



UNIVERSITEIT VAN AMSTERDAM

The predictive power of house price forecasting models

Student name: Martijn Duijster

Student number: 0475211

Thesis supervisor: Dhr. Prof. Dr. J.B.S. Conijn

MSc Business Economics, Finance & Real Estate (dual track)

University of Amsterdam

Abstract

Different models commonly used in house price forecasting are tested on their prediction accuracy. Using data from two house price indices for the Netherlands for the period 1995-2016, AR, ARIMA, ADL, VAR and error-correction models are selected to forecast house price index changes. Using rolling window pseudo-out-of-sample forecasting, the models estimate house price changes one up to six quarters ahead. RMSFE, MAFE and percentage correct signs are used to evaluate the findings. The ADL model significantly outperforms all other models in terms of RMSFE.

Statement of originality

This document is written by Martijn Duijster who accepts full responsibility for its contents. I declare that the text and work presented in this thesis is original and no other sources than those mentioned in the text and references have been used in its creation.

Contents

1. Introduction.....	4
2. Literature review	6
2.1. International research in real estate price forecasting	6
2.1.1. Univariate regression models.....	6
2.1.2. Multiple regression models	7
2.1.3. Vector autoregression and (vector) error correction models	8
2.2. Dutch research in real estate price forecasting	9
2.3. House price dynamics	10
2.3.1. Market characteristics	10
2.3.2. House price determinants.....	11
2.4. Forecasting models.....	15
2.4.1. ARIMA models.....	15
2.4.2. Autoregressive distributed lag models.....	16
2.4.3. VAR models.....	16
2.4.4. (Vector) error-correction model.....	17
3. Hypothesis.....	17
4. Methodology and data.....	18
4.1.1. Methodology	18
4.1.2. Rolling window forecasting.....	18
4.1.3. Forecast evaluation	19
4.1.4. Data	20
5. Results.....	23
5.1. Data preparation, model selection	23
5.1.1. Deflating and differencing	23
5.1.2. Deseasonalizing	24
5.1.3. Time series regression assumptions.....	24
5.1.4. Variable selection.....	25
5.1.5. Lag Selection	26
5.1.6. The models.....	28
5.2. Discussion of results.....	31
6. Limitations, issues and recommendations	33

7. Conclusion	34
References	35
Appendices	38

1. Introduction

The characteristics of real estate markets create the rare opportunity of predicting the future. Unlike most financial markets, the market for residential real estate is inefficient (Case and Shiller, 1988). Informational inefficiency manifests itself as random noise, inertia and predictability. In the long term, house prices adjust towards the equilibrium level as supply and demand are balanced. In the short term, the stock of housing is fixed and prices are mostly determined by demand. The demand for housing is dependent on numerous factors which are mostly related to economic circumstances. Major influences on demand are interest rates, income levels, wealth and unemployment. Price movement is usually gradual as the market takes on new information at a slow rate. The inertia in house prices is modeled as autocorrelation. This entails that lagged values of house prices can contain information on future values.

House price models can be used to incorporate these various influences to predict future house price movements. These models range from being as simple as an autoregressive AR(1) model, up to more complex systems of equations and concepts like machine learning. Models incorporate different levels of information. Univariate models only use information contained within the dependent variable itself. Multivariate models also employ explanatory variables that can contain predictive content. Systems of equations such as vector autoregressive (VAR) models use two or more dependent variables to more accurately predict the variables of influence as to generate a more precise estimate. Error-correction models consider the relationship between variables with a common stochastic trend to model an equilibrium value which helps predict the dependent variable. The latter two models can also be combined into one model to form a vector error correction (VEC) model.

Which model is most appropriate to forecast house prices is foremost dependent on economic theory and the circumstances of the research. Models should be chosen based upon their ability and characteristics to simulate the appropriate economic influences. Although it can be argued that the best models are those that most accurately consider all economic theory and influences, this adds complexity to these models which makes them vulnerable as the chances on wrong handling increases with this complexity.

Instead of judging models based their link with economic theory, another question one could ask is which model is most successful? Complexity comes at a cost and that is why a common philosophy in modelling is to keep it simple. Both approaches seem to make sense, but how is this balanced?

In looking for the tradeoff between complexity and simplicity, this research focusses on the primary goal of forecasting models; to make successful forecasts. This leads to the research question; Which type of house price forecasting model is most successful in accurately predicting house price changes in the Netherlands?

The motivation for this question lies in the gap present in the literature, especially when considered for the Dutch housing market. The Dutch literature on house price forecasting has a focus on error-correction models. In a more general sense, there certainly is literature comparing different models in their forecasting capabilities, but these models often portray only minor differences between them. This research aims to test the forecasting ability of a wide range of models to see which ones can produce the tightest fit to the observed values within the settings of the Dutch housing market.

To answer the research question, the most widely used models in house price forecasting are identified and reviewed to create a set of models that will be used to test the hypothesis. These models need the right set of variables to accurately forecast house prices. Two house price indices are used as a proxy for house price movements. The research uses quarterly data ranging from 1995q1 to 2016q4. The house price indices used are created by Statistics Netherlands (CBS) and the National Realtors Association (NVM). Testing on more than one index will increase validity of the research. As variable selection is a likely source of bias, a broad overview of commonly used variables is presented as a starting point for variable selection. Dependent upon availability, a set of variables will be selected, some of which may measure similar influences.

Models using explanatory variables need to be provided the same set of variables, so that the influence of the model itself can be explored. The initial selection will be reduced by testing for statistical properties that are desirable for use in the models. The number of lags for each model is optimized and the final models are presented. Each model is used to make pseudo out-of-sample forecasts using a rolling window technique to minimize dependence on specific circumstances over time. A sample spanning 10 years is used to make predictions for 1 up to 6 quarters ahead, the window then moves ahead one quarter and the process is repeated, resulting in 41 forecasts per model specification. Three goodness-of-fit measure are used to indicate the forecast accuracy. The root mean squared forecast error RMSFE, the mean absolute forecast error MAFE and percentage of correct signs. The hypothesis is tested based on the RMSFE as this measure is most widely used in the literature and penalizes large errors. The results are discussed as well as the limitations of the research. Finally, a conclusion is drawn to express the most relevant findings of the research.

2. Literature review

The empirical literature on house price forecasting is explored to identify the most commonly used models as a starting point for the research and to support its theoretical basis. The sections are ordered by increasing complexity of the models and from international to local to create oversight in the findings.

2.1. International research in real estate price forecasting

2.1.1. Univariate regression models

Stevenson (2007) compares the forecasting ability of ARIMA models using two UK office rent indices. ARMA(1,0) up to ARMA(3,3) specifications are used to test for the best performing specification. Special interest is given to the results of the specifications obtained from the information criteria. A rolling window technique with a sample window covering 13 years is used to make four-quarter ahead predictions. The forecast accuracy is measure by the mean absolute error, the mean squared error and the error variance. The findings indicate that selection of lags based on information criteria may not result in the best forecast performance.

Crawford and Fratanoni (2003) compare the forecasting performance of ARIMA, GARCH and Markov-switching models on US house prices indices from 1979 to 2001. ARIMA models us a

combination of autoregression on the lagged values and the lagged error terms. GARCH specifications model the variance using autoregressive lags and lags of the squared error. Markov-switching models are dynamic, meaning that they can change the underlying autoregressive specification per period of similar nature, known as regimes. This theoretically enables them to adjust to the right circumstances. Both in- and out-of-sample forecasts are conducted. The forecasts are compared based on the RMSFE and R^2 . In sample, the regime switching models are found to perform much better compared to the ARIMA and GARCH models. Out-of-sample, the ARIMA models displayed the best forecast performance. The findings in the switching models are contested by several studies like Miles (2008), who argues that Markov-switching models are not fit for forecasting.

2.1.2. Multiple regression models

Case and Shiller (1990) conduct time-series cross section regressions with US data from 1970 to 1987 in four major cities. They use explanatory variables and lagged values of the dependent variable to model house price changes. They start out with a wide range of variables and conclude that price inertia, construction costs, real income, and population changes are of influence on house prices. They conclude that the housing market is inefficient because excess returns can be made. This is due to the time it takes for information to be incorporated into the price. In the short run, autocorrelation is positive, followed by negative correlation at longer horizons. This signals that it takes time to adjust to a new equilibrium, which can be overshoot, followed by a reversal of the price movement.

Rapach and Strauss (2009) perform house price forecasts using autoregressive distributed lag ADL models and AR benchmarks. ADL models combine autoregression with cross sectional analysis and is also known as time-series cross-sectional analysis. Though some claim that an ADL model is limited to one independent variable, in this analysis ADL model can include more than one independent variable. The aim of the research is to estimate differences in forecastability across US states. They examine this across several US states from 1995 to 2006 using pseudo out-of-sample, 4 and 8 quarters ahead forecasts. National, regional and state level variables are used to form ADL models for each state. In evaluating the forecasts, the root mean squared forecast error RMSFE is used to measure the differences between the observed values and the forecasts. This is a widely used measure of fit and will be discussed the section on forecast

evaluation. The authors find large differences between the accuracy of the forecasts. In states with relatively high price growth, ADL models offer only little to no improvements over the AR model. This has the implication that the variables used contain no extra information on future house prices. Overall, house prices in coastal states (which displayed larger price growth) are therefore more difficult to predict than interior states.

2.1.3. Vector autoregression and (vector) error correction models

Brooks and Tsolacos (2002) model UK retail rents using univariate time series and vector autoregressive (VAR) models. VAR models consist of several interconnected time-series equations where the dependent variables and their lags are included in the other equations so that they exert mutual influence. They aim to identify the best performing specifications for different forecasting horizons and use different methodologies to evaluate the forecasts. The models include an autoregressive model, a long-term mean model, a random walk model and a VAR model. For the VAR model, variables from the literature are chosen as a starting point. These are then filtered for their suitability for use in a VAR model using the Dickey-Fuller test to ensure stationarity. The Granger causality test for testing predictive content in the variables in both directions is used to select variables that are fit to use in a VAR model. Information criteria are used to set lag lengths. The models are tested using the root mean squared forecast error RMSFE, the mean absolute forecast error MSFE, and percentage correct signs. To their surprise, the AR model showed to have the best forecasting results.

Zhou (1997) uses a demand model for the US housing market in the form of a VAR with an error-correction term (known as a vector error-correction model VECM). It is found that sales and house prices are cointegrated, which means they share a stochastic trend. Such long-term relationships can benefit the model because it is known where the variables will stand relative to each other in the long term. This relationship is used to produce a VAR with sales and price as both dependent and independent variables with the inclusion of an error-correction term. The model is used to forecast sales and house prices from 1991 to 1994 using a sample from 1970 to 1990. The model is found to provide a close fit for both sales and price.

2.2. Dutch research in real estate price forecasting

This section is organized around the article by Francke et al. (2009), because it evaluates error-correction models used by Dutch institutions. In between the discussion of the articles by Francke et al, the other articles and models are described.

Francke et al. (2009) evaluate existing house price models for the Netherlands. The models can be categorized as demand and supply-and-demand models. In the short-term, the stock of houses is fixed and demand variables like interest rates, income and borrowing capacity determine house prices. In the long run (more than 10 years), the supply of houses can adjust and house prices are the equilibrium value determined by the demand and supply variables. A distinction is made between stock and flow variables. Stock variables are those fixed in the short run such as wealth levels, demographics, and the beforementioned stock of houses itself. Flow variables are adjustable in the short-term and include interest rates, income and consumption. The study considers error-correction models, particularly one developed by Kranendonk and Verbruggen (2008) and one by Boelhouwer et al (2001).

The error-correction model by Kranendonk and Verbruggen (2008) consists of a long-term cointegration relationship and a short-term relationship with the error-correction term and long-term equilibrium. In the long-term the signs of the estimated variables income, interest, wealth and housing stock are all in line with economic theory, except perhaps household financial stock, which negatively influences house prices. In the short-term variable signs for income, interest rate and wealth all follow economic theory except unemployment, which is positively related to house prices.

Francke et al. (2009) review the model and test the residuals of the long-term equilibrium equation using the Dickey-Fuller test and find no evidence of cointegration (in contrast to the original authors using the Johansen test for cointegration). Another violation of the cointegration assumptions that not all variables in the cointegration relationship are non-stationary. These results therefore question the specification of an error-correction model by Kranendonk and Verbruggen (2008).

Boelhouwer et al. (2001) employ an error-correction model which includes variables to model speculative, seasonal, income and interest rate effects along with the error-correction term. Using

the model, predictions are produced for the period 2000-2010 (prudence is applied by naming them directional outcomes for the forecast horizon over two years, though in fact they are presented as percentage change, real and nominal sales prices). If the out-of-sample forecast is compared to the actual developments observed after the article was published, a different path of house prices can be seen. The model predicts declining growth from 2001, a decline in house prices from 2002 to 2007 and steady prices from there on. In reality, the depicted decline was stalled for another five years while a bubble was building. Though the forecast generally predicts the opposite sign of the actual house price development, the authors may not have anticipated further bubble build up which is a deviation from the long-term equilibrium price. With hindsight, the authors may not have predicted the actual house price but anticipated earlier adjustment toward equilibrium prices as happened starting in 2007.

Francke et al. (2008) re-estimate this model with similar results and do not mention any major flaws in the model. A comparison with the model by Kranendonk and Verbruggen (2008) reveals that the latter more closely resembles a cointegrated error-correction model. This is because Boelhouwer et al. (2007) only employ stationary variables.

Francke et al. (2008) combines aspects of both models to create a new set of models which are compared to each other. The models differ by the inclusion of a linear trend and drift, next to the random walk model. Preference is given to the random walk specification, which has the lowest standard error. Forecasts are produced for the out-of-sample period of 2010 to 2015. The proposed scenarios differ from the now observed reality in that an increase in inflation, mortgage rate and income are expected. Income is accurately predicted by the recession scenario, but inflation and mortgage rate have decreased since 2012. This causes the three models in the recession scenario to predict higher house prices in nominal terms than observed. This signifies the importance of the future expected values of explanatory variables. Other than this, no flaws in the models or prediction process can be appointed.

2.3. House price dynamics

2.3.1. Market characteristics

Case and Shiller (1989) demonstrate that the market for real estate is inefficient. Their follow-up paper (1990) demonstrates that it is possible to exploit these inefficiencies to produce forecasts.

The real estate market differs in several ways from efficient financial markets. In efficient markets such as described by Fama (1970), market prices fully reflect all available information. The property market is highly heterogeneous because of the physical location and other property characteristics. This is associated with frictions such as asymmetric information, transaction costs, carrying costs and illiquidity. In practice, this means that these markets are slow to respond and information is incomplete. Wheaton (1990) models how search costs play a central role in establishing the reservation prices for buyers and sellers through the vacancy rate. The market price is a result of bargaining between the reservation prices of both buyer and seller. The real estate market shows various causes of market inefficiencies which result in slow equilibrium adjustments. Because of these characteristics, the housing market can be successfully forecasted up to a limited degree.

2.3.2. House price determinants

In this section, the key influences on house prices are discussed to identify variables that can be used in the models. Because of the wide availability of literature on this topic, special attention is given to Dutch research as this could highlight market specific price dynamics for the Netherlands. A difficulty in selecting variables for forecasting in relation to causal effect studies is that the entire spectrum of house price influences should be considered.

In a study to identify house price determinants and their changing composition over time, Dröes and van de Minne (2015) use housing supply, construction costs, GDP per person, labor as a percentage of total population, the opportunity cost of capital, and unemployment to identify the leading influences over time. They find that currently, income and interest rates are the key determinants of house prices, this is in line with the consensus view in the literature. Income and interest rates are attributed to be the most influential forces driving real house prices. Another study of house price determinants for the Netherlands by Verbruggen et al. (2004) shows another set of variables that can explain house prices. For the short term these variables are income, interest, households, rental prices, financial wealth, and stock of houses. Starting out with these two sets of variables, we can now explore in further detail which house price determinants are appropriate for the models.

As described in the previous section, frictions cause real estate market to be inefficient. Therefore, prices do not instantly adjust, but take time to move to the new equilibrium. The

unresponsiveness of prices is manifested as autocorrelation which is the same as inertia. The lagged values of the dependent variable can therefore help to predict the future values. Univariate models are based on this concept and use the information which is present in past values.

Findings of Verbruggen et al. (2004) show that real wages have a significant impact on house prices. Together with household wealth it is said to explain 75% of the price increase of house prices between 1992 and 2000. Income is often measured as gross domestic product (GDP), sometimes expressed per person or household. Other studies use disposable income (Francke et al. 2009), which may be a more appropriate measure because taxes are also considered. Income has a positive effect on house prices, the relationship is deemed so fundamental that it is often modelled to be cointegrated with house price. This entails that they move together in the long term, or put more precisely; they have a common, stochastic trend. Usually this means that variables are integrated of order 1, or are first difference stationary. The cointegrating coefficient is the coefficient of the error correction term and is chosen so that it eliminates the common trend from the difference. Gallin (2006) however, claims that income and house prices are not cointegrated and therefore do not share a long-term trend. This implies that using an error-correction model using these two variables would be inappropriate.

Interest rates have gained importance as determinants of house prices ever since debt financing of houses became more widespread (Dröes and van de Minne, 2015). Interest can be viewed as the cost of borrowing. If interest rates drop and borrowing therefore becomes cheaper, the user costs of housing goes down. The demand for housing increases and prices will increase accordingly. De wit et al. (2013) research how changes in the mortgage rate affect price and turnover and find it has a gradual impact on house prices. The levels of debt financing in the Netherlands are amongst the highest in the world because of mortgage interest deductibility (Schilder, 2012). The fiscal stimulation of debt financing increases borrowing levels, which in turn cause higher levels of house prices. Because the mortgage rate is closely connected to house prices, it is selected as a potential variable. Long- and short-term interest rates are also included, this creates the opportunity to calculate their difference called the interest spread. There is a discussion in the literature about the predictive value of the spread. According to Stock and Watson (2003) there is some evidence of predictive content of up to two years ahead contained in the spread and it

outperforms all other measured indicators. This may well improve the forecast performance of the tested models and should therefore be included as well.

Labor is one of the driving forces of the economy, income from labor is the primary force for consumer spending. The performance of the labor market therefore has a particularly large impact on the overall economy. Measures for the labor market such as the unemployment rate are amongst the main indicators to see how well the economy is doing. Unemployment has a direct effect on the housing market. Schnure (2005) estimates that a 1% increase in the unemployment rate decreases house prices by 1% in the U.S.. Though the magnitude of unemployment rate changes to the Dutch housing market could be different, it could be indicator for forecasting house prices. Prolonged changes in the unemployment rate can also indicate the stance of the business cycle, which often coincides with changes in house prices. The unemployment rate is expressed as a percentage of the working force, both of which are included in the variable selection.

Stock market fluctuations represent changing expectations about the future income generated from the shares that make up the stock market. As an indicator of general economic circumstances, the stock market may be positively linked with the housing market. Case et al. (2005) argue that stock market fluctuations can affect house prices in several ways. The wealth effect entails that stock markets influence household wealth, which in turn is a factor in the amount of money spent on housing and therefore house prices. This view is in accordance with housing as a consumption good. From the perspective of an investment good, the amount of money invested in housing is competing with funds invested in the stock market. If the stock market outlook shows a less favorable future, investments could be transferred to the housing market therefore increasing housing prices. Though the argumentation for the sign of the relation between the stock market and the housing market covers both possibilities. It is thought that stock market fluctuations indicate economic growth which is associated with an increase in house prices. This is confirmed by van den End and Kakes (2002) which find a generally positive coherence between share prices and house prices but also see instances of contrary behavior. A variable portraying the AEX index with dividend reinvestment is included in the selection to function as an indicator for house prices.

Demographics can affect the demand for housing, the supply is set by the availability of housing over the different submarkets. If the balance of supply and demand is disrupted, this will affect house prices. The Dutch housing market is characterized by an inelastic supply due to limited space in popular locations and severe zoning restrictions. Francke and van de Minne (2014) find that in countries with local supply constraints such as the Netherlands, population growth will result in higher house prices. To include such effects, variables used in the research to model the effect of demographics and housing supply are included. Population, households, housing stock, construction cost index, and building permits are included in the potential variables to account for the demand and supply of housing from a demographic perspective. A concern for the Netherlands is the ageing population and a smaller share of working people to support an increasing amount of retired people. The effect of an ageing population will decrease house prices (Takáts, 2012) (van Dam et al. 2016). Variables are included to address changes in demographics, but due to the relatively short forecasting horizon, it seems unlikely that these will significantly impact house price forecasts on a national level.

Besides the previously mentioned factual measures, perceptions of future house prices can be related to actual house prices. Housing market sentiment is considered to have an impact on house price developments. Huang's (2014) research involves the Good Time To Buy Index which quantifies perceptions of the future housing market. The results show that the index can signal housing busts three quarters in advance. In the Netherlands, Boumeester (2014) concludes that the Eigen Huis Market Indicator has a reasonable coherence with the actual house price developments. The Eigen Huis Market Indicator was introduced in 2004, and therefore does not cover the time span of the sample period. Fortunately, the Consumer Economic Survey of the Central Bureau of Statistics Netherlands contains the question about the propensity to buy a house in the coming two years. Out of all possible replies, the share of positive responses is taken to construct a variable that measures the propensity to buy a house in the coming two years. As an additional measure consumer confidence is included though it is said not to correspond with the Eigen Huis Market Indicator.

2.4. Forecasting models

2.4.1. ARIMA models

By combining an AR(p) model with an MA(q) model, integrated of order (i), the ARIMA model is constructed. The current value of a variable in an autoregressive model depends on the past values of the variable plus an error term. The term p, in AR(p), denotes the number of lags considered in the model. The noise term is denoted as u_t .

$$y_t = u + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + u_t$$

It is desirable property for an AR(p) process to be stationary. This means that the autocorrelation will converge to zero after a limited number of lags. If this is not the case a unit root may be present, which is undesirable for forecasting purposes.

The moving average model is a linear regression on the current and lagged error terms. The error terms are assumed to be independent of each other, to follow a normal distribution, with constant mean and constant variance.

$$y_t = u + u_1 + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q}$$

A finite MA process is always stationary, has a constant mean, a constant variance and the autocovariance can be different from zero up to lag q and will be zero thereafter.

An autoregressive moving average model ARMA(p,q) is created by combining an AR(p) with a MA(q) model. In this combined model, the current value is linear dependent on its own past values and the current and past values of the white noise term.

$$y_t = u + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} + u_t$$

The white noise term is normally distributed, has a mean of zero, constant variance. The I in ARIMA stands for integrated. Because the variable underlying the ARMA model needs to be stationary, differencing sometimes needs to be applied to meet this demand. An ARMA(p,q) model of a variable differenced d times is therefore the same as an ARIMA(p,d,q) model. The orders of the model could be determined graphically using autocorrelation function ACF and partial autocorrelation PACF plots or with information criterions like Akaike's information criterion AIC of Schwarz's Bayesian information criterion SBIC.

2.4.2. Autoregressive distributed lag models

The market for single-family homes is known to have market inefficiencies which cause prices to slowly adjust to new price levels (Case and Schiller, 1988). This characteristic can be exploited to forecast price movements in this market. If new information is slowly incorporated into prices, then adding lagged variables can transfer this information to the model thereby improving upon a model only capturing information from its own past values.

The autoregressive distributed lag model combines an AR(p) model with lagged values of explanatory variables. Exactly which explanatory variables and how many lags should be used depends on economic theory and the information present in the lags. An ADL model with p lags for the dependent variable y, k explanatory variables x with q lags will look like this.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \alpha_{11} x_{t-1} + \dots + \alpha_{1q} x_{1t-q} + \alpha_{k1} x_{1t-1} + \dots + \alpha_{kq} x_{kt-q} + u_t$$

2.4.3. VAR models

Vector autoregressive models (VAR) are widely used to research house prices dynamics. A VAR is a system of equations in which each dependent variable is explained by its own lags and the lags of other variables, which themselves are the dependent variable in the system of equations. It combines features of univariate time series and simultaneous equations. It was introduced by Sims (1980) to capture the dynamics of multiple time series.

VAR models depict economic dependencies as a system of variables that are influenced by their own past values and the past values of the other variables. This implies that all variables in the system are endogenous, they are created within the model. The variables used in such a model must therefore show two-way causality, not only must variable A influence variable B, but also the other way around. This puts a restriction on the variables to be used. If exogenous variables are added to the model, it is no longer a true VAR but is often referred to as a VARX. This use may not be in the spirit of the true VAR according to Sims (1980) who states that a Var should be as unrestricted as possible. Another condition for a VAR to be unrestricted, is that the number of lags for each equation will have to be the same length. A VAR with different lag lengths per equation can be viewed as setting the coefficient of left out lags to zero. An obvious drawback of

using a VAR model is that many parameters need to be estimated. This can be troublesome in case of a relatively small sample size, as the degrees of freedom will run out, resulting in large standard errors (Brooks, 2014, pp-330-334).

To select the lag length of a VAR an adaptation of the Akaike information criterion AIC can be used. The multivariate Akaike information criterium MAIC trades-off weights the trade-off between the fall in the residual sum of squares RSS as more lags are added. The lowest MAIC value should be selected to come to the ideal number of lags in the VAR model.

2.4.4. (Vector) error-correction model

If two variables X_t and Y_t that are integrated of order one are linearly combined, the difference between them will usually also be integrated of order one. But if a linear combination X_t and Y_t is integrated of order zero, X_t and Y_t are said to be cointegrated. The coefficient Θ for which $Y_t - \Theta X_t$ is integrated of order zero is called the cointegrating coefficient. The term $Y_t - \Theta X_t$ is called the error correction term. Modelling cointegration into the model can help improve forecast accuracy. If Θ is chosen so that it eliminates the common trend, the error correction term can be included as an additional variable in the model because it is stationary. This can be used in an ADL model to create an error correction model or a VAR model to create a VEC model.

$$\begin{aligned}\Delta Y_t &= \beta_{10} + \beta_{11}\Delta Y_{t-1} + \dots + \beta_{1p}Y_{t-p} + \gamma_{11}\Delta X_{t-1} + \dots + \gamma_{1p}\Delta X_{1t-p} + \alpha_1(Y_{t-1} - \Theta X_{t-1}) \\ &\quad + u_{1t} \\ \Delta X_t &= \beta_{20} + \beta_{21}\Delta Y_{t-1} + \dots + \beta_{2p}Y_{t-p} + \gamma_{21}\Delta X_{t-1} + \dots + \gamma_{2p}\Delta X_{1t-p} + \alpha_2(Y_{t-1} - \Theta X_{t-1}) \\ &\quad + u_{2t}\end{aligned}$$

3. Hypothesis

The research question; which type of house price forecasting model is most successful in accurately predicting house price changes in the Netherlands? Is translated into hypotheses so it can be quantifiably tested and is divided into 2 sets 5 sub hypotheses for statistical purposes.

Hypothesis 1_{a,b,c,d,e}: The AR/ARMA/ADL/VAR/error-correction model can predict house price changes modeled by the CBS index significantly more accurate than all other models.

Hypothesis 2_{a,b,c,d,e}: The AR/ARMA/ADL/VAR/error-correction model can predict house price changes modeled by the NVM index significantly more accurate than all other models.

4. Methodology and data

This paper aims to find out which models can most accurately predict future house price developments within the specified settings. AR, ARMA, ADL, VAR and error-correction models are considered in various appearances to optimize their forecasting capability. The forecasts are conducted in rolling windows as to minimize bias from differences in economic circumstances. Forecast horizons are made from one up to six quarters to check how the different models perform over an increasing forecast horizon. The accuracy is compared using the most established methods. For hypothesis testing, pairwise t-tests are performed on all combinations of models per index, using the average RMSFE for the six different forecast horizons. This paper adds to the existing forecast comparison literature that hypothesis testing is used to test for significant differences between goodness-of-fit measures. To add to the understanding of the forecast accuracy, the mean absolute error and percentage correct signs are also reviewed. This will reveal which models are most accurate in forecasting the change in house prices. In this process, all procedures and tests are described to ensure statistical validity.

4.1.1. Methodology

4.1.2. Rolling window forecasting

The goal of this research is to evaluate how forecasts generated using different models fit the observed house price changes. Pseudo out-of-sample forecasting combines the ability to compare forecast with the actual values with the real-life restriction that future values cannot be used in forecasting models. It entails that a part of the sample is used to estimate the model's parameters and the following part is used to compare the forecast with the actual data.

A problem with choosing a certain fixed time window to forecast is that the circumstances of that period are of great influence on the fit between the forecast and observed values. Forecast testing at a time of big price changes may lead to different conclusions than in periods of economic stability. To overcome this problem and avoid arbitrarily choosing forecast periods that will influence the evaluation, rolling window forecasting will be used. This technique is a useful tool in comparative forecast evaluation. In a rolling window, the in-sample period length used to estimate the parameters of the model is fixed. In accordance with the Brooks and Tsolacos (2000 p.11), a 10 year in-sample window length is chosen.

The forecast horizon is also of influence in measuring the accuracy of the forecasts. Choosing to forecast a certain number of periods ahead in the future will favor certain models over others. Choosing the length of the forecast horizons and windows is arbitrary. It is a compromise between having too few periods ahead and not exploiting all forecasting power the model possesses and choosing too long of a horizon, resulting in ever more random forecasts. As a lot of earlier work uses forecasting horizons of up to one or two years, a compromise is made of one and a half years consisting of six increasing steps of one quarter (Brooks and Tsolacos (2000) 1 to 8, Rapach and Strauss (2007) 4 and 8, Stevenson (2007) 1 to 4).

The number of steps for the rolling window is limited to 41. The first six quarter ahead forecast will use the 10 year in-sample estimation period of 1995q2 to 2005q1 to make a forecast for the 'out-of-sample' forecast evaluation period 2006q3. The whole process then moves forward one step (one quarter) and the next forecast will be made until step 41, when the in-sample period of 2005q3 to 2015q2 is used to make the forecast for period 2016q4.

4.1.3. Forecast evaluation

The error measures used for forecast evaluation are of great importance in determining the best fitting model. Almost all research in forecast evaluation uses error measures that penalize large errors by squaring them. This paper will follow the convention of penalizing large errors but the point is made that this can be viewed as an arbitrary, though logical choice. The most widely used error measures in the literature are the mean squared forecast error MSFE and the root mean squared forecast error RMSFE. Both error measures penalize large errors by squaring them. The difference in interpretation is that the RMSFE is the square root of the MSFE. The choice is made for the RMSFE as this will result in a more comprehensive oversight of the results as the error measure removes the quadratic term.

Besides the RMSFE, the mean absolute forecast error MAFE and the percentage correct signs predictions are often used as complimentary forecast measures. The MAFE measures the average of the absolute differences between the predictions and observations. The percentage correct sign prediction measures the percentage of times the forecast predicts the right direction of the observation changes and was introduced Psearan and Timmermann (1992). The motivation for including these measures is that they penalize large errors less heavily. The RMSFE

disproportionally penalizes large errors, the MAFE proportionally penalizes errors and the percentage correct sign does not take the size of the errors into account.

4.1.4. Data

The literature on forecasting in real estate has identified variables of influence on prices which are taken as a starting point for creating the models in question. They are gathered from various sources and have been altered to enable their use in the models. The sample period of the quarterly data ranges from 1995q1 to 2016q4 creating 88 observations per variable. The availability of two different house price indices can help to provide insight whether the measurement of house price movements itself influences the models and variables used. A brief description of the variables and their sources is given and a table of descriptive statistics is presented in Appendix 1, Table 1.

CBS house price index (CBS)

The Central Bureau of Statistics provides a house price index based on the sale price appreciation method. In short, this method keeps track of the ratio of the average sale price and the average WOZ value (waardering onroerende zaken), which is a certified valuation for tax purposes. This functions as a control for variations in price level attributable to quality differences of houses being sold in a certain period. Because the basis of the index are transactions taken from the land registry, there is a lag time of approximately two months in the CBS house price index (De Haan et al, 2009).

NVM house price index (NVM)

The Dutch Association of Realtors NVM also lists a quarterly index of house prices. The index is based on the transactions recorded by their members which account for roughly 75% of all transactions in the Netherlands. Another difference with the CBS index is that there is no lag because the transactions are recorded at the time of sale. A third difference is that the NVM applies weights based on houses sold instead of housing supply. This has the consequence that there is a relative overrepresentation of apartments and urban area housing as they have a larger share in the number of sales than in the supply of houses (NVM, 2017).

Mortgage rate (CBS) long- and short-term interest rates and the term spread (OECD)

Due to the lack of a continuous series of a quarterly mortgage rate from 1995 up to 2016, two series from the CBS are combined so that this variable can be considered. The average mortgage rate series ranging from 1993 to 2003 is combined with the series mortgage rate ranging from 1999 to 2016. The two variables show a minor deviation of a maximum of half a percent so that a gradual average was used in the overlapping period. The 10-year government bond rate is used as a proxy for the long-term interest rate. The three-month Euribor rate is the interest rate at which in interbank loans between primary banks in the EU are priced. It is widely recognized as a good measure for the short-term interest rate. Using both these variables, the term spread is calculated as the difference between the long- and the short-term interest rates.

Gross Domestic Product (OECD, DataStream) and mode income (CBS)

The annual gross domestic product of the Netherlands is taken from DataStream and fitted to the quarterly frequency by applying compound growth per quarter. To add relevance to the variable it is also presented as GDP per household by dividing it by the number of households. Though GDP is often used in house price modelling, a more relevant measure of income may be the mode of income in the Netherlands as it is less influenced by extreme values that do influence GDP. Mode income could prove to have larger predictive powers than GDP measures because it is less influenced by fluctuations in the top income groups which are not that relevant for forecasting house price index changes. These income variables can improve the models by adding information about income that may influence house prices.

Stock (DataStream)

AEX total return is taken from DataStream as a proxy for the entire stock market and economic climate. Dividend is incorporated in this series to get a more realistic series than the actual AEX value over time. The stock variable may be good indicator for the stance of the economy and business cycles.

Construction costs, building permits, houses and new houses (CBS)

Construction cost can influence house prices as constructing new houses can be a substitute for existing housing. The series is provided by the CBS and indexed with base year 2010. The number of building permits is an indicator of the new housing supply and presumably leads the variable new houses because of the lag time of construction. The construction of new houses is

relatively inelastic in the Netherlands due to land use restrictions. The variables are used to capture the influences of the supply of houses and especially of newly built housing.

Unemployment and labor (CBS)

Unemployment states the percentage unemployed people relative to the working population which is given by the variable labor. The unemployment rate is a widely used measure in reviewing the stance of the economy. Long lasting trends in unemployment and the working population can signal the stance of business cycles. These variables represent the conditions of the labor market and overall economy which are relevant in forecasting house prices.

Total mortgage size (CBS)

This variable contains the total size of all residential mortgages in millions. For the years up to 2000, only annual data is available. The data is interpolated using a simple average. Using exponential growth using log specification would lead to ‘flat’ rates per annum in case the variable would be log difference specified.

Population and households (CBS)

The CBS provides an annual series of the number of households and the population size in the Netherlands. The series are interpolated to estimate the missing values in-between the annual values. These two variables have the role of describing demographics and thereby the demand for housing.

Propensity to buy and consumer confidence (CBS)

Percentage buy gives the percentage of households that intend to buy a house within the next two years. Consumer confidence is the balance of positive to negative answers of respondents about their feel about economic conditions. Both are survey data variables from the CBS are included to check if they may have predictive qualities. These variables are included to cover how people feel about the future conditions in the housing market.

Consumer Price Index and percentage inflation (CBS)

The consumer price index is used to deflate prices to real levels so that the inflation component of price changes over time is eliminated. The percentage inflation is used to deflate percentage interest rates to the current levels.

5. Results

5.1. Data preparation, model selection

This section is dedicated to the selection of variables and optimization of the models for forecasting. It should be noted that the models are optimized for using the entire sample period of q2 1995 to q4 2016, but due to the rolling window regression methodology, the actual sample size will be 10 years in each regression made in the rolling window process. This means the model may not always be optimal for the specific sample, and may even cause issues.

Every model specification requires 6 steps ahead times 41 rolling windows therefore requiring 246 regressions per specification. This number is multiplied by the different archetype models (AR, ARMA, ADL, VAR and EC) and each different number of lags tested per model. Because of these quantities, most test are performed on the entire observation period instead of all the subsamples used for the regressions. Another consequence of the large amount of regressions due to the rolling windows technique is that is nearly impossible to test all the residuals. To create the forecast models, some of the variables are transformed to a more appropriate form. Subsequently the time series regression assumptions are tested. The variables are selected and the appropriate number of lags for each model is determined. Finally, the model specifications are presented that are used to test the forecast performance of each type.

5.1.1. Deflating and differencing

Using the consumer price index, the financial variables are deflated to real prices so that they can be compared across time. This results in the variables real price CBS, real price NVM, real GDP, real mode income, real stock, real construction costs. The mortgage- and long-term interest rates are deflated using the percentage change of the CPI, thereby creating the variables real mortgage rate and real interest rate.

The log difference is applied to all variables where appropriate. Some interest rates depict negative values which prevents logarithmic specification as log functions do not allow for

negative numbers. In these cases, the simple difference is used as these series are quoted as percentages. Consumer confidence is also exempt from log specification due to the nature of the data. These transformation is made to induce stationarity, which is a desirable property for time series regression. The logarithmic scaling has the advantage over percentage differences that a +0.5 change followed by a -0.5 change brings it back to the same value and is widely applied in similar research.

5.1.2. Deseasonalizing

Graphical inspection of the variables reveals seasonality in the NVM price index. It is most obviously visible in the autocorrelation function plot. A limited amount of autoregression is no problem for time series regression assumptions, but a seasonal effect of this size must be corrected for. This is done by applying a moving-average filter which adjusts the smooths out the peaks that occur yearly. The filter applied has weights of (1,2,7,2,1). The weights of the moving average are chosen based on satisfying tests results for autocorrelation and stationarity tests yet minimizing the impact to the variable. The effects of de-seasoning are shown in Figure 1 in Appendix 2. It shows a graph of $\Delta \log$ real house price NVM before and after the procedure. Figures 1c and 1e depict the autocorrelation functions of the untreated and treated variable. Further mentions of the NVM house price index variable concern the deseasonalized version unless explicitly mentioned.

5.1.3. Time series regression assumptions

To ensure the validity of the research, the time series regressions assumptions have to be met.

- 1) The noise term has a conditional mean zero
- 2) (a) The data have a stationary distribution
(b) Autocorrelation becomes insignificant as the amount of time between observations increases
- 3) Large outliers are unlikely
- 4) No perfect multicollinearity

Results of the following tests can be found in Appendix 2,3 and 4. Variables that are no longer eligible to be used in the models are not displayed in the subsequent tests as to maintain oversight of the results. The first condition that the expected value of the error term is zero, this is checked

after conducting regressions. The amount of regressions makes it very impractical to display all these results. The Dickey-Fuller test is used in testing for stationarity. The type of test per variable (trend or no trend) is decided on the nature of the variable visual and visual inspection. Plots of some the variables given in Figures 2a to al in Appendix 5. Autocorrelation plots are inspected to ensure limited autocorrelation. These can be seen in Figures 1a and 1e. There are no large outliers, this is checked graphically using the variable plots in Figures 2a to al in Appendix 5. There is no multicollinearity, the correlation diagrams of Table 3 in Appendix 4. show no correlations above 0.8. Only the correlations for the variables used in the models are displayed.

5.1.4. Variable selection

To make the comparisons between the models as fair as possible, the same set of variables must be included in each multivariate model. If the model would be fitted with different variables it would be impossible to make a fair judgement on the model itself, it would be biased by the variables it incorporates. The choice for a certain set of variables will likely favor one model over another. That is why the selection of variables is set up with the aim of minimizing arbitrary choices and let the data decide which variables to include, as suggested by Brooks and Tsolacos, (2000).

The explanatory variables that have passed the statistical tests to meet the time series regression assumptions are tested for their mutual influence on the house price variables. Granger causality statistics indicate whether lagged variables of one variable can help predict the other. The term causality in this sense does not imply economic theory causation. It merely confirms that the variables contain information that is useful for forecasting the other variable beyond the information contained in the past values of the variable itself. The results of the Granger causality tests are given in Table 4a and b. The number of lags for each dependent variable is based on the number of significant autocorrelation lags (6 for the CBS series, 4 for the NVM series), obtained from the autocorrelation plots given in Figures in Appendix 4.

Some of the variables collected measure the same influences. Therefore, only one of them should be selected to be included in the model. If more than one of these variables display significant two-way causation in the Granger causality tests, the variable with the lowest P-value for Granger causation towards the dependent variable in the test is selected. It is expected that the variable

with the lowest P-value will be better able in getting the model to obtain the closest fit between the observations and the forecast values.

For the variable of interest $\Delta \log$ real price CBS, the following variables have significant two-way Granger causality using the 6 lags from the autocorrelation plot; propensity to buy, $\Delta \log$ real mode income, $\Delta \log$ unemployment, $\Delta \log$ transactions. The variables building permits, $\Delta \log$ real mortgage rate, number of houses, $\Delta \log$ real GDP, $\Delta \log$ GDP per household, $\Delta \log$ total mortgage debt, $\Delta \log$ labor, $\Delta \log$ construction costs, $\Delta \log$ consumer confidence and $\Delta \log$ households all have significant one-way Granger causality at the 5% level.

For the model using the NVM index, other variables contain information on the future values of $\Delta \log$ real price NVM. The results of the 4-lag Granger causality tests indicate that Δ term spread, $\Delta \log$ real GDP, $\Delta \log$ real GDP per household, $\Delta \log$ transactions and $\Delta \log$ unemployment all show two-way Granger causality. Because $\Delta \log$ GDP and $\Delta \log$ GDP per household have the same P-value for Granger causality towards $\Delta \log$ real price NVM, the choice is made to drop $\Delta \log$ GDP because it has less significant Granger causality from $\Delta \log$ real price NVM. The variables showing one-way Granger causality towards $\Delta \log$ real price NVM are building permits, $\Delta \log$ real mode income, $\Delta \log$ real total mortgage $\Delta \log$ real stock $\Delta \log$ labor, $\Delta \log$ construction costs and Δ consumer confidence.

It may seem like a waste of good data not to use these variables in the models, as including some of these endogenous variables could result in a better fitting model. However, the aim of this research is to identify the archetype model that produces the best forecast performance. So although the models could easily be expanded to form so called ARIMAX and VARX type models, this would distract from the main research question and is therefore better suited for a separate research paper.

5.1.5. Lag Selection

In the search for models with the best forecasting ability for the Dutch housing market it is important to consider the number of lags included in each model as they may contain valuable information on future house prices. In the literature, the most common method for lag selection is using information criteria. Both the Akaike Information Criterion AIC and the Bayesian Information Criterion BIC are considered and in case of different outcomes, the results of both information criteria will be tested. A problem with selecting the number of lags based on

information criteria however is that they do not always select the specification with the best forecast performance Chaplin (1999). To overcome this problem, a simple 3 step algorithm is followed to search for better fitting model specifications. Step 1, models with the number of lags obtained by the AIC and BIC are tested, added with testing for lags 1 and 4 to include the lowest number of lags and annual effects. Step 2, to the best fitting model, measured by the lowest average RMSFE over 6 steps, one lag is added and subtracted (up to lag 1) and tested again. Step 3, if step 2 does not show improvement, the model with the selected number of lags that has the lowest RMSFE is used in evaluating the model type, otherwise step 2 is repeated. An advantage of this method is that it limits the number of results to display which can become overwhelming. A disadvantage is that it does not consider the accuracy of individual step ahead forecast but rather the average of forecast steps. To avoid this, all individual step ahead forecasts have been carefully monitored to avoid dropping the potentially best forecast models. The motivation of using this method is simply to find the best fitting forecast model and simultaneously avoid bias by arbitrarily choosing a model for the comparison.

The decision whether to include lag specifications outside the values given by the information criteria is taken on balancing the cost of potential data-mining versus the cost of not choosing the optimal lag specification for the forecast. Judgement that the bias caused by not choosing the optimal number of lags is likely to outweigh the bias of potential data mining has determined to add the additional lag specifications.

The results of the lag selection process to find the best lag specifications have resulted in different models than specified by the AIC and BIC. A partial explanation for this can be the different sample size used for the tests of the AIC and BIC versus the testing of the RMSFE. But mostly it is because the AIC and BIC are tests to determine whether the marginal benefit of including additional lags will outweigh the marginal cost of added estimation uncertainty. The RMSFE measures the squared errors of the fit between the observed and the forecast values.

Cointegration

For the error-correction model, the Johansen test for cointegration is used to look for cointegrating relations with the house price models. The simplest cointegrating relationship found is used for the error correction model (in number of variables and not exceeding rank 1). The CBS real house price has no cointegrating relationships with a single variable but the

combination of unemployment, mode income and transactions together cointegrate with the real house price variable. Within the NVM model, the variable transaction is found to have a cointegrating relationship with house prices. The equations are regressed to use the residuals as the error correction term ECT.

5.1.6. The models

The process of transforming and testing variables for their suitable characteristics regarding the time series assumptions, selecting the ones with two-way Granger causality and optimizing the number of lags has resulted in the following models.

For forecasting the CBS house price index changes;

AR (4)

$$\Delta \ln CBS_t = \beta_1 \Delta \ln CBS_{t-1} + \dots + \beta_4 \Delta \ln CBS_{t-4} + \varepsilon_t$$

ARMA (1,1)

$$\Delta \ln r CBS_t = \beta_1 \Delta \ln r CBS_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

ADL (4)

$$\begin{aligned} \Delta \ln r CBS_t = & \beta_1 \Delta \ln r CBS_{t-1...4} + \beta_2 PTB_{t-1...4} + \beta_3 \Delta \ln r Income_{t-1...4} \\ & + \beta_4 \Delta \ln Unemployment_{t-1...4} + \beta_5 \Delta \ln Trans_{t-1...2} + \varepsilon_t \end{aligned}$$

VAR (2)

$$\begin{aligned} \Delta \ln r CBS_t = & \beta_1 \Delta \ln r CBS_{t-1...2} + \beta_2 PTB_{t-1...2} + \beta_3 \Delta \ln r Income_{t-1...2} \\ & + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \varepsilon_t \end{aligned}$$

$$\begin{aligned} PTB_t = & \beta_1 \Delta \ln r CBS_{t-1...2} + \beta_2 PTB_{t-1...2} + \beta_3 \Delta \ln r Income_{t-1...2} + \beta_4 \Delta \ln Unemployment_{t-1...2} \\ & + \beta_5 \Delta \ln Trans_{t-1...2} + \mu_t \end{aligned}$$

$$\begin{aligned} \Delta \ln r Income_t = & \Delta \ln r CBS_{t-1...2} + \beta_2 PTB_{t-1...2} + \beta_3 \Delta \ln r Income_{t-1...2} \\ & + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \omega_t \end{aligned}$$

$$\Delta \ln Unemployment_t$$

$$\begin{aligned} = & \beta_1 \Delta \ln r CBS_{t-1...2} + \beta_2 PTB_{t-1...2} + \beta_3 \Delta \ln r Income_{t-1...2} \\ & + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \varphi_t \end{aligned}$$

$$\Delta \ln sTrans_t = \beta_1 \Delta \ln rCBS_{t-1...2} + \beta_2 PTB_{t-1...2} + \beta_3 \Delta \ln rIncome_{t-1...2} \\ + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \varphi_t$$

ECM (2)

$$\Delta \ln rCBS_t = \beta_1 \Delta \ln rCBS_{t-1...2} + \beta_2 PTB_{t-1...2} + \beta_3 \Delta \ln rIncome_{t-1...2} \\ + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \beta_5 ECT_{t-1...2} + \varepsilon_{t-1}$$

Where ECT are the residual values of the function

$$CBS_t = PTB + \beta_2 RealModeIncome + \beta_4 Unemployment_t + \varepsilon_{t-1}$$

For forecasting the NVM house price index changes

AR(4)

$$\Delta \ln NVM_t = \beta_1 \Delta \ln NVM_{t-1} + \dots + NVM + \varepsilon_t$$

ARMA (1,1)

$$\Delta \ln rNVM_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

ADL (1)

$$\Delta \ln rNVM_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1} + \beta_2 \Delta \ln rSpread_{t-1} + \beta_3 \Delta \ln rGDPHH_{t-1} \\ + \beta_4 \Delta \ln Unemployment_{t-1} + \beta_5 \Delta \ln Trans_{t-1} + \varepsilon_t$$

VAR (2)

$$\Delta \ln rNVM_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1...2} + \beta_2 \Delta \ln Spread_{t-1...2} + \beta_3 \Delta \ln rGDPHH_{t-1...2} \\ + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \varepsilon_t$$

$$\Delta \ln Spread_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1...2} + \beta_2 \Delta \ln Spread_{t-1...2} + \beta_3 \Delta \ln rGDPHH_{t-1...2} \\ + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \varepsilon_t$$

$$\Delta \ln rGDPHH_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1...2} + \Delta \ln Spread + \beta_3 \Delta \ln rGDPHH_{t-1...2} \\ + \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \varepsilon_t$$

$$\Delta \ln Unemployment_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1...2} + \beta_2 \Delta \ln Spread_{t-1...2} + \beta_3 \Delta \ln rGDPHH_{t-1...2} + \\ \beta_4 \Delta \ln Unemployment_{t-1...2} + \beta_5 \Delta \ln Trans_{t-1...2} + \varepsilon_t$$

$$\Delta \ln rTrans_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1..2} + \Delta \ln Spread + \beta_3 \Delta \ln rGDPHH_{t-1..2} \\ + \beta_4 \Delta \ln Unemployment_{t-1..2} + \beta_5 \Delta \ln Trans_{t-1..2} + \varepsilon_t$$

ECM (2)

$$\Delta \ln rNVM_t = \beta_0 + \beta_1 \Delta \ln rNVM_{t-1..2} + \beta_2 \Delta \ln rSpread_{t-1..2} + \beta_3 \Delta \ln rGDPHH_{t-1..2} \\ + \beta_4 \Delta \ln Unemployment_{t-1..2} + \beta_5 \Delta \ln Trans_{t-1..2} + \beta_5 ECM_{t-1..2} + \varepsilon_{t-1}$$

Where ECT are the residual values of the function

$$\ln NVM_t = \beta_1 \ln Transactions_t$$

5.2. Discussion of results

Using information criteria may be the wrong approach in finding the best fitting forecast models (Stevenson 2007). The models that have the best fit in terms of the RMSFE do not always have the number of lags the information criteria support. Tables 6a and d provide average of the RMSFE for 1 up to 6 quarter ahead in the rolling window forecasts. In the appendix additional forecast measures are displayed in the remainder of table 6.

Looking at the results of the best forecasting specification of 5 widely used models, several relations can be spotted. The most obvious may be that as the forecast horizon increases, the size of the errors also usually increases, which is not surprising as it is more difficult to make prediction further into the future. The models each follow their own trend regarding the increase of the error size. The model types used for both house price indices show somewhat of a pattern between the CBS and NVM estimates, though is hard to spot this without formal testing. From Table 6, it seems that for the CBS data, the ADL model has produced the most accurate forecast. It is a little harder to see which model is successful using the NVM data. Therefore, a pairwise t-test is performed between sets of forecasts of the same index to see if there are significant differences between them.

Table 6a. Root mean squared forecast error of $\Delta \log$ real HPI CBS

Model	Forecast horizon					
	1	2	3	4	5	6
AR (4)	0.000165	0.000184	0.000213	0.000237	0.000320	0.000337
ARMA (1,1)	0.000164	0.000185	0.000229	0.000258	0.000328	0.000340
ADL (4)	0.000135	0.000147	0.000150	0.000192	0.000211	0.000221
VAR(2)	0.000138	0.000184	0.000244	0.000354	0.000467	0.000527
ECM (2)	0.000126	0.000160	0.000177	0.000220	0.000300	0.000299

Table 6d. Root mean squared forecast error of $\Delta \log$ real HPI NVM

Model	Forecast horizon					
	1	2	3	4	5	6
AR (4)	0.000138	0.000295	0.000423	0.000433	0.000455	0.000455
ARMA (1,1)	0.000398	0.000538	0.000635	0.000609	0.000689	0.000689
ADL (1)	0.000096	0.000185	0.000312	0.000346	0.000417	0.000456
VAR (2)	0.000095	0.000200	0.000383	0.000433	0.000465	0.000477
ECM (2)	0.000130	0.000271	0.000416	0.000435	0.000487	0.000495

Table 7 reads as follows; the values displayed in the tables are P-values of pairwise t-test using the sample of six forecast horizons. The top row models are tested against the left column models the hypothesis: $\text{RMSFE}_{\text{top row model}} < \text{RMSFE}_{\text{left column model}}$. The AR (4) model is therefore thought to better forecast the CBS house price index changes than the ARMA (1,1) in the setting of this research.

Testing the hypothesis if there is a model outperforming all other models is indicated by an entire column of P-values below the significance level. Here we can see that this is the case and the hypothesis by looking at the column of the ADL (4) model. The null hypothesis is rejected at a 5% significance level in favor of the alternative hypothesis: The ADL model produces significantly better fitting forecasts than the other tested models for the CBS house price index. Looking at the second table displaying the P-values regarding the NVM house price changes, the same conclusion can be drawn. The ADL model significantly produces forecasts with a better fit measured by the RMSFE than all other model.

Table 7a. P-value of pairwise t-tests average RMSFE CBS

	AR (4)	ARMA (1,1)	ADL (4)	VAR(2)	ECM (2)
AR (4)	-	0.9617	0.0036	0.9574	0.1381
ARMA (1,1)	0.0383	-	0.0025	0.9465	0.0359
ADL (4)	0.9964	0.9975	-	0.9831	0.9784
VAR(2)	0.0426	0.0535	0.0169	-	0.0166
ECM (2)	0.8619	0.9641	0.0216	0.9834	-

Table 7d. P-value of pairwise t-tests average RMSFE NVM

	AR (1)	ARMA (1,1)	ADL (1)	VAR(2)	ECM (2)
AR (1)	-	0.9999	0.0087	0.1153	0.7042
ARMA (1,1)	0.0000	-	0.0000	0.0001	0.0000
ADL (1)	0.9913	0.9999	-	0.9823	0.9991
VAR(2)	0.8847	0.9999	0.0177	-	0.9875
ECM (2)	0.2958	0.9999	0.0009	0.0125	-

6. Limitations, issues and recommendations

This thesis aims to identify the best fitting models for future house prices changes measure by the indices. This does not imply that the model with the best fit also produces the best models to use in practice. Goodness-of-fit measure can quantify one model over another, but this says very little about the actual quality of the model.

The variables selected for the purpose may not have been the best variables in general for forecasting purposes, but manually selecting a set of models, bias would be created as the results may have been very different for another set of variables. This process has presented restrictions on the use of variables. All models should be equipped with the same set, otherwise there would have been little sense in evaluating their relative performance. Selection bias is inevitable in such a research, the inclusion for a VAR model necessitates two-way causality between the variables employed, giving preferential treatment to this model because of this.

Another form of bias is the time scale suited for the different models. The strengths of the VAR and error correction model is that they also dynamically predict variables to be used in forecasting the house price changes. This is much more beneficial in the long term than over a maximum of 6 quarter, which is probably why the ADL model has been working relatively well.

Restrictions if the form of data availability influenced the data gathering process. Potentially useful variables were omitted due to limited data availability. The use of available data was further restricted due to the time span of the sample period, as a lot of relevant variables are not available for the full period. This causes omitted variable bias in the models. The severity of this problem is reduced because the aim of the research is not to test the influences of the variables on house prices, but rather the performance of models using the with and without the same set of variables. Not having the optimal set of variables will therefore relatively favor the univariate models as they do not depend on explanatory variables.

This thesis should be regarded as an attempt to increase understanding in the formation of forecasting models and the evaluation thereof. It has no external validity about the performance of the models in question.

7. Conclusion

In filling the gap in the literature about forecast accuracy between most of the common models used in house price forecasting, this research has selected models from the archetypes AR, ARIMA, ASL, VAR, an error correction. In the selection process, it has become clear that the number of lags specified by information criteria may not have the best forecasting accuracy.

The question of preference for complexity versus simplicity has not been answered by the results of this thesis. The ADL model which can be seen as the halfway point between simplicity and complexity has significantly outperformed all other models. The circumstances of the research may have contributed towards this as a forecast horizon in steps of 1 to 6 quarters ahead may not be enough for the dynamic models to show off their true potential. This is however the cost of an analysis over a broad spectrum.

The research results also reveal that selecting the most accurate forecasting model is sensitive to arbitrary choices made in the research. The selection of variables, lags, and forecast measures may all influence the results.

This research has introduced pairwise t-testing for significant differences in goodness-of-fit measures. This idea could be expanded upon, as a majority of forecast evaluation research does not consider hypothesis testing in judgements of forecast accuracy.

References

- Boelhouwer, P., Haffner, M., Neuteboom, P., & de Vries, P. (2001). *Koopprijsontwikkeling en de fiscale behandeling van het eigen huis*. Delft: OTB research institute TU Delft.
- Boumeester, H. (2014). *Nadere analyse verloop en betekenis van de Eigen Huis Marktindicator*. Delft: OTB - TU Delft.
- Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control, revised ed.* Holden-Day.
- Brooks, C. (2014). *Introductory econometrics for finance*. Cambridge university press.
- Brooks, C., & Tsolacos, S. (2000). Forecasting models of retail rents. *Environment and Planning A*, 32(10), 1825-1839.
- Case, K. E., & Shiller, R. J. (1988). The efficiency of the market for single-family homes.
- Case, K. E., & Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. *Real Estate Economics*, 18(3), 253-273.
- Case, K. E., Quigley, J. M., & Shiller, R. J. (2005). Comparing wealth effects: the stock market versus the housing market. *Advances in macroeconomics*, 5(1).
- Chaplin, R. (1999). The predictability of real office rents. *Journal of Property Research*, 16(1), 21-49.
- Dam, F. van, F. Daalhuizen, C. de Groot, M. van Middelkoop en P. Peeters (2013). *Vergrijzing en ruimte*. Den Haag: Planbureau voor de Leefomgeving.
- de Haan, J., van der Wal, E., & de Vries, P. (2009). The measurement of house prices: A review of the sale price appraisal ratio method. *Journal of Economic and Social Measurement*, 34(2, 3), 51-86.
- De Wit, E. R., Englund, P., & Francke, M. K. (2013). Price and transaction volume in the Dutch housing market. *Regional Science and Urban Economics*, 43(2), 220-241.
- Dröes, M., & van de Minne, A. (2015). Time-varying determinants of long-run house prices. *ASRE working paper*.
- End, J. W., & Kakes, J. (2002). *De samenhang tussen beurskoersen en huizenprijzen*. Netherlands Central Bank, Monetary and Economic Policy Department.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), 383-417.
- Francke, M., & van de Minne, A. (2013). The effects of demographic changes and supply constraints on Dutch housing prices. *ASRE Research Papers*.

- Francke, M., Vujic, S., & Vos, G. A. (2009). Evaluation of house price models using an ECM approach: the case of the Netherlands. manuscript University of Amsterdam.
- Gallin, J. (2006). The long-run relationship between house prices and income: evidence from local housing markets. *Real Estate Economics*, 34(3), 417-438.
- Huang, M. (2014). Bubble-like housing boom–bust cycles: Evidence from the predictive power of households' expectations. *The Quarterly Review of Economics and Finance*, 54(1), 2-16.
- Kranendonk, H., & Verbruggen, J. (2008). *Are houses overvalued in the Netherlands?* (No. 200). CPB Netherlands Bureau for Economic Policy Analysis.
- Liow, K. H. (2006). Dynamic relationship between stock and property markets. *Applied Financial Economics*, 16(5), 371-376.
- NVM, (2017). *Analyse Woningmarkt 1^e kwartaal 2017*. Available at nvm.nl
- 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.
- Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4), 461-465.
- Rapach, D.E. and Strauss, J.K. (2007) Forecasting real housing price growth in the Eighth District states. *Regional Economic Development*, Vol.3, No.2, pp.33-42.
- Schilder, F. P. W. (2012). Essays on the economics of housing subsidies. Thela Thesis.
- Schnure, C. (2005): Boom-bust cycles in house: The changing role of financial structure. IMF Working paper no. 05/200, International Monetary Fund: Washington DC.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1-48.
- Stevenson, S., 2007. A comparison of the forecasting ability of ARIMA models. *Journal of Property Investment & Finance* 25 (3), 223-240.
- STOCK, J. H., & Watson, M. W. (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature*, 41, 788-829.
- Sutton, GD. 2002: Explaining changes in house prices. *BIS quarterly review*, Bank for International Settlements: Basel, Switzerland, September, pp 46–55
- Takáts, E. (2012). Aging and house prices. *Journal of Housing Economics*, 21(2), 131-141.
- Tse, R.Y.C. (1997) An application of the ARIMA model to real-estate prices in Hong Kong. *Journal of Property Finance*, Vol.8, No.2, pp.152-163.

Verbruggen, J. P. (2005). *Welke factoren bepalen de ontwikkeling van de huizenprijs in Nederland?* (No. 81). Centraal Planbureau.

Wheaton, W. C. (1990). Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy*, 98(6), 1270-1292.

Appendices

Appendix 1.

Table 1. Descriptive statistics

Variables	Obs	Mean	Std. Dev.	Min	Max
Consumer price index	88	85.0	10.52	67.2	100.7
Percentage inflation	88	1.90	0.90	0.00	4.39
House price index CBS	88	81.6	20.59	36.6	106.7
House price index NVM	88	198.1	43.44	98.6	250.8
Mortgage rate (%)	88	4.24	1.23	1.97	7.79
Annualized Agreed Rate (%)	88	2.36	1.61	-0.30	5.13
Long term interest rate (%)	88	3.77	1.70	0.05	7.56
Short term interest rate (%)	88	2.36	1.61	-0.30	5.13
Term spread (%)	88	1.42	0.90	-0.50	3.54
GDP (in billions)	88	540.28	116.35	320.02	700.94
GDP per household	88	88479	6213	73576	100278
Mode income	88	29600	4426	22235	36875
Total mortgage (in millions)	88	480401	170351	176759	670473
Stock (index)	88	881.6	282.7	280.8	1435.2
Unemployment rate (%)	88	5.39	1.46	2.80	8.50
Labour (in thousands)	88	8323	490	7339	8958
Transactions	88	43540	9965	22111	60874
Construction costs (index)	88	117.3	17.43	89.5	143.8
Building permits	88	2863	1027	1008	4940
Propensity to buy	88	9	2.35	5	14
Consumer confidence	88	-7	18.58	-40	26
Houses (in thousands)	88	7011	406	6253	7677
Households (in thousands)	88	7135	378	6469	7770
Population (in thousands)	88	16285	464	15424	17056
Real HPI CBS	88	95.59	18.44	54.82	120.16
Real HPI NVM	88	232.7	36.13	147.7	284.6
Real deseasonalized HPI NVM	88	232.7	35.87	149.1	282.9
Real mortgage rate (%)	88	2.34	1.15	0.18	5.39
Real AAR (%)	88	0.45	1.37	-2.75	3.21
Real long term interest (%)	88	1.87	1.50	-1.22	5.16
Real short term interest (%)	88	0.45	1.37	-2.75	3.21
Real GDP (in billions)	88	632.85	69	476.84	726.23
Real GDP per household	88	88479	6213	73576	100278
Real mode income	88	34964	1053	33258	36955
Real total mortgage (in millions)	88	552821	144856	264744	716969
Real stock (index)	88	1036.1	296.9	418.4	1638.3
Real construction costs (index)	88	138.6	4.42	131.6	147.1
Δ log real HPI CBS	87	0.0063	0.0176	-0.0458	0.0500
Δ log real HPI NVM	87	0.0057	0.0166	-0.0409	0.0501
Δ log deseasonalized HPI NVM	87	0.0057	0.0159	-0.0367	0.0476
Δ log real mortgage rate	87	-0.0164	0.3617	-1.0488	2.0296
Δ real AAR	87	-0.0426	0.4972	-1.8330	1.2274
Δ real long term interest	87	-0.0641	0.4770	-1.4660	1.1570
Δ real short term interest	87	-0.0426	0.4972	-1.8330	1.2274
Δ term spread	87	-0.0214	0.3985	-0.9214	1.9963
Δ log real GDP	87	0.0043	0.0098	-0.0316	0.0264
Δ log real GDP/HH	87	0.0022	0.0098	-0.0340	0.0245
Δ log real mode income	87	0.0012	0.0064	-0.0142	0.0164
Δ log real total mortgage	87	0.0106	0.0127	-0.0145	0.0351
Δ log real stock	87	0.0133	0.1192	-0.3565	0.2768
Δ log unemployment	87	-0.0052	0.0842	-0.2231	0.2603
Δ log labour	87	0.0023	0.0026	-0.0046	0.0074
Δ log transactions	87	0.0079	0.2048	-0.5021	0.4407
Δ log construction costs	87	0.0008	0.0077	-0.0169	0.0208
Δ log building permits	87	-0.0059	0.1800	-0.5561	0.3536
Δ log propensity to buy	87	-0.0033	0.1205	-0.3365	0.3365
Δ consumer confidence	87	0.08	6.63	-18	16
Δ log houses	87	0.0024	0.0006	0.0013	0.0036
Δ log households	87	0.0021	0.0005	0.0007	0.0033
Δ log population	87	0.0012	0.0004	0.0004	0.0019

Appendix 2.

Figure 1a. ACF of $\Delta \log$ house price CBS

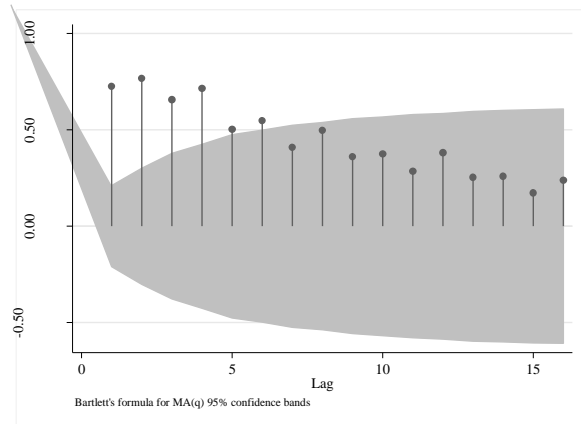


Figure 1b. PACF of $\Delta \log$ house price CBS

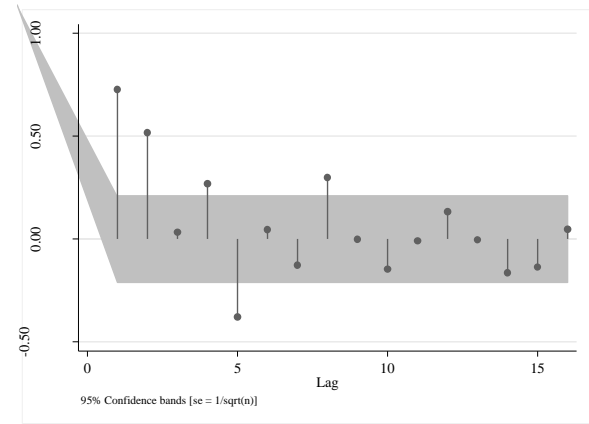


Figure 1c. ACF of $\Delta \log$ house price NVM

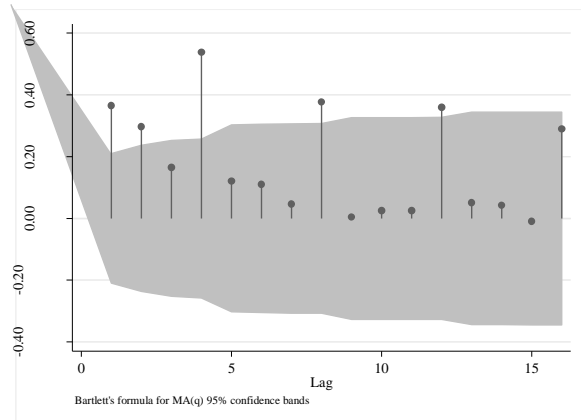


Figure 1d. PACF of $\Delta \log$ house price NVM

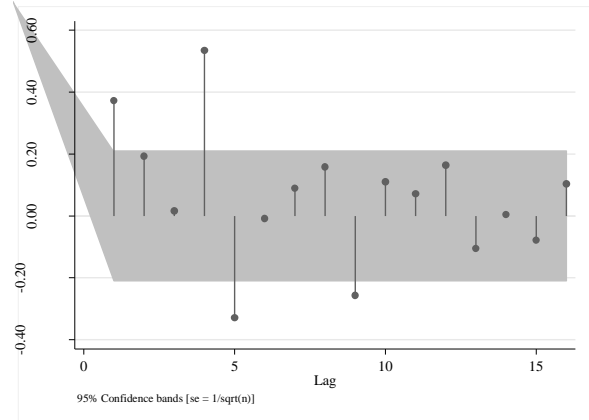


Figure 1e. ACF of deseasonalized $\Delta \log$ house price NVM

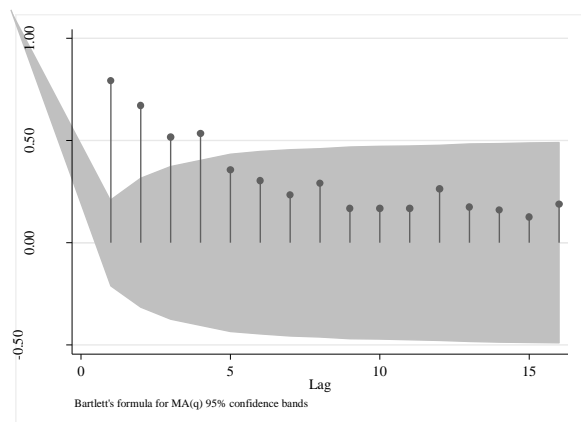
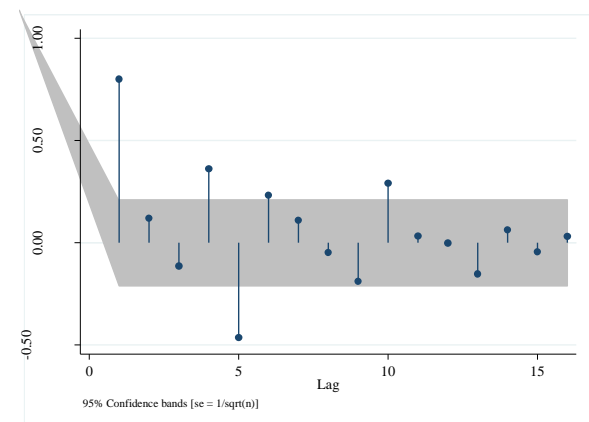


Figure 1d. PACF of deseasonalized $\Delta \log$ house price NVM



Appendix 3.

Table 2. Dickey-Fuller test statistics

Variables	DF-test statistic	P-value
Real HPI CBS (trend)	-1.526	0.8202
Real HPI NVM (trend)	-2.052	0.5728
Real mortgage rate (trend)	-2.840	0.1827
Real AAR (trend)	-2.263	0.4547
Real long term interest (trend)	-2.729	0.2241
Real short term interest (trend)	-2.263	0.4547
Term spread (trend)	-2.256	0.4586
Real GDP (trend)	-1.209	0.9087
Real GDP per household (trend)	-1.261	0.8972
Real mode income (trend)	-2.025	0.5877
Real total mortgage (trend)	1.660	1.0000
Real stock (trend)	-2.245	0.4643
Unemployment rate (trend)	-2.405	0.3768
Labour (trend)	-0.832	0.9628
Transactions (trend)	-0.530	0.9822
Real construction costs (trend)	-1.912	0.6484
Building permits (trend)	-4.101	0.0063
Propensity to buy (trend)	-3.921	0.0114
Consumer confidence (trend)	-1.370	0.8695
Houses (trend)	-4.475	0.0017
Households (trend)	-1.300	0.8880
Population (trend)	-1.383	0.8658
$\Delta \log$ real HPI CBS	-3.647	0.0049
$\Delta \log$ deseasonalized HPI NVM	-3.016	0.0334
$\Delta \log$ real mortgage rate	-8.687	0.0000
Δ real AAR	-8.589	0.0000
Δ real long term interest	-7.888	0.0000
Δ real short term interest	-8.589	0.0000
Δ term spread	-6.069	0.0000
$\Delta \log$ real GDP	-6.930	0.0000
$\Delta \log$ real GDP/HH	-6.953	0.0000
$\Delta \log$ real mode income	-7.437	0.0000
$\Delta \log$ real total mortgage	-3.730	0.0037
$\Delta \log$ real stock	-8.946	0.0000
$\Delta \log$ unemployment	-8.454	0.0000
$\Delta \log$ labour	-5.528	0.0000
$\Delta \log$ transactions	-8.339	0.0000
$\Delta \log$ construction costs	-12.461	0.0000
$\Delta \log$ building permits	-12.780	0.0000
$\Delta \log$ propensity to buy	-14.054	0.0000
Δ consumer confidence	-6.937	0.0000
$\Delta \log$ houses (trend)	-2.695	0.2380
$\Delta \log$ households	-3.058	0.0298
$\Delta \log$ population (trend)	-1.308	0.8860

Appendix 4.

Table 3a Correlation diagram of CBS model variables

	$\Delta \log$ real HPI CBS	Propensity to buy	$\Delta \log$ real mode income	$\Delta \log$ unemployment
$\Delta \log$ real HPI CBS	1			
Propensity to buy	0.74	1		
$\Delta \log$ real mode income	-0.23	-0.27	1	
$\Delta \log$ unemployment	0.32	0.00	0.15	1
$\Delta \log$ transactions	0.02	0.05	-0.04	-0.18

Table 3b. Correlation diagram NVM model variables

	$\Delta \log$ real HPI NVM	Δ term spread	$\Delta \log$ real GDP/HH	$\Delta \log$ unemployment
$\Delta \log$ real HPI NVM	1			
Δ term spread	-0.13	1		
$\Delta \log$ real GDP/HH	0.43	-0.46	1	
$\Delta \log$ unemployment	-0.47	0.18	-0.07	1
$\Delta \log$ transactions	0.30	-0.11	0.06	-0.55

Appendix 5.

Figure 2a. CBS price index

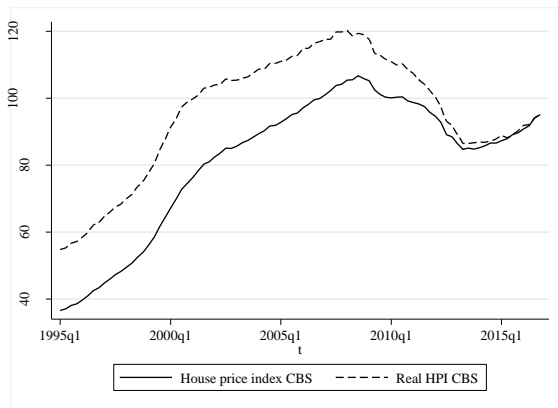


Figure 2b. NVN price index

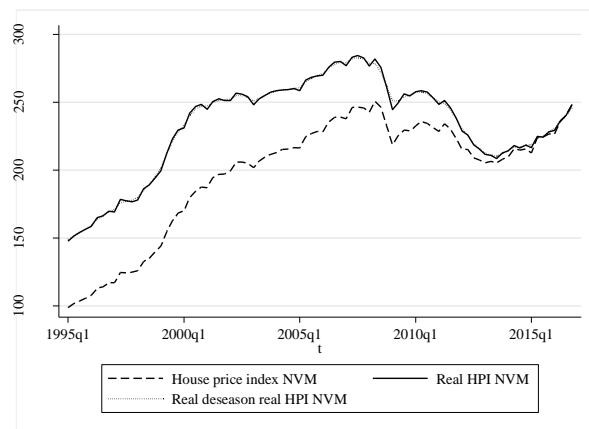


Figure 2c. $\Delta \log$ real house price CBS

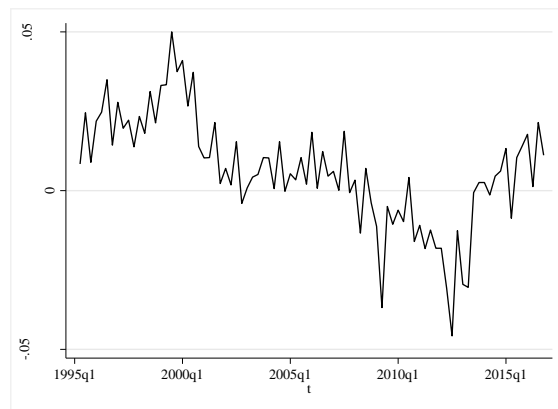


Figure 2d. $\Delta \log$ (des) real house price CBS

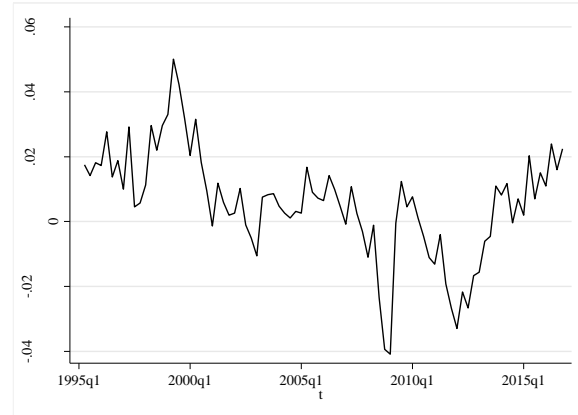


Figure 2e. Unemployment

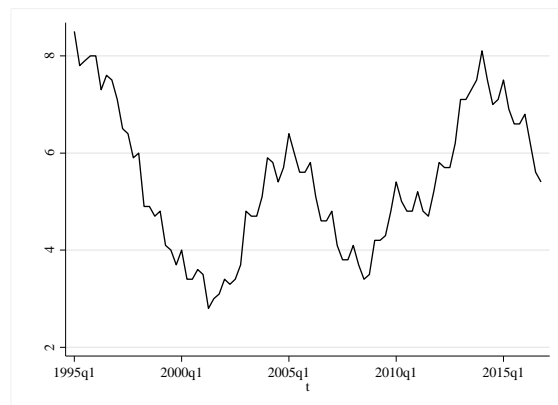


Figure 2f. $\Delta \log$ unemployment

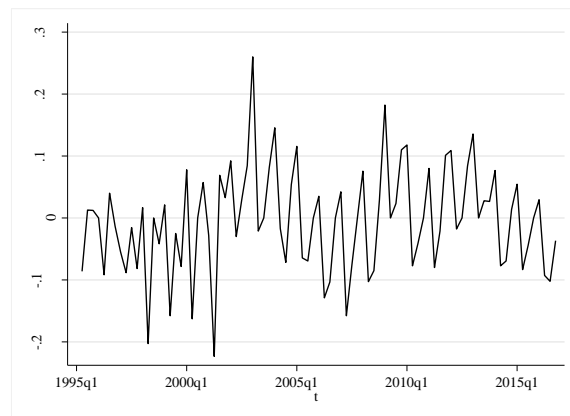


Figure 2g. Term spread

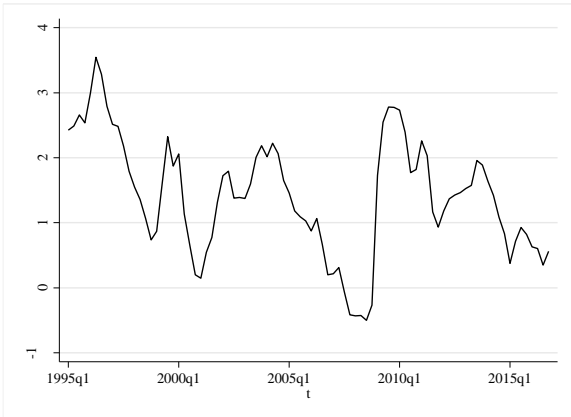


Figure 2h. Δ Term spread

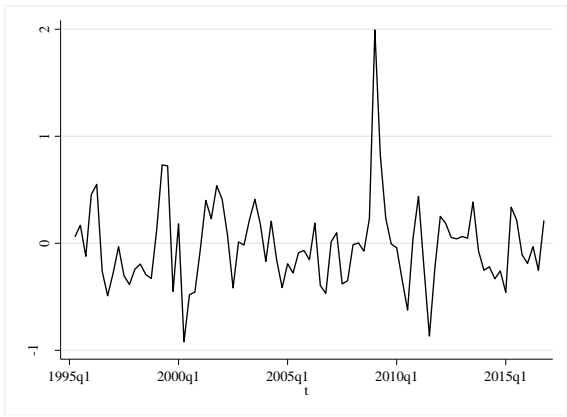


Figure 2i. Real GDP per household

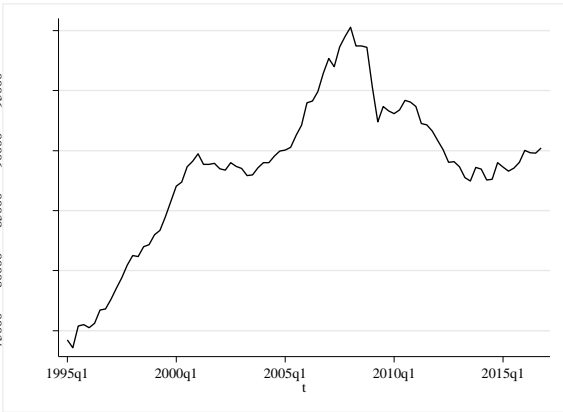


Figure 2j. Δ log real GDP per household

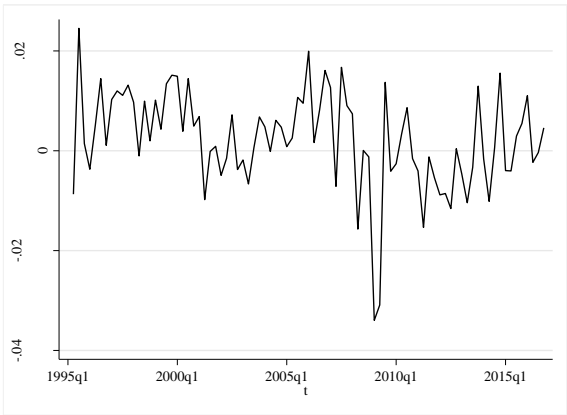


Figure 2k. (smoothed) transactions

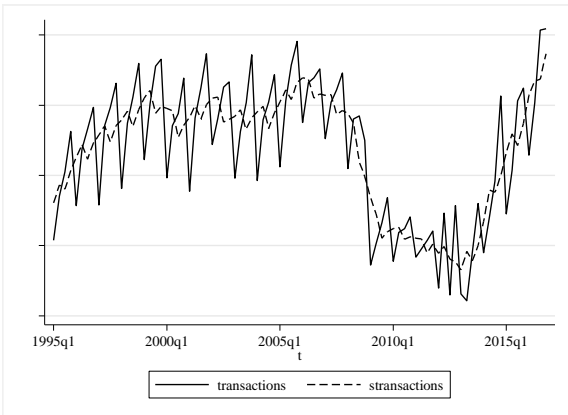
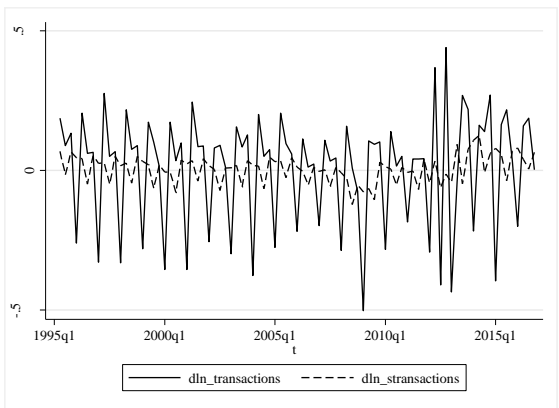


Figure 2l. Δ log (des) transactions



Appendix 6

Table 4a. 6 lag Granger causality tests with $\Delta \log$ real HPI CBS

Variables		Variables	P-value
$\Delta \log$ real HPI CBS	→	Building permits	0.1620
Building permits	→	$\Delta \log$ real HPI CBS	0.0000 *
$\Delta \log$ real HPI CBS	→	Propensity to buy	0.0060 *
Propensity to buy	→	$\Delta \log$ real HPI CBS	0.0060 *
$\Delta \log$ real HPI CBS	→	Houses	0.2420
Houses	→	$\Delta \log$ real HPI CBS	0.0050 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ real mortgage rate	0.3110
$\Delta \log$ real mortgage rate	→	$\Delta \log$ real HPI CBS	0.0400 *
$\Delta \log$ real HPI CBS	→	Δ real AAR	0.9790
Δ real AAR	→	$\Delta \log$ real HPI CBS	0.1080
$\Delta \log$ real HPI CBS	→	Δ real long term rate	0.9660
Δ real long term rate	→	$\Delta \log$ real HPI CBS	0.2940
$\Delta \log$ real HPI CBS	→	Δ real short term rate	0.9790
Δ real short term rate	→	$\Delta \log$ real HPI CBS	0.1080
$\Delta \log$ real HPI CBS	→	Δ term spread	0.7320
Δ term spread	→	$\Delta \log$ real HPI CBS	0.4280
$\Delta \log$ real HPI CBS	→	$\Delta \log$ real GDP	0.7560
$\Delta \log$ real GDP	→	$\Delta \log$ real HPI CBS	0.0010 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ real GDP/HH	0.7180
$\Delta \log$ real GDP/HH	→	$\Delta \log$ real HPI CBS	0.0010 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ real mode income	0.0000 *
$\Delta \log$ real mode income	→	$\Delta \log$ real HPI CBS	0.0010 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ HPI total mortgage	0.1240
$\Delta \log$ HPI total mortgage	→	$\Delta \log$ real HPI CBS	0.0060 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ real stock	0.2650
$\Delta \log$ real stock	→	$\Delta \log$ real HPI CBS	0.2590
$\Delta \log$ real HPI CBS	→	$\Delta \log$ unemployment	0.0000 *
$\Delta \log$ unemployment	→	$\Delta \log$ real HPI CBS	0.0070 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ labour	0.6120
$\Delta \log$ labour	→	$\Delta \log$ real HPI CBS	0.0020 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ transactions	0.0000 *
$\Delta \log$ transactions	→	$\Delta \log$ real HPI CBS	0.0000 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ construction costs	0.1240
$\Delta \log$ construction costs	→	$\Delta \log$ real HPI CBS	0.0400 *
$\Delta \log$ real HPI CBS	→	Δ propensity to buy	0.0470
Δ propensity to buy	→	$\Delta \log$ real HPI CBS	0.0230
$\Delta \log$ real HPI CBS	→	Δ consumer confidence	0.2410
Δ consumer confidence	→	$\Delta \log$ real HPI CBS	0.0020 *
$\Delta \log$ real HPI CBS	→	$\Delta \log$ households	0.3200
$\Delta \log$ households	→	$\Delta \log$ real HPI CBS	0.0320 *

Appendix 6

Table 4b. 4 lag Granger causality tests with $\Delta \log$ real HPI NVM

Variables	Variables	P-value
$\Delta \log$ HPI NVM	→ Building permits	0.0730
Building permits	→ $\Delta \log$ HPI NVM	0.0347 *
$\Delta \log$ HPI NVM	→ Propensity to buy	0.2860
Propensity to buy	→ $\Delta \log$ HPI NVM	0.2850
$\Delta \log$ HPI NVM	→ Houses	0.1820
Houses	→ $\Delta \log$ HPI NVM	0.3660
$\Delta \log$ HPI NVM	→ $\Delta \log$ real mortgage rate	0.7840
$\Delta \log$ real mortgage rate	→ $\Delta \log$ HPI NVM	0.8870
$\Delta \log$ HPI NVM	→ Δ real AAR	0.1460
Δ real AAR	→ $\Delta \log$ HPI NVM	0.0750
$\Delta \log$ HPI NVM	→ Δ real long term rate	0.9410
Δ real long term rate	→ $\Delta \log$ HPI NVM	0.9570
$\Delta \log$ HPI NVM	→ Δ real short term rate	0.1460
Δ real short term rate	→ $\Delta \log$ HPI NVM	0.0750
$\Delta \log$ HPI NVM	→ Δ term spread	0.0010 *
Δ term spread	→ $\Delta \log$ HPI NVM	0.0400 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ real GDP	0.0010 *
$\Delta \log$ real GDP	→ $\Delta \log$ HPI NVM	0.0000 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ real GDP/HH	0.0000 *
$\Delta \log$ real GDP/HH	→ $\Delta \log$ HPI NVM	0.0000 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ real mode income	0.0740
$\Delta \log$ real mode income	→ $\Delta \log$ HPI NVM	0.0080 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ total mortgage	0.4990
$\Delta \log$ HPI total mortgage	→ $\Delta \log$ HPI NVM	0.0010 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ real stock	0.7250
$\Delta \log$ real stock	→ $\Delta \log$ HPI NVM	0.0000 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ unemployment	0.0000 *
$\Delta \log$ unemployment	→ $\Delta \log$ HPI NVM	0.0010 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ labour	0.2990
$\Delta \log$ labour	→ $\Delta \log$ HPI NVM	0.0350 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ transactions	0.0000 *
$\Delta \log$ transactions	→ $\Delta \log$ HPI NVM	0.0420 *
$\Delta \log$ HPI NVM	→ $\Delta \log$ construction costs	0.0680
$\Delta \log$ construction costs	→ $\Delta \log$ HPI NVM	0.0010 *
$\Delta \log$ HPI NVM	→ Δ propensity to buy	0.7700
Δ propensity to buy	→ $\Delta \log$ HPI NVM	0.7650
$\Delta \log$ HPI NVM	→ Δ consumer confidence	0.0010 *
Δ consumer confidence	→ $\Delta \log$ HPI NVM	0.2290
$\Delta \log$ HPI NVM	→ $\Delta \log$ households	0.4770
$\Delta \log$ households	→ $\Delta \log$ HPI NVM	0.5090

Appendix 7

Table 5a. Information Criteria CBS model

CBS	AIC	BIC
AR	4	4
ARIMA	4,4	4,4
ADL	2	1
VAR	5	2
ECM	2	1

Table 5b. Information Criteria NVM model

NVM	AIC	BIC
AR	5	5
ARIMA	5,3	5,3
ADL	1	1
VAR	4	2
ECM	2	1

Appendix 8

Table 6b. Mean absolute forecast error of Δ log real HPI CBS

Model	Forecast horizon					
	1	2	3	4	5	6
AR (4)	0.010583	0.010804	0.011927	0.012845	0.014721	0.015000
ARMA (1,1)	0.010662	0.010831	0.012197	0.013868	0.015473	0.015393
ADL (4)	0.009450	0.010175	0.010336	0.011406	0.011352	0.011627
VAR(2)	0.008574	0.010145	0.012263	0.013690	0.015819	0.016154
ECM (2)	0.008318	0.009526	0.010534	0.011762	0.013438	0.013012

Table 6c Percentage correct sign of Δ log real HPI CBS

Model	Forecast horizon					
	1	2	3	4	5	6
AR (4)	83%	76%	68%	68%	59%	51%
ARMA (1,1)	73%	71%	71%	61%	56%	49%
ADL (4)	76%	73%	66%	63%	66%	61%
VAR(2)	90%	78%	71%	63%	59%	54%
ECM (2)	88%	80%	83%	78%	71%	71%

Table 6e. Mean absolute forecast error of Δ log real HPI NVM

Model	Forecast horizon					
	1	2	3	4	5	6
AR (4)	0.009065	0.013816	0.016882	0.017306	0.017715	0.017796
ARMA (1,1)	0.016861	0.017282	0.020033	0.019922	0.021396	0.021311
ADL (1)	0.007588	0.009520	0.013467	0.014783	0.016108	0.016613
VAR (2)	0.007118	0.010946	0.015442	0.01675	0.017850	0.018193
ECM (2)	0.008044	0.010145	0.013392	0.014888	0.016992	0.017524

Table 6e. Percentage correct sign of Δ log real HPI NVM

Model	Forecast horizon					
	1	2	3	4	5	6
AR (4)	78%	59%	46%	41%	37%	41%
ARMA (1,1)	66%	68%	49%	56%	41%	39%
ADL (1)	83%	76%	61%	54%	51%	49%
VAR (2)	83%	66%	46%	37%	20%	24%
ECM (2)	88%	78%	80%	71%	68%	71%

Appendix 9

Table 7b. P-value of pairwise t-tests average MAFE CBS

	AR (4)	ARMA (1,1)	ADL (4)	VAR(2)	ECM (2)
AR (4)	-	0.9773	0.005	0.5941	0.0003
ARMA (1,1)	0.0227	-	0.0044	0.2509	0.0001
ADL (4)	0.995	0.9956	-	0.9628	0.7244
VAR(2)	0.4059	0.7491	0.0372	-	0.0088
ECM (2)	0.9997	0.9999	0.2756	0.9912	-

Table 7c. P-value of pairwise t-tests average % correct CBS

	AR (4)	ARMA (1,1)	ADL (4)	VAR(2)	ECM (2)
AR (4)	-	0.0336	0.5	0.8184	0.9973
ARMA (1,1)	0.9664	-	0.9181	0.9644	0.9998
ADL (4)	0.5	0.0819	-	0.6723	0.999
VAR(2)	0.1816	0.0356	0.3277	-	0.985
ECM (2)	0.0027	0.0002	0.001	0.015	-

Table 7e. P-value of pairwise t-tests average MAFE NVM

	AR (1)	ARMA (1,1)	ADL (1)	VAR(2)	ECM (2)
AR (1)	-	0.9984	0.0024	0.0493	0.0116
ARMA (1,1)	0.0016	-	0.0002	0.0024	0.0003
ADL (1)	0.9976	0.9998		0.9924	0.9837
VAR(2)	0.9507	0.9976	0.0076		0.0483
ECM (2)	0.9884	0.9997	0.0163	0.9517	-

Table 7f. P-value of pairwise t-test average % correct NVM

	AR (1)	ARMA (1,1)	ADL (1)	VAR(2)	ECM (2)
AR (1)	-	0.7534	0.9991	0.1749	0.9995
ARMA (1,1)	0.2466	-	0.9901	0.1364	0.9992
ADL (1)	0.0009	0.0099	-	0.0078	0.9956
VAR(2)	0.8251	0.8636	0.9922	-	0.9954
ECM (2)	0.0005	0.0008	0.0044	0.0046	-