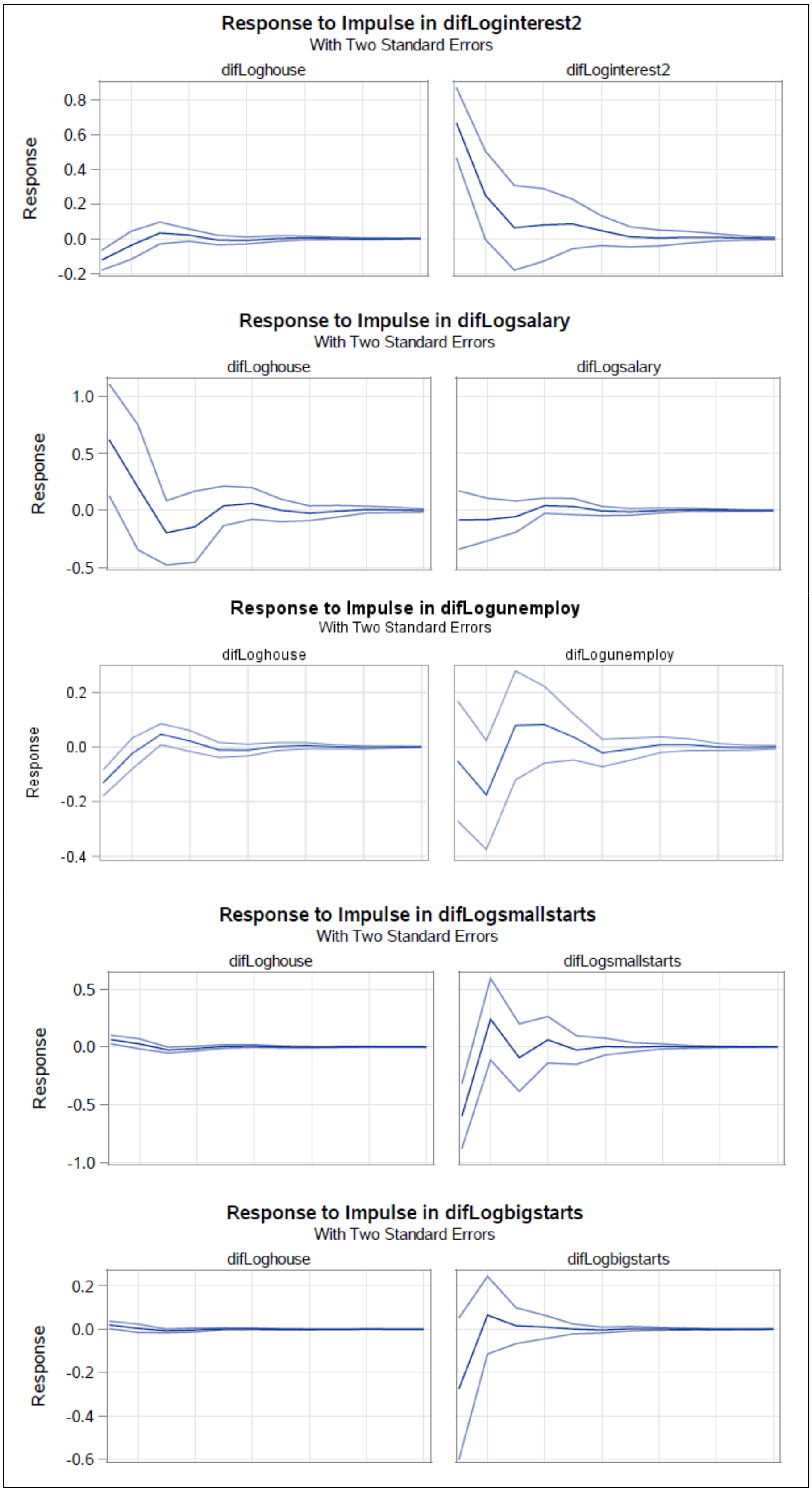

FIGURE 8.1: IMPULSE RESPONSE FUNCTIONS

Model Parameter Estimates						
Equation	Parameter	Estimate	Standard Error	t Value	Pr > |t|	Variable
difLoghouse	CONST1	0.00377	0.00334	1.13	0.2643	1
	AR1_1_1	0.23084	0.09210	2.51	0.0155	difLoghouse(t-1)
	AR1_1_2	0.01967	0.00873	2.25	0.0286	difLogbigstarts(t-1)
	AR1_1_3	0.06340	0.01823	3.48	0.0011	difLogsmallstarts(t-1)
	AR1_1_4	-0.13185	0.02384	-5.53	0.0001	difLogunemploy(t-1)
	AR1_1_5	0.61931	0.24515	2.53	0.0147	difLogsalary(t-1)
	AR1_1_6	-0.12396	0.02857	-4.34	0.0001	difLoginterest2(t-1)
difLogbigstarts	CONST2	0.02688	0.06236	0.43	0.6682	1
	AR1_2_1	0.02164	1.72165	0.01	0.9900	difLoghouse(t-1)
	AR1_2_2	-0.27497	0.16313	-1.69	0.0981	difLogbigstarts(t-1)
	AR1_2_3	0.01972	0.34078	0.06	0.9541	difLogsmallstarts(t-1)
	AR1_2_4	0.37608	0.44558	0.84	0.4027	difLogunemploy(t-1)
	AR1_2_5	1.09257	4.58260	0.24	0.8125	difLogsalary(t-1)
	AR1_2_6	0.79411	0.53404	1.49	0.1433	difLoginterest2(t-1)
difLogsmallstarts	CONST3	0.02308	0.02558	0.90	0.3712	1
	AR1_3_1	-0.77354	0.70612	-1.10	0.2786	difLoghouse(t-1)
	AR1_3_2	-0.06475	0.06691	-0.97	0.3378	difLogbigstarts(t-1)
	AR1_3_3	-0.60313	0.13977	-4.32	0.0001	difLogsmallstarts(t-1)
	AR1_3_4	0.23605	0.18275	1.29	0.2024	difLogunemploy(t-1)
	AR1_3_5	-0.00635	1.87950	-0.00	0.9973	difLogsalary(t-1)
	AR1_3_6	0.15168	0.21903	0.69	0.4918	difLoginterest2(t-1)
difLogunemploy	CONST4	-0.00714	0.01539	-0.46	0.6449	1
	AR1_4_1	0.43951	0.42498	1.03	0.3060	difLoghouse(t-1)
	AR1_4_2	-0.05911	0.04027	-1.47	0.1484	difLogbigstarts(t-1)
	AR1_4_3	-0.35066	0.08412	-4.17	0.0001	difLogsmallstarts(t-1)
	AR1_4_4	-0.05059	0.10999	-0.46	0.6476	difLogunemploy(t-1)
	AR1_4_5	-0.04889	1.13118	-0.04	0.9657	difLogsalary(t-1)
	AR1_4_6	-0.35818	0.13182	-2.72	0.0090	difLoginterest2(t-1)
difLogsalary	CONST5	0.01169	0.00174	6.72	0.0001	1
	AR1_5_1	-0.15225	0.04801	-3.17	0.0026	difLoghouse(t-1)
	AR1_5_2	0.00428	0.00455	0.94	0.3512	difLogbigstarts(t-1)
	AR1_5_3	0.01739	0.00950	1.83	0.0733	difLogsmallstarts(t-1)
	AR1_5_4	-0.02767	0.01243	-2.23	0.0305	difLogunemploy(t-1)
	AR1_5_5	-0.08371	0.12779	-0.66	0.5154	difLogsalary(t-1)
	AR1_5_6	-0.00163	0.01489	-0.11	0.9133	difLoginterest2(t-1)
difLoginterest2	CONST6	-0.02537	0.01184	-2.14	0.0370	1
	AR1_6_1	1.58523	0.32678	4.85	0.0001	difLoghouse(t-1)
	AR1_6_2	0.00904	0.03096	0.29	0.7715	difLogbigstarts(t-1)
	AR1_6_3	0.06211	0.06468	0.96	0.3416	difLogsmallstarts(t-1)
	AR1_6_4	0.04719	0.08457	0.56	0.5794	difLogunemploy(t-1)
	AR1_6_5	-0.68428	0.86982	-0.79	0.4352	difLogsalary(t-1)
	AR1_6_6	0.66710	0.10137	6.58	0.0001	difLoginterest2(t-1)

FIGURE 8.2: FINAL MODEL

Bibliography

Eurostat. Handbook on residential property prices indices (rppis). pages 1,16,25,26,57,67,74, 2013a. URL http://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF.

NRK. Slik har norske boligpriser utviklet seg gjennom de siste 200 årene. August 2012. URL https://www.nrk.no/norge/boligpriser-gjennom-200-ar-1.8273402.

R. J. Shiller. Derivatives markets for home prices. page 4, April 2008. URL http://www.nber.org/papers/w13962.pdf.

J. Whelan and K. Msefer. Economic supply demand. pages 6–9, January 1996. URL https://ocw.mit.edu/courses/sloan-school-of-management/15-988-system-dynamics-self-study-fall-1998-spring-1999/readings/economics.pdf.

J. Tirole. The theory of industrial organization. page 12, 1994.

M. Silver. Why house price indexes differ: Measurement and analysis. page 4, January 2012. URL https://www.imf.org/external/pubs/ft/wp/2012/wp12125.pdf.

P. Boug and Y. Dyvi. Modag – en makroøkonomisk modell for norsk økonomi. pages 191–198, January 2009. URL https://www.ssb.no/a/publikasjoner/pdf/sos111/sos111.pdf.

Eurostat. Handbook on residential property prices indices (rppis). pages 1–179, 2013b. URL http://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF.

L. Jiang, P. C. B. Phillips, , and J. Yu. A new hedonic regression for real estate prices applied to the singapore residential market. page 3, December 2014. URL http://cowles.yale.edu/sites/default/files/files/pub/d19/d1969.pdf.

StatisticsNorway. House price index, 2016. URL https://www.ssb.no/en/priser-og-prisindekser/statistikker/bpi/kvartal/2017-01-12?fane=om#content.

StatisticsNorway. Modag and kvarts, 2017a. URL http://www.ssb.no/forskning/makrookonomi/konjunkturanalyser-og-prognoser/modag-og-kvarts.

StatisticsNorway. About us, 2017b. URL https://www.ssb.no/en/omssb/om-oss.

Alexander Stäubert. Consumer confidence index. January 2017. URL http://www.tns-gallup.no/kantar-tns-innsikt/forventningsbarometeret-1-kv-2016/.

Frode O. Andersen. Supply and demand. March 2015. URL https://www.regjeringen.no/en/topics/european-policy/areas-cooperation/statistics/id686230/.

D. H. Jacobsen and B. E. Naug. Hva driver boligprisene?//what drives the housing prices? August 2004. URL http://www.norges-bank.no/Upload/import/publikasjoner/penger_og_kreditt/2004-04/jacobsen.pdf.

J. E. Silvia, A. Iqbal, K. Swankoski, S. Watt, and S. Bullard. Economic and business forecasting: Analyzing and interpreting econometric results. pages 1–382, 2014.

D. N. Gujarati. *Basic Econometrics*. McGraw-Hill, 4th edition, 2004. ISBN 978-0072478525.

J. H. Stock and M. W. Watson. *Introduction to Econometrics*. Pearson, 3rd edition, 2011. ISBN 978-0133487701.

G. Elliott, T. J. Rothenberg, and J. H. Stock. Efficient tests for an autoregressive unit root. *Econometrica, 64:4*, pages 813–836, July 1996. URL http://scholar.harvard.edu/files/stock/files/efficient_tests_for_an_autoregressive_unit_root.pdf.

A. Iqbal. Iqbal lecture notes, advanced forecasting in business and economics. November 2016.

T. Srivastava. Deaths by swimming pool drowning vs. nicholas cage films and other spurious correlations. June 2015. URL https://www.analyticsvidhya.com/blog/2015/06/establish-causality-events/.

A. Goldman. Deaths by swimming pool drowning vs. nicholas cage films and other spurious correlations. May 2014. URL http://www.wnyc.org/story/spurious-correlations/.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37:3*, pages 424–438, August 1969. URL http://www.jstor.org/stable/1912791?seq=1#page_scan_tab_contents.

APPENDIX: DATA SOURCES

(The names are URL links)

Both measures of interest rates:

Interest rates (Statistics Norway)

For banks' average lending rate to all choose "loans in total", "Total" and "banks"

For mortgage companies' average lending rate to households choose "loans in total" "Households" and "Mortgage Companies"

Unemployment Rate (Statistics Norway)

Choose "Persons (per cent)" "Unemployed", "Both Sexes" and "15-74 years"

Income (Statistics Norway)

Choose "Seasonally adjusted", "households" and "disposable income"

Salary (Statistics Norway)

Choose "Index on monthly average earnings" and "all industries"

Housing Starts (Statistics Norway)

Choose "Building permits started. unit:dwellings". own calculations based on the different categories of "type of buildings".

Expectations Barometer (Finance Norway)

Choose the spreadcheet on the right side of the page.

APPENDIX: RMSE PROCEDURE

		REAL		VAR INSAMPLE	
	House	loghouse	House	loghouse	difloghouse
12/31/2014	173.1	5.153869462			
3/31/2015	179.5	5.190175208	174.983645	5.164692513	0.01082305
6/30/2015	185.3	5.221976133	179.6091818	5.190783278	0.026090766
9/30/2015	184.6	5.218191322	182.9924799	5.209445059	0.01866178
12/31/2015	181	5.198497031	185.3892077	5.222457441	0.013012382
3/31/2016	187.8	5.235377567	187.8493429	5.235640274	0.013182833
6/30/2016	195.5	5.275560379	190.7619547	5.25102634	0.015386066
9/30/2016	199.3	5.294811227	193.826376	5.266962789	0.015936449
12/31/2016	199.3	5.294811227	196.8014235	5.282195218	0.015232428
				20.39746224	
				32.38541176	
				2.584120802	
				19.26514467	
				0.002434719	
				22.44907334	
				29.96055926	
				6.242884738	
				4.081774913	

Figure 3: RMSE EXAMPLE: FORECAST OF 8 PERIODS AHEAD, FIRST PERIOD 2015:1

2013:1	4.94407
2013:2	7.86746
2013:3	7.030014
2013:4	5.669054
2014:1	2.646746
2014:2	3.082365
2014:3	4.809011
2014:4	3.947956
2015:1	4.081775
Average	4.897606

Figure 4: AVERAGE OF ALL RMSE FOR THAT PARTICULAR MODEL