ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Specialization Financial Economics

Googling the Dutch Housing Market

Predicting movements in the housing market with internet search behaviour

Author: J.A. Blokland LLM
Student number: 374743
Thesis supervisor: Dr. J.J.G. Lemmen
Second assessor: Dr. J.C.M. Kil
Finish date: September 2020

PREFACE AND ACKNOWLEDGEMENTS

In this master thesis I analyze my favorite industry: real estate. Real estate is a sector where market changes are incorporated slowly. The current corona crisis illustrates this. Experts predict that the impact of the corona crisis will be reflected in housing prices, only as of nine months after the corona virus hit our society. The results of my master thesis show that the attention of potential home buyers and home sellers can be noticed months before a transaction actually takes place. It surprises me that the internet, even though it is already used for decades, is only occasionally used as a means to catch investor attention, especially in an industry where news about the market is incorporated with a huge delay. I understand that this can be complicated for individual homebuyers. Professional investors, however, could take more advantage by analyzing internet search behaviour to earn higher returns.

I want to thank my thesis supervisor dr. J.J.G. Lemmen in particular. His insights have been very helpful, and he was always available to help me when I needed him.

ABSTRACT

This master thesis focuses on the predictability of the housing market in the Netherlands. I use a sample of housing transactions in Amsterdam an Utrecht in the period January 2007 – June 2019. I find that Google search volume indices – overall – are significantly and positively correlated with the monthly average transaction price, volume and excess return. For example, an 1% increase in the search query ‘huis te koop’ leads to an increase of 0.156% of the future average house price in Amsterdam and 0.351% in Utrecht, corresponding to an increase of the future average house price of € 520.09 in Amsterdam and € 966.71 in Utrecht. An increase of 1% of the search query ‘makelaar’ leads to an increase of 1.092% of the future monthly average transaction volume in Amsterdam and 1.381% in Utrecht, corresponding to seven more monthly transactions in Amsterdam and four more monthly transactions in Utrecht. I also find that Google search volume indices are positively correlated with liquidity, suggesting that liquidity is high when there is more investor attention. The main results suggest the presence of price pressure, which is strengthened by the robustness checks: Google search volume indices are stronger correlated with the monthly average transaction price and volume during the financial crisis compared to the period after the financial crisis. Furthermore, overvaluation is significantly and positively correlated with attention. The results of this research have some useful implications. The keywords I use in this research, are positively and significantly correlated with the transaction price and are – in most of the cases – Granger causing the transaction price. This means that (retail) investors can predict the housing market with a free and user-friendly tool: Google Trends. in a market that is known for its inefficiency, this tool can therefore be very valuable.

Keywords:
Price pressure, price efficiency, investor attention, internet search behaviour, housing market

JEL Classification:
B26, C32, D84, D91, G12, G14, R31

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

During the crisis in the housing market that started in 2008, we have seen that houses were valued too high (Tu et al. 2018; Sornette & Woodard, 2010). In the period prior to the collapse of the housing market, beginning in 1998, house prices were increasing enormously. The acceleration in the rise of the prices of houses was unprecedented. The rise of housing prices eventually led to a low number of failures in the market. Low market failures resulted in reduced perceived risk and a higher availability of funds. Capital markets demanded less loan requirements, which led to an increase of the number of potential owners. The increase in potential owners led to an increase in prices. As a result, a vicious circle was originated (Van Haaren, 2020). Finally, the acceleration in house prices turned into an asset bubble. Houses were bought at a price that did not correspond to their intrinsic value. The prices could not be explained by economic factors such as construction costs (Jarsulic, 2010).

After the financial crisis, prices started to rise again (Nijskens & Lohuis, 2019). In the past five years we can observe substantial increases in prices which show similarities compared to the period prior to the collapse of the housing market in 2008. This raises the question whether houses currently are overvalued? In this thesis I describe my research on the housing market in the Netherlands and its efficiency. I do this by testing whether using Google search data (obtained with Google Trends) increases price efficiency and/or price pressure.

According to research by the International Monetary Fund, 30% of the increase in housing prices in the Netherlands in the period 1997-2007 cannot be explained by the development of fundamental factors (International Monetary Fund, 2008). From a large sample of countries, only Ireland showed a larger percentage (Kranendonk & Verbruggen, 2008). This implies that among the countries included in the sample, the risk of a downfall of house prices is highest in the Netherlands (and Ireland). The fact that a large part of the increase in housing prices cannot be explained by fundamentals may be caused by the typical owner-friendly housing market in the Netherlands. Dutch governmental policies have often encouraged people to own a house instead of renting one. Liberal policies concerning home ownership came into power from the moment Dries van Agt became prime minister (1977). Van Agt's administration stimulated home ownership (Van der Schaar, 1987). Even though there is a lack of construction of new houses, the government still encourages to own a house. This creates a mismatch between policy and market behaviour, resulting in an increasing housing shortage (Boelhouwer, 2005). The Dutch government has encouraged people to own a house in various ways for years (Tu et al., 2018). The mortgage interest relieve program is the most important one. The Dutch government makes it possible for homeowners to deduct their mortgage interest payments from their taxable income

(Boelhouwer & Hoekstra, 2009). The reduction in tax payments makes it attractive to own a dwelling. The mortgage interest tax relief system in the Netherlands is not unique in Europe. However, comparable schemes in other European countries are less generous. Additionally, in the Netherlands the loan-to-value ratio, which reflects the loan compared to the value of the house, is high as it is over 100 percent. It is the highest loan-to-value ratio in Europe. In other European countries this ratio is below 100 percent, meaning that homeowners need to bring in own cash in order to buy a house (Boelhouwer & Hoekstra, 2009). Also, the Netherlands has a high household indebtedness compared to gross disposable income (Nijskens & Lohuis, 2019). In 2019, this ratio was 216.52%. Only in Denmark this ratio is higher (Statista, 2020b). Another policy in the Netherlands that attracts new homeowners is the possibility to take out an interest-only mortgage. This mortgage allows the borrower to only pay the interest due at the end of the month. In 2017, 55% of the population used this type of mortgage. *De Nederlandsche Bank* (Dutch Central Bank; DNB) has stated that this percentage is too high since the risk of households not being able to redeem their mortgage is high (DNB, 2017). The policies to encourage homeownership lead to detrimental risks. Encouraging homeownership can lead to overvalued prices. Besides that, the high loan to value ratio implies that banks in the Netherlands are facing default risks towards relatively larger loans, compared to other European countries. The high percentage of interest-only mortgages signals a higher probability of bad debts. Moreover, certain other characteristics of the Dutch housing market can create a bubble. Glaeser et al. (2008) find that the chance of a bubble taking place is more likely for housing markets where less land is available for development. This could be the case in a densely populated country like the Netherlands.

The housing market is not known for its efficiency. Case and Shiller, for example, find that information tends to be incorporated into housing prices with a lag. Furthermore, they find that additional forecasting variables (rising construction costs, real income growth per capita, and increasing adult population) only explain a small part of the variation, causing upward movements of housing prices, meaning that housing prices are inefficient (Case & Shiller, 1990).

There are various explanations for the lack of price efficiency in the housing market. Firstly, the housing market is 'ruled' by information asymmetry. It is almost inevitable that sellers of properties have more information than buyers (Levitt & Dubner, 2005). Secondly, market participants often lack rationality. Housing prices are prone to various behavioural biases (Salzman & Zwinkels, 2017). Investor sentiment is an important factor in property decision making (Gallimore & Gray, 2002). Thirdly, the housing market depends on liquidity. This list of explanations is non-exhaustive, i.e. also other explanations can be thought of. For example, institutional changes, such as favourable mortgage regulations, can explain why people buy more houses in a certain period (Shiller, 2007).

As shown, a relatively high level of inefficiency is present in the housing market. Besides that, the housing market is prone to price pressure. The theory of price pressure, which is intertwined with the efficiency theory, is related to the fact that buyers only buy assets (e.g. houses) that catch their attention. On the other hand, sellers only sell their house when the attention in the housing market is positive (Genesove & Mayer, 2001). Observable measures that are associated with buyer attention are news, unusual trading volume and extreme returns. An event that catches attention is often captured by the news. Consequently, the transaction volume increases when the news reaches the house buyers. Meaningful news will often affect buyers' beliefs, resulting in more houses being bought than usual. Important news may lead to significantly positive or negative returns (Barber & Odean, 2008). The price pressure theory challenges the efficient market theory because it suggests that investors can earn excess returns by taking advantages of overpricing or underpricing without taking any additional risks (Barberis et al., 1998).

The theories of price efficiency and price pressure lead, in relation to investor attention, to different outcomes. From a price efficiency point of view, investor attention leads to an increase in price efficiency since investor attention facilitates the information incorporation into prices, which results in better market efficiency. From a price pressure perspective, investor attention leads to a drift towards certain assets that are prone to attention while other assets are neglected. Price pressure therefore leads to price inefficiency. So, according to the price efficiency theory, investor attention mitigates the underreaction effect to news while the price pressure theory states that investor attention may strengthen the overreaction effect to news. However, one could also argue that the detection of price pressure eventually leads to an increase in price efficiency since the investor is now aware of price pressure and therefore can anticipate this information.

As of April 2020, 4.57 billion people were actively using the internet (Statista, 2020a), which is a substantial percentage of the world population. Analyzing internet search behaviour can therefore provide a lot of accurate and representative information on whether house prices are over- or undervalued and whether house prices are prone to sentiment and attention. The analysis of internet search behaviour can be conducted by the use of Google Trends, which is a tool that analyses Google search data. With Google Trends, data of search queries and their search volume can be obtained in the form of a Google Search Volume Index (GSVI) (Braun, 2016). In other research, Google Trends already proved to be a valuable instrument. Hu et al. (2020), for example, find by analyzing Google search data that governments should strengthen the public awareness of COVID-19 nationally to reinforce the public alertness and sensitivity to the corona virus (Hu et al., 2020). Research about the relation between search behaviour on Google (as a proxy for sentiment) and housing prices is limited (Meshcheryakov, 2018). My thesis can therefore be a relevant contribution to the literature about price efficiency and price pressure in the housing market.

In this master thesis, I investigate whether it is possible to detect price movements in the Dutch housing market using Google search data. In order to do so, Google Search Volume Indices are used as independent variables in a regression model to test whether search behavior has an impact on the monthly average transaction price, volume and excess return. The research question in this thesis is:

Is internet search behaviour a good predictor of movements in the Dutch housing market?

Two main datasets are collected: a dataset of the search queries from Google and a dataset of transaction prices in the Netherlands. The collected sample is a sample between January 2007 – June 2019. The dataset of search queries contains a list of search queries such as *'funda'* (the largest internet platform in the Netherlands where people sell their houses), *'huis te koop'* (house for sale) and *'hypotheek'* (mortgage) which are transformed by Google Trends to search statistics in the form of a Google Search Volume Index. The data is a time-adapted index with values between 0 and 100, where 100 is the highest value in GSVI, meaning that search volume was highest in this period (Braun, 2016). The output looks as follows: for each search query Google provides a timeseries with the preferred frequency (Meshcheryakov, 2018).

The dataset of transaction prices in the Netherlands is obtained from the Dutch Realtor Association (*Nederlandse Coöperatieve Vereniging van Makelaars en Taxateurs in onroerende goederen U.A. NVM*). This dataset contains variables such as the list price, the transaction price and the number of days a dwelling is on the market. Next, I transform the variables into standardized variables, which means that the values are rescaled in such a way that the mean is zero and the standard deviation is one. With standardized values it is easier to compare variables with each other. Additionally, I smooth the dataset out for seasonality. An example of this is displayed in figure 1. The trend line of Google searches on 'Funda' is plotted together with the house price development in Utrecht and Amsterdam.

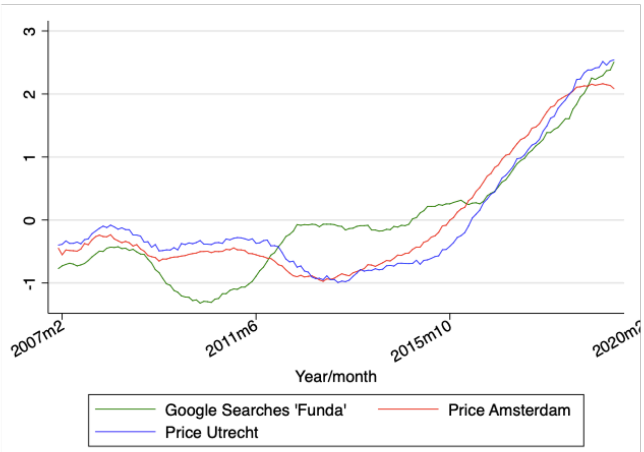

Figure 1. Google Searches and House Prices.

The Google Trends data are provided as a time series with a certain frequency. Before I start processing the data, I apply an Augmented Dickey-Fuller test for non-stationarity (Dickey & Fuller, 1979). I also apply a normality test to control for skewness and Kurtosis. Data that is too skewed or too peeked, is transformed to ln-values.

Sometimes it is difficult to decide in which direction causality goes. This may also apply on this research. The causality problem often arises when information is slowly incorporated or because possible causal variables are omitted (Granger, 1969). I therefore firstly perform a Granger causality test to find out whether Google search behavior affects movements in the housing market or vice versa (van Dijk & Francke, 2018).

I apply an autoregressive regression model to test whether search behavior has an impact on housing prices (inspired by van Veldhuizen, Vogt & Voogt, 2016). I regress the monthly change in GSVI on the monthly average transaction price, transaction volume and excess return. I also regress the monthly change in GSVI on different liquidity measures. In this model I test whether GSVI has statistical power in forecasting house prices. Next, I test a model with control variables. I test for the population growth, the real GDP growth, the number of households, the mortgage interest rate, the average rent and the CPI price index. I also apply an efficiency test with a weak-form test, by regressing the lagged values of the search queries on the dependent variables. With the use of this analysis I can detect whether historical values of the search query have an impact on the transaction price.

I find that most of the search queries are Granger causing the monthly average transaction price. Google search volume indices are significantly correlated with the monthly average transaction price, volume and excess return. Adding the control variables leads to a small decrease in significance. Transaction prices and volumes are higher when there is more attention on the internet. Some lagged values of the search queries have a significant relation with the monthly average transaction price, volume and excess return. This implies that a random walk is not present and that prices are not fully efficient. I find that the monthly average transaction price, trading volume and excess return are positively correlated with liquidity. I also find that search queries are positively and significantly correlated with the liquidity measures, suggesting that liquidity is higher when investor attention is higher. The results suggest the presence of price pressure. This suggestion is strengthened by the robustness checks. Attention is more significantly correlated with the monthly average transaction price and volume during the financial crisis, compared to the period after the financial crisis. Furthermore, attention is positively and significantly correlated with the difference between transaction price and the WOZ value (valuation of the dwelling by the local government) and therefore overvaluation. Google search volume indices thus seem to be good predictors of movements in the housing market.

The key contribution of this master thesis is that I use a unique dataset of transaction prices in the Netherlands. As far as I know, research on the relation between Google search behaviour (GSVI) and movements in the housing market in the Netherlands using NVM data has never been done before. Interesting data about the spread between list and transaction data can be connected to liquidity and liquidity related search queries. Besides that, as stated earlier, research on the relation between internet search behaviour and housing prices in general is limited. Van Dijk and Francke (2018) test the relation between internet search behavior, housing prices, and liquidity in the Dutch housing market. However, instead of using Google search data as proxy for internet search behavior, they use the number of clicks on www.funda.nl. Furthermore, the sample period in their research is 2011-2014 and is therefore shorter. Van Veldhuizen, Vogt and Voogt (2016) also test the relation between search behavior and transactions on the housing market, but they have used data from the Dutch Central Statistics Office instead of the NVM. The implications of my master thesis are useful for decision-making of house buyers and investors. (Retail) investors can predict the housing market with a free and user-friendly tool: Google Trends. in a market that is known for its inefficiency, this tool can therefore be very valuable.

The next chapter is dedicated to the literature review in which I discuss the theories of price efficiency and price pressure and how the housing market is related to these theories. In the third chapter, I discuss the two main data sources and how the data is collected. In the fourth chapter, I explain the methods. Besides explaining the main regression model, I also explain other import tests such as the efficiency tests and the Granger causality tests. In the fifth chapter, I describe the results. In the sixth chapter I focus on the interpretation of the results. I conclude in chapter seven. In chapter seven, I also touch upon some limitations of this master thesis.

CHAPTER 2 Literature review

2.1 Introduction

According to the literature, the housing market has often been subject to over- or undervaluation. In contrast to what academics believe, housing markets are not efficient (Salzman & Zwinkels, 2017). Market inefficiency is caused by information asymmetry, irrational decision-making and illiquidity. Also, price pressure can lead to a disordered housing market. In this chapter, I discuss the theoretical background of my research. First, I discuss the price efficiency theory. Next, I describe the price pressure theory. Thereafter, I elaborate on the characteristics of the housing market and how price efficiency and price pressure are connected to these characteristics.

2.2 Price efficiency

2.2.1 Definition

An efficient market provides signals for resource allocation (Fama, 1970). Investors on the stock market decide whether they should buy a certain stock and buyers in the housing market decide whether it is a good decision to buy or to rent a house. These decisions are made under the assumption that the prices of products, stocks and houses fully reflect all available information. Various definitions of an efficient market can be given. One of the well-known modern economists, Eugene Fama, came up with a widely accepted definition. Fama describes a market that contains a lot of market participants and active competition. Information is (almost) costlessly available. In these circumstances an efficient market should lead to prices that fully reflect all available information (Fama, 1965). Three main conditions of market efficiency are derived from Fama's definition. Firstly, there are no transaction costs involved when assets are traded. Secondly, all available information is freely available for all market participants. Lastly, all market participants agree on the implications of the current information for the current price and the distribution of the future prices. These conditions, however, are not necessary. Even the price of a particular transaction in which transaction costs are involved can fully reflect all necessary information (Fama, 1970). Although these definitions are not necessary for market efficiency, they are often a source of it.

Efficient market models also assume that information is directly incorporated in asset prices. The words *fully reflect* – used by Fama – are an example of this implication. These models both assume that diffusion of public information is instantaneous and that investors directly act upon this information (Merton, 1987). Individual events causing an over/underreaction can take place, but on the long run these reactions will be in equilibrium (Fama, 1998). Efficient market models also assume that investors act rationally in their decision-making process (Kishore, 2006).

Following the classic price efficiency theory of Fama (1965), investment market return follows a random walk pattern. This means that returns would not be predictable. This is explained by the arbitrage theory, meaning that rational investors react when prices deviate from their intrinsic value. In this way a market equilibrium is maintained. Prices reflect all available information. Additionally, there is no *free lunch*, meaning that investors cannot earn excess returns greater than the returns warranted by its risk. When buying a house, this means that the transaction price is the reflection of all available information and that the homebuyer cannot earn an abnormal return in the future.

2.2.2 Attention

In order to process information instantaneously, investors need to pay full attention to all information that is provided. In reality, investors have limited attention. One cannot observe every piece of information. We experience selective attention, meaning that we might think that we absorb all information but in fact we only obtain information about the details we are focused on (Hirshleifer et al., 2009). This can be explained by psychological arguments. Humans have selective attention and this selective attention is governed by different rules and are clarified by different mechanisms. Also, the intensity and amount of attention differs among humans. This implies selection, i.e. not all human beings pay attention to the same things. Human beings process information differently. It is easy to pay attention to one object, but it is harder to pay attention to multiple objects (Kahneman, 1973).

Attention is not only about selection, but also about amount and intensity. Different mental activities require different demands from the mental capacity. A decision about which can of Coca Cola one should buy in the supermarket is easier than a decision about which house one should buy. When supply of attention does not match demand, performance may not be optimal. Additionally, performance may fail because the supply of information was not sufficient (Kahneman, 1973). This is also applicable in the content of buying a house. A lot of information needs to be processed. A potential homebuyer can choose between dozens or maybe even hundreds of houses. And when the choice for a particular house is made, still a lot of information needs to be processed. What are the characteristics of the house? Does the house fulfil the requirements to accommodate a whole (future) family? What is a fair price? How does the housing market function? How attention is connected to price efficiency, is explained in the next section.

2.2.3 Attention and price efficiency

Even if the conditions mentioned in the previous section apply, the required information first needs to attract investors before prices *fully reflect* all necessary information (Ben-Rephael et al., 2017). Prices can only fully reflect all the information if investors pay attention to the information provision. Some authors argue that asset prices adjust slowly to new information, implying that one must investigate returns over a long horizon to get a picture of market efficiency (Fama, 1998). In reality, investors have limited attention. The existence of freely available public information does not directly imply that information is immediately reflected in asset prices. How quickly information is incorporated depends on the channels that investors have at their disposal and which channels they actually use (Drake et al., 2012). A lack of investor attention implies that prices are often prone to underreaction to news. Different proxies can be used when testing the impact of investor attention on price efficiency. Barber and Odean (2008) use extreme returns and trading volume to proxy investor attention. Ben-Rephael et al. (2017) use news articles and search frequency as proxy for investor attention. Da et al. (2011) and Drake et al. (2012) use search frequencies in Google.

Dellavigna and Pollet (2009) find that investors on the stock market react less often on announcements made on Friday compared to announcements made on other weekdays. Additionally, the abnormal trading volume around the announcement day is 10 percent lower for Friday announcements than for other weekdays announcements. Eventually, investors detect the mispricing and adapt the information. These findings suggest that investors are distracted from work-related activities before the weekend and that efficiency increases with investor attention. A comparable conclusion can be drawn from the research done by Hirshleifer et al. (2009). They find that in periods when a relatively large number of firms is releasing new information, stock prices react less heavily compared to periods when a smaller number of firms is releasing new information. Limited investor attention through distraction therefore leads to underreactions on the market. Also, Ben-Rephael et al. (2017) find that prices underreact to news releases when investors fail to pay sufficient attention. Ben-Rephael et al. draw a clear difference between institutional and retail attention. Institutional attention, measured with search behavior on Bloomberg terminals, enhances price efficiency. Retail attention, measured with Google search behavior, does not enhance the incorporation of information.

Drake et al. (2012) test the impact of information demand on the market response to earnings news by analyzing Google search data. In the first phase, they examine the timing of Google searches around earnings announcements. They find that the demand of information on Google starts to increase two weeks prior to the earnings announcements. The abnormal search volume is increasing as the announcement date is approaching and experiences a spike at the earnings announcement date and reverts to its normal level in two weeks after the announcement date. Next, Drake et al. investigate which factors influence investor information demand. They do this by regressing abnormal search volume on potential explanatory variables. Abnormal search volume is positively correlated with firm-specific coverage by the press, which implies that Google search data is a complement to alternative news channels. Furthermore, investors seek more information when they trade in illiquid assets such as real estate. Other findings are that abnormal search volume is higher for higher spreads and higher idiosyncratic volatility. These results imply that investors seek more information when costs of information are high (i.e. for illiquid assets) and when the potential benefit is higher.

In the second phase, Drake et al. examine the impact of information demand on the relation between market activity and earnings information. They find that when investors demand more information in the period prior to an announcement, market activity reflects more upcoming earnings information in the period before the announcement. Consequently, the market response on the announcement date is less significant. These results suggest that the market reaction on announcements is partly preempted when the demand for information in the period prior to the announcement is high. The findings of Drake et al. also imply that markets are actually not efficient, since it costs time and money for investors to find information. However, price efficiency is increasing with information demand before the announcement date. The results cannot prove that the increase in price efficiency is directly related to Google search behavior. The research, however, displays that investors use internet to obtain information about the market.

The research cited in this section was done by analyzing stock price performances. The results, however, could also apply on the real estate sector. Maybe the results apply even more since investors seek more information when they trade in illiquid assets (Drake et al., 2012). Potential homebuyers- and sellers use the internet to obtain information about the housing market. News releases related to the housing market, for example news related to the mortgage interest rate, may lead to an increase in search activity on the internet.

2.3 Price pressure

2.3.1 Definition

Another theory, is the theory of price pressure. The price pressure theory can be viewed as a behavioural finance alternative to the theory of price efficiency (DeBondt & Thaler, 1985). This theory focuses on the fact that buyers only buy assets (e.g. houses) that catch their attention. Barber and Odean (2008) mention that investors (buyers), especially retail investors, face a large decision problem when they want to buy an asset. Gathering information is beneficiary. A potential homebuyer can choose between hundreds of options. Sellers on the other hand, only have to sell their own assets. Sellers will only sell their house when the housing market is subject to positive attention (Genesove & Mayer, 2001). High retail investor attention intensifies price overreactions (Huo et al., 2006).

Overconfidence and investor attention are related. Two types of overconfidence can be distinguished in the literature. First, overconfidence relates to investors who overestimate the precision of their private information signals and who underestimate the information signals that are publicly received (Daniel et al., 1998). The second type of overconfidence refers to the phenomenon that investors belief that success in assets' performance is caused by their skill and personal characteristics, (self-attribution bias). These investors do not regard luck as part of their success and overestimate their own abilities (Daniel et al., 1998). They believe that they are better than the average investor (Malmendier & Tate, 2005). Overconfidence is greater in small and illiquid markets, such as the housing market. Shiller (2016), for example, finds that overconfidence is a significant determinant of behaviour in the housing market. The explanation of this phenomenon is that it is cheaper and easier to obtain information about large and liquid assets than information about small and illiquid assets.

Observable measures that are associated with buyer attention are news, unusual trading volume and extreme returns. An event that catches attention is often also captured by the news. Consequently, the transaction volume increases when the news reaches the potential house buyers. Significant news will often affect buyer's beliefs, eventually resulting in more houses being bought than usual. Important news may lead to significant positive or negative returns (Barber & Odean, 2008).

So, following the price pressure theory, assets like real estate are prone to overreaction when they receive a lot of investor attention. Investors are biased to one direction which means that the release of positive information leads to positive returns in the future. Current positive news can therefore attribute in predicting positive excess returns in the future (Barberis et al., 1998). One can observe investor sentiment, referring to a set of beliefs about future capital gains that are not necessarily justified by the intrinsic value of the asset (Baker & Wurgler, 2007).

2.3.2 Evidence

Barber and Odean (2008) examine the relation between investor attention and the buying and selling behavior of individual investors. Three measures are used to catch investor attention: news, unusual trading volume and extreme returns. Barber and Odean find that individual investors show buying behavior based on investor attention. They are net buyers on days with high-volume trades, days with extremely negative and positive one-day returns and when firms are in the news. Barber and Odean furthermore find that the price pressure effect is larger for retail investors than for professional investors. Professionals have more time, money and resources to analyse the assets they may invest in. It is therefore less likely that they invest in assets only based on investors' attention. They will also pick assets that do not get a lot of attention. Retail investors, on the other hand, have less time and sophisticated knowledge while they have to pick assets from a large set of available alternatives (Da et al., 2011). The findings of Barber and Odean certainly apply on the housing market, where most of the market participants are retail buyers. Genesove and Mayer (2001) find that the spread between the list price and the transaction price of houses is twice as large for domestic buyers as for professional investors. They also find that trading volumes fall when house prices fall, what cannot be explained by perfect asset models. These results imply that sellers only sell when the price of their house is increasing, which they may substantiate with the fact that sellers observe the profit made earlier by other sellers.

Da et al. (2011) use Google search data to measure investor attention. They use search queries, which are translated into a Google Search Volume Index (GSVI) by Google Trends, to find a relation between investor attention and asset prices. Firstly, they investigate the relation between the GSVI and other investor attention proxies such as extreme returns, turnovers and news. In a vector autoregression framework they find that the GSVI is actually leading the alternative proxies. This implies that internet search behavior is an earlier detector of price changes than conventional investor attention proxies.

Secondly, Da et al. examine which type of attention the GSVI is capturing. They find strong evidence that the GSVI is catching retail investors attention. Lastly, Da et al. test the theory of Barber and Odean (2008). Barber and Odean believe that retail investors are net buyers of attention-grabbing assets, which therefore leads to positive price pressure of these assets. Da et al. find a strong price pressure effect with the GSVI data on the short run and a reversal effect in the long run (Da et al., 2011).

Conclusions about attention-based decision making can be generalized to a variety of economic situations such as the process of buying a house (Barber & Odean, 2008). Implications of the provided evidence are that buyers make buying decisions based on information that recently caught their attention and that the use of Google Trends can help us to detect attention-based decision making. How this exactly works, is explained in the data section.

2.4 The balance

Efficient asset prices are prices that fully reflect all available information. Evidence shows that an increase in attention leads to efficient prices since investors have more information available about the assets they might want to invest in. Distraction leads to underperformance of asset prices and therefore inefficient prices. Drake et al. (2012) find evidence that an increase in search activity on Google is a proxy for information demand. Consequently, Google search activity increases price efficiency. Other research, on the other hand, shows that investor attention is a signal of price pressure. This implies that prices deviate from their intrinsic value when assets gain a lot of attention. Da et al. (2011) find evidence that internet search behaviour catches investor attention and the price pressure effect of investor attention. Eventually, prices correct in the long run. Consequently, markets are in balance and efficient. The mentioned results do not direct in one particular direction.

Factors that influence the price pressure and price efficiency effect of investor attention on asset prices, are for example liquidity and type of investor. Cited research shows that illiquid assets are more prone to price pressure than liquid assets. Retail investors are more prone to irrational decision making than professional investors. Barber and Odean (2008) find that the impact of investor attention on extreme returns is strong for individual investors while it is not for professional investors. Ben-Rephael et al. (2017) find that the search behaviour of a professional investor is a good proxy for an *increase* in price efficiency and they explicitly state that this does not apply for retail investors. The housing market is known for its illiquidity. Besides that, the decision of buying a house is often a decision of a retail investor. In the next section I explain how this is related to the cited literature.

2.5 The housing market

2.5.1 Introduction

With all good intentions, housing markets cannot be defined as efficient according to Fama's definition of efficient markets. As discussed in previous section, the housing market is an illiquid market. Besides that, a large part of the housing market consists of retail investors. Market participants in the housing market often lack rationality (Kishore, 2006). The housing market is also prone to information asymmetry, which makes prices impossible to fully reflect all information.

Case and Shiller find that information tends to be incorporated with a lag into housing prices. Furthermore, they find that additional forecasting variables (construction costs, real per capita income growth, and increase in the adult population) only explain a small part of the variation, meaning that housing prices are inefficient (Case & Shiller, 1990). Quigley finds that a simple model of economic fundamentals only explains between 10 and 40 percent of changes in housing prices (Quigley, 1999). Farlow finds that supply-side factors and demand-side factors related to household formation and population size only explain a small part of the variation in housing prices and that they cannot explain the enormous volatility in housing prices (Farlow, 2004a). In this section I explain the characteristics of the real estate market and how they relate to the theories of price efficiency and price pressure.

2.5.2 Information asymmetry

There are various explanations of the lack of price efficiency in housing markets. One of them is information asymmetry. Gallimore et al. (2000) find that real estate investment behaviour differs from normative behaviour. Normative decision making requires a lot of information and information processing which is associated with higher costs (Salzman & Zwinkels, 2017). It is almost inevitable that sellers of properties have more information than buyers (Levitt & Dubner, 2005). Buyers only have access to superficial information and need to decide whether they buy a house only after a visual inspection of half an hour. Sellers on the other hand also have access to advanced information such as latent problems (Wong et al., 2011).

Transaction prices of comparable houses differ among each other. This can be caused by differences among buyers. Qiu, Tu and Zhao (2020) find that heterogeneity among home buyers explains the variation in transaction prices. For example, the difference between a local buyer and a non-local buyer. Non-local buyers typically pay more for a particular house than local buyers. Also, demographic characteristics of buyers, such as income, age, education, explain diversity in transaction prices. Buyers who are older, have a lower income and who have less education are willing to pay less. Other characteristics, however, can mitigate the non-buyer effect. For example, an experienced home buyer may have advanced negotiation skills.

Information asymmetry in the housing market is not only caused by heterogeneity among buyers but also by heterogeneity among the assets. Home buyers do not easily obtain information about hidden flaws such as unclean water supply or a leaking roof (Wong et al., 2011). Also, visible features cause difference in housing prices. Dwellings are almost never identical (Thwaites & Wood, 2003). Even identical terraced houses in a particular area can differ among themselves. Some of them face valuable modifications while others lack maintenance. One can therefore only partly support a bid price on historical transaction prices. In less heterogenic housing markets, meaning that willingness to pay for units with the same observable quality and characteristics varies little among buyers, buyers have little need to inspect many houses for sale for hidden flaws in search of their best match. This implies that an increase of houses for sale has only a small effect on the quality. In more heterogenic housing markets, however, house buyers have to inspect a lot of units in the search for their best match and in order to reduce their idiosyncratic risk. An increase of dwellings for sale will therefore lead to a significant quality improvement (Nenov et al., 2015). To conclude, it is hard to compare properties on their characteristics. Simply taking the mean of transaction prices is not an appropriate method to access the price of a particular dwelling (Thwaites & Wood, 2003).

2.5.3 Irrational decision making – investment function

In order to obtain efficient prices, investors have to take all available information into account and have to make their decisions rationally (Kishore, 2006). Investment decision making tend therefore to be a normative process. Market participants in the housing market often lack rationality. Investor sentiment is an important factor in property decision making (Gallimore & Gray, 2002). Herd behaviour has a big impact on the real estate market (Shiller, 2005). Additionally, the housing market has not only an investment but also a consumption function (Shiller, 2007). Market participants are not only professional investors, but also household buyers. Their approach of valuing a house can deviate from valuing a house based on fundamentals (Salzman & Zwinkels, 2017).

According to Farlow (2004b), *over-optimism* is the most important bias in the housing market. Over-optimism relates to an overly optimistic opinion about future price movements. Farlow finds that buyers believe that buying a house does not involve a great risk but they, on the other hand, also believe that house prices show a yearly increase, on average, of more than 11 percent. Farlow (2004b) argues that media strengthen overconfidence by only publishing over-optimistic information and neglecting negative information. A bias related to over-optimism, is *overconfidence.* Overconfidence relates to the underestimation of risks. Buyers evaluate different outcomes of a certain investment decision, forget about the negative outcomes and overestimate the positive outcomes. Domestic investors often are overconfident (Kishore, 2006). Shiller (2016) states that overconfidence is a significant determinant of behaviour in the housing market.

Momentum effect is the bias that refers to the phenomenon where observed price movements in the housing market are used as a bias for future price expectations. Case (2003) examines whether price increases lead to an increase in people buying a house. More than 80 percent of the examined people says that they are encouraged to buy a house by historical price increases. This finding implies that buyers neglect to take fundamentals into account. People follow rumours and the news. Media can have a big impact. The presence of the momentum effect implies the presence of *herd behaviour.* The housing boom in 2007 is considered a speculative boom. The high prices were not created by estimating the real value of properties, but by the enthusiasm of the market participants. *Herd behaviour* is one of the driving factors of this enthusiasm (Shiller, 2016). Potential buyers are convinced that prices keep increasing and feel the fear of regret of not buying a house (Kishore, 2006).

A lot more other biases can be detected in the housing market. Farlow (2004b) detects the presence of *confirmation bias* meaning that buyers attribute their success to wise investment decision making while bad investments are blamed on bad luck. *Regret theory* is the bias which explains that people make investment decisions because they want to avoid having regret for not making a particular decision (Farlow, 2004b).

As discussed earlier, efficient prices reflect all available information, implying that deviating prices move back to their intrinsic value as soon as possible. This means that future house prices are unpredictable. A free lunch is impossible. Taking into account the behavioural perspective of the housing market, these assumptions loose strength. Prices are predictable in boom periods. Case and Shiller (1990), for example, find that price changes observed in one year are persistent in the following year. Fundamentals only explain a small part of the variation in price changes. Also, Quigley (1999) and Farlow (2004a) find that fundamentals only explain a small part of changes in property prices. The earlier mentioned psychological biases such as the *momentum effect* can explain the deviation from efficient prices. An overall conclusion that can be drawn, is that the housing market is inefficient and that price pressure effect is applicable the (Dutch) housing market.

Prices are predictable and persistently deviate from their intrinsic values. The strong presence of biases as *herd behaviour* and *overconfidence* make housing markets prone to price pressure. Daniel, Hirshleifer and Subrahmanyam (1998) show that overconfidence and attention are correlated. More attention would imply higher trading volumes and increasing excess return. This I will test with the first three hypotheses.

Hypothesis 1

H_0 *Housing related Google search volume indices changes are not correlated with changes in the Dutch housing transaction prices*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the Dutch housing transaction prices*

Hypothesis 2

H_0 *Housing related Google search volume indices changes are not correlated with changes in transaction volume on the Dutch housing market*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the transaction volume on the Dutch housing market*

Hypothesis 3

H_0 *Housing related Google search volume indices changes are not correlated with changes in excess return on the Dutch housing market*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the excess return on the Dutch housing market*

After testing the first three hypotheses, I test whether price efficiency is increasing in investor attention or whether attention leads to an increase of the price pressure effect. This I do with the fourth hypothesis:

Hypothesis 4

H_0 *Internet search behaviour does not increase price efficiency on the Dutch housing market*

H_α *Internet search behaviour does increase price efficiency on the Dutch housing market*

2.5.4 Irrational decision making – consumption function

Buying a house is not only an investment function, but also a consumption function. The biases explained in previous section, mainly explain why buyers believe that prices of (their) houses will increase or decrease. Behaviour, however, can also be explained by other idiosyncratic characteristics such as emotions. It makes intuitively sense: a particular house that belongs to a certain family for centuries, is worth more to this particular family as to a random investor.

The consumption function of buying a house depends on residential mobility. Brown and Moore (1970) describe two phases of residential mobility. First, a change in factors as family structure and income must cause dissatisfaction about the current house. Thereafter, the family must decide whether they move or adjust their current dwelling. Age, income, education and professional career are also factors that determine residential mobility (Mulder & Hooimeier, 1999). Also, macroeconomic fundamentals such as the mortgage interest rate determine housing mobility (Dieleman, 2001).

The ultimate decision of buying a house is often determined by unexplainable emotions and feelings (Levy et al., 2008). Hidalgo and Hernandez (2001) find that for home buyers social connection is more important than physical characteristics of a particular house. The choice for a particular house is also correlated with image. A house has a symbolic function. Buying a particular house can improve the buyer's image in society and boost his self-image (Sirgy et al., 2005). Females are often more emotionally driven than males (Manzo, 2003). In boom periods, buyer's decision making is also determined by feelings of fear and desperation. Potential buyers offer significant more money for a particular house than sellers initially are asking. A rational valuation is not useful for them because they will otherwise not get a chance to buy this particular house (Shiller, 2016).

2.5.5 Liquidity

It is generally known that housing markets are illiquid, compared to other traditional financial assets (Morawski, 2008). When observing liquidity, one can distinguish two types of liquidity: *trading liquidity* or *asset liquidity* and *funding liquidity* or *market liquidity* (Hull, 2018; Morawski, 2008; van Dijk & Franke, 2018). The *trading perspective* refers to the ease an asset can be converted into cash (Morawski, 2008). Shares on an exchange index, for example, can be bought and sold every second during opening hours of the market. Bonds are not publicly traded. Bond transactions take place over the counter, meaning that buyer and seller privately negotiate on a particular deal. Trading costs are higher, and it takes longer before a transaction takes place. Bonds are therefore less liquid than shares (Hull, 2018). If someone wants to buy or sell a house, a whole process needs to be initiated. It can take weeks or months before a transaction in the housing market finally succeeds. Besides that, there are transaction costs involved, which are way higher than transaction costs in other financial asset markets. The real estate broker gets a commission and the notary sends a bill. Another important costs that need to be taken into account, are opportunity costs of not investing in an alternative asset. Opportunity costs increase with the instability of prices. Also, the fact that the housing market is a heterogenic market, makes information costs to increase. Houses are never identical so for every transaction process new information needs to be gathered. All these costs together makes the housing market, from a trading perspective, an illiquid market. It is hard to liquidate a particular asset (house) in a short period of time with low transaction costs. Trading liquidity decreases when the order size increases, when transaction costs increase and when the time to execute increases (Morawski, 2008). The *funding perspective* relates to the financeability in the market. Is there enough money to fund certain transactions? How is the income position of potential buyers? How eager are banks to provide a mortgage?

Uncertainty about the listing period and the marketing period cause illiquidity in the housing market. Economic downturns can lead to unemployment and deteriorated income situations. These macroeconomic imbalances have an impact on the liquidity in the housing market. People do not have the money to buy a house and it will be harder to obtain mortgages (Taltavull de La Paz & White, 2016). On the other hand, economic boom periods have a positive impact on macroeconomic fundamentals (e.g. higher incomes). This will also lead to macroeconomic imbalances and eventually to a new equilibrium. Liquidity is bad in periods of economic downturn while liquidity improves in economic boom periods.

(Data on) traditional liquidity measures (e.g. applicable in the stock market) are relatively easy to find. Think for example about the bid-ask spread (Morawski, 2008). Since the stock market is a public market and transparent about its transactions, it is easy to measure liquidity in this market. In the housing market measuring liquidity is a bit more complex. In a market where market makers or dealers are absent, there is no bid-ask spread. However, although there are no market makers in the housing market, one can still detect a particular spread: the spread between the list price and the transaction price. Jud et al. (1995) find that the difference between the listing price and the selling price is an appropriate indicator of market liquidity.

Zhu et al. (2019) find that *days on the market* is an important liquidity measure. Days on the market refers to the number of days a particular asset is officially for sale (i.e. the listing period). For sellers this is an important number, because this number implies how fast they can liquidate their house and therefore how fast they can receive money. For buyers it is also an important number. It reflects the popularity of a house. If, on average, the number of days on the market decreases, the number of potential buyers is probably higher.

Another method to measure liquidity in the housing market is the downward adjustment of the list price. If the housing market becomes less liquid, potential buyers have less money to spend. Overall, sellers need to lower the list price in order to be able to sell their house.

Is liquidity positively or negatively correlated with price efficiency? A lot of theories state that liquidity is positively correlated with price efficiency. Illiquidity implies the existence of more transaction costs. If traders do not actively search information in illiquid markets, this may imply that prices deviate by large amounts from their intrinsic values (Tetlock et al., 2007). Alternatively, one could argue that liquidity is a proxy for noise trading. This will have a negative impact on price efficiency. The alternative view may be applicable in the housing market. As discussed earlier, in economic boom periods, it is easier to get a mortgage. Besides that, we can also detect psychological biases and emotions in economic boom periods. I will test this with the following hypotheses.

Hypothesis 5:

H_0 *Liquidity is not correlated with the monthly average transaction price*

H_α *Liquidity is positively correlated with the monthly average transaction price*

Hypothesis 6:

H_0 *Liquidity is not correlated with the monthly transaction volume*

H_α *Liquidity is positively correlated with the monthly transaction volume*

Hypothesis 7:

H_0 *Liquidity is not correlated with the monthly excess return*

H_α *Liquidity is positively correlated with the monthly excess return*

Lastly, I test whether the level of liquidity can be predicted by investor attention using Google search volume indices. In hypotheses 5, 6 and 7 I test the relation between liquidity and the monthly average transaction price, volume and excess return. If Google search queries have predictive power in which direction (il)liquidity goes, one also has an indication in which way the housing market will move. I test this with the following hypothesis:

Hypothesis 8:
H_0 *Liquidity is not correlated with housing related Google search volume indices changes*
H_α *Liquidity is positively correlated with housing related Google search volume indices changes*

2.5.6 Conclusive remarks

Explanations of lacking efficiency are inexhaustible. Other explanations can be offered too. For example, institutional changes, such as a favourable mortgage regulation, which certainly is applicable on the Dutch housing market, can explain why people buy more houses in a certain period (DeBondt, 1995; Shiller, 2007). Also, the lack of new additions to stock can cause bubbles in the housing market. Countries with less land available for development (like the Netherlands) are less able to absorb demand shocks with new construction (Glaeser et al., 2008; Nijskens & Lohuis, 2019).

In this section I discussed the characteristics of the housing market. In general, the housing market cannot be seen as an efficient market. Only a small part of the variation in price changes of housing prices can be explained by fundamentals. Information asymmetry and irrational decision making are the main causes of prices deviating from their fundamental value. Sellers and real estate agents have more information than buyers. Buying a house is both an investment as a consumption function. Both functions deliver explanations of irrational decision making by homebuyers. Psychological biases as overconfidence and herd behaviour explain why buyers believe that the price of their house is going to increase. Buying a house is a ‘one-way bet’. The buyer believes that owning a house provides him a profit in the future while he will not bear any risk (Ferguson, 2008). Emotions and feelings play a big role in homebuyers’ decision making, which makes investment decisions irrational and unexplainable.

CHAPTER 3 Data

3.1 Introduction

In this section I explain the data collection process. Besides that, I describe the collected data. I consult two main data sources: Google search data and data on transactions in the housing market of Amsterdam and Utrecht. The Google Search data is obtained with the use of Google Trends. The data from the Dutch realtor association is requested via a contact at the Erasmus University Rotterdam.

3.2 Google search data

3.2.1 Google search data

Why Google search data as a proxy for investor attention? Internet users use a search engine to find information on the internet and Google is the favourite search engine. Additionally, Google search data is an acknowledged attention measure: when a person searches for a house on the internet, he is unquestionably paying attention to this house. Therefore, aggregate information of search queries is a good market-wide measure of attention (Da et al., 2011). Furthermore, Google search data is an excellent measure to quantify attention compared to alternative news sources such as newspapers. It is hard to measure investor attention by analyzing how people read their newspaper. Where are these people located? Where exactly are they interested in? Besides that, when analyzing Google search behaviour, one directly knows that people are interested in a particular topic (Wu & Deng, 2015). Additionally, information about fundamentals provided by governments and banks is provided with a lag. People's reaction on market events is, however, directly visible on the internet.

3.2.2 Google trends

What exactly is Google Trends? Google, which is the market leading search engine, collects and saves the search behaviour of its users. Google Trends is a tool by which data of search queries and their search volume can be obtained in the form of a Search Volume Index (SVI) (Braun, 2016). Google Trends analyses have various advantages. At first, its comprehensiveness. The Google Trends dataset covers information about almost every country where internet is available. Secondly, the data are uploaded continuously in real time. Furthermore, its customization is a useful feature. Google Trends allows request in search inquiries for multiple search terms. Additionally, the data can be delivered in each desired frequency (e.g. monthly, weekly). Lastly, the information provided by Google Trends is freely accessible for everyone (Meshcheryakov, 2018).

In other research, Google Trends already proved to be a valuable instrument. Hu et al. (2020), for example, found by analyzing Google Search data that governments should strengthen the public awareness of COVID-19 nationally to reinforce the public vigilance and sensitivity to the coronavirus (Hu et al., 2020). Another example is the research by Moat et al. (2013). Moat et al. find that in Google search data patterns can be found that may be interpreted as “early warning signs” of stock market moves. Da, Engelberg and Gao (2011) find that Google search data is positively correlated with other proxies for investor attention and that they possess additional power when they are used for predicting future stock prices.

As shown, Google Trends analyses can inexhaustibly be used in practically every kind of industry. Research about the relation between search behaviour on Google (as a proxy for buyer/investor sentiment) and housing prices is, however, limited (Meshcheryakov, 2018). Wu and Brynjolfsson (2015) find that adding search queries data to their baseline model leads to an improvement of the explanatory power of the model. It outperforms the model predictions of the National Association of Realtors. Van Veldhuizen et al. (2016) investigate the relation between lagged search queries of the word ‘hypotheek’ (mortgage) and the current transaction volume in the Dutch housing market. They find a significant relationship between the current and lagged search query ‘hypotheek’ and the transaction volume on the Dutch housing market. They also find that searches of six and nine months ago are positively associated with the current transaction volume.

3.2.3 Data collection

Google is providing search information from 2004 onwards. However, the information of the first couple of years is not accurate. The transaction data provided by the NVM are measured until June 2019. I therefore use a sample period of January 2007 – June 2019. Google search data is accessible on a weekly and monthly basis. However, weekly data is only available for the last five years. Since the observation period is January 2007 – June 2019, I use monthly data. Selecting the search queries, I found inspiration in the paper written by Wu and Deng (2015), who use the query ‘house for sale’. Van Veldhuizen et al. (2016) regress the lagged search queries of the word ‘hypotheek’ (mortgage) on the transactions value in the Dutch housing market. Wu and Brynjolfsson (2015) use the words ‘real estate,’ ‘housing sales,’ ‘home staging,’ and ‘home inspection’. Beracha and Wintoki (2013) use the words ‘real estate i’ and ‘rent i’ to detect attention (where i is the name of the city).

The largest website that shows houses for sale in the Netherlands, is www.funda.nl. The owner of Funda is the NVM, the organisation that provides the data on housing transactions (van Dijk & Francke, 2018). In a survey by Conclusr 93% of the participants responded ‘funda’ on the question what the first housing website is that they were thinking about (Conclusr, 2014). Since people are consulting this website often, I use ‘funda’ as search queries. Further I look at ‘huis te koop’ (house for sale), ‘huis verkopen (to sell a house), ‘makelaar’ (real estate agent), ‘hypotheek’ (mortgage) and ‘hypotheek afsluiten’ (taking mortgage).

3.2.4 Descriptive statistics

Figures 2 and 3 show the trend line for Google searches 'funda' and 'huis te koop'. Both variables are standardized values, meaning that the values are rescaled such that the mean is zero and the standard deviation is one. Now it is easier to compare values with each other (Van Veldhuizen et al., 2016). I also correct for seasonal effects since the number of housing sales depend on the time of the year. I use the moving average of six months before, the current month and the six months after. The formula for standardizing data is as follows:

$$z_i = \frac{X_i - \mu_i}{\sigma_i} \tag{1}$$

z_i reflects the standardized value of variable *i*. X_i reflects the number that needs to be transformed. μ_i displays the mean of variable i and σ_i reflects the standard deviation of variable *i*. The trend line, smoothed out for seasonality, is black (bold) while the unseasoned line is light grey, reflected on the background. The GSVI trend lines of 'funda' and 'huis te koop' look as follows:

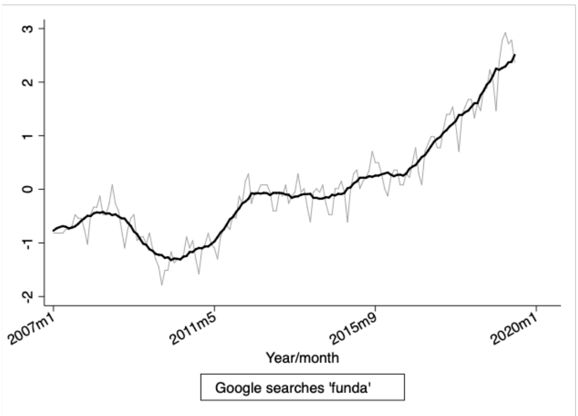

Figure 2. Google searches 'funda'.

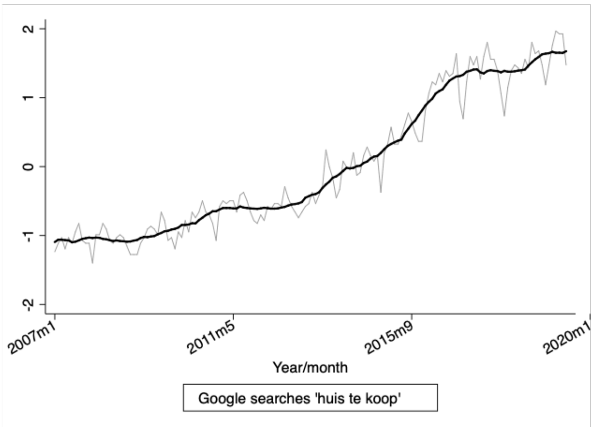

Figure 3. Google searches 'huis te koop'.

Both the lines of 'funda' and 'huis te koop' show an increasing pattern. The lowest search value for 'funda' is 32 and is reached in December 2009. The highest value is 100 and is reached at March 2019. 'The lowest value for 'huis te koop' is 18 and is reached in December 2007. The highest value is 100 and is also reached at March 2019. This implies that GSVI does not change while the house price probably changes. Table 1 also displays the descriptive statistics of the other search queries.

Table 1. Descriptive statistics search queries.

	N	Mean	Median	St.Dev.	Min	Max	Skewness	Kurtosis
'funda'	150	57.8	57	14.44	32	100	.90	3.50
'huis te koop'	150	52.08	42	24.36	18	100	.52	1.81
'huis verkopen'	150	71.04	71	10.67	41	100	-.012	3.09
'makelaar'	150	43.26	32	20.98	21	100	1.29	3.45
'hypotheek'	150	71.17	73	12.34	47	100	0.01	2.07

The observation period is 12.5 years, from January 2007 until June 2019. 'N' is the amount of monthly observations, where one observation is the monthly average search intensity. Mean and St.Dev. are the means and standard deviations over the whole observation period. Min and max are the minimum and maximum monthly average searches intensities. The Google search volume index (GSVI) values contain a value between 0 and 100, where 100 is the highest value in SVI. The values are relative values.

3.2.5 Normality test

Skewness reflects the extent to which the distribution function of a particular variable is symmetrical. If the skewness value of a particular variable is lower than -1 or higher than +1, this is an indication that the distribution of the variable is skewed. Kurtosis measures the extent in which the distribution is peaked. A kurtosis value higher than 1 indicates that the distribution is too peaked. If a distribution is too peaked and/or too skewed, this is an indication that the variable is not normally distributed (Hair et al., 2017). Table 1 shows the Skewness and Kurtosis values of the search queries. The search term 'makelaar' does have a value higher than 1 (1.29), meaning that the distribution is not symmetrical. Also 'funda' has a high value (0.90). Every variable in the sample has a Kurtosis value higher than 1. Since the skewness and Kurtosis tests indicate that the variables are not normally distributed, I cannot regress the GSVI variables on the dependent variables. I therefore transform the values to ln-values, which I discuss later. I also use a Shapiro-Wilk test of normality to test whether the variables are normally distributed. When performing the Shapiro-Wilk test of normality, I find that none of the variables are normally distributed.

3.2.6 Augmented Dickey-Fuller test

The Google Trends data is provided as a time series with a certain frequency. Before I start processing the data, I apply an Augmented Dickey-Fuller test for non-stationarity (Dickey & Fuller, 1979). If the time series is non-stationary, or at least weak stationary, the time series is typified with a time-invariant mean and variance. A non-stationary time series is a time series with a deterministic or stochastic trend. In testing for non-stationarity, the null hypothesis states that a random walk trend is present while the alternative hypothesis states the opposite. I perform Augmented Dickey Fuller tests for all the variables I use in the regression model. In order to do so, I compute the rate of change (Da et al., 2011; Meshcheryakov, 2018). The rate of change is computed as follows:

$$\Delta GSVI_{j,t} = \frac{GSVI_{j,t} - GSVI_{j,t-1}}{GSVI_{j,t-1}} \quad (2)$$

$$\Delta GSVI_{j,t} = \ln\left(GSVI_{j,t}\right) - \ln(GSVI_{j,t-1}) \quad (3)$$

j represents a keyword in the search query and t represents the time period (month). None of the null hypotheses are rejected, meaning that all the variables are non-stationary. I conclude that a unit root is present in the time series. The lagged value then does not provide information about the rate of change in the future value.

3.3 NVM Data

3.3.1 Data collection

The Dutch realtor association, NVM, did not want to share data about the transactions in the Netherlands as a whole. I therefore requested data about the transactions in the housing market of Amsterdam and Utrecht. The housing markets in Amsterdam and Utrecht are very popular and known for its enormous growth (Savills, 2019). Furthermore, the NVM only provided data up and including June 2019. I therefore cannot investigate the impact of the corona crisis in the current, interesting, period.

I drop observations of transactions of houses which are partly rented, since the discount factor of these transaction might be different. The raw data show some unconventional spots. For example, the average transaction value per month, in Utrecht, in 2007, lays between € 260,000 and € 280,000. However, in April 2007 this value is almost € 1,700,000. Also, some transactions are connected to a 0 or 9999 m^2 dwelling, which probably is a system mistake. I therefore winsorize the data at 1% and 99%. Based on the monthly average transaction prices, I calculate the rate of change in the time series. I do this by using the following formula:

$$\Delta HP_{j,t} = \frac{HP_{j,t} - HP_{j,t-1}}{HP_{j,t-1}} \quad (4)$$

$$\Delta HP_{j,t} = \ln(HP_{j,t}) - \ln(HP_{j,t-1}) \quad (5)$$

3.3.2 Descriptive data

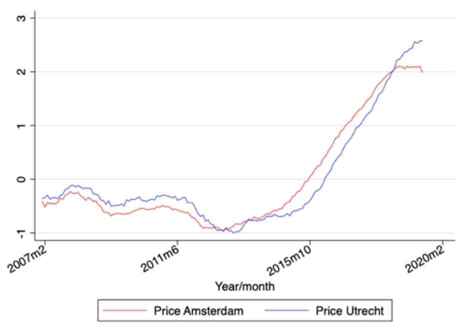

Figure 4. Price Trend House Prices.

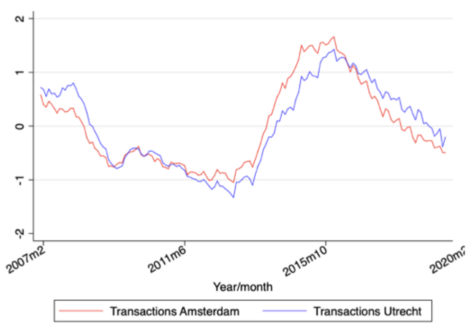

Figure 5. Number of Transactions.

Figure 4 shows the trend lines of the monthly average transaction prices in Utrecht and Amsterdam. Figure 5 shows the trend line of the monthly number of transactions. The lines are corrected for seasonal effects. The mean transaction price in Amsterdam in the period January 2007 – June 2019 is € 333,390.40. The lowest monthly average transaction value is € 246,972.70 and is reached in September 2012. The highest monthly average transaction value is € 509,229.60 and is reached in June 2019. On average, 658 monthly transactions took place in Amsterdam. In January 2013, only 393 transactions took place. The highest amount of transactions, 1139, was reached in June 2015.

The monthly average transaction price in Utrecht, during the period January 2007 – June 2019 is € 275,415.50. The lowest monthly average transaction price in Utrecht is € 203,712.00, reached in January 2013. The highest monthly average transaction price in Utrecht is reached in April 2019 and is € 401,008.50. On average, 297 monthly transactions took place in Utrecht. The lowest amount of monthly transactions, 141, was recorded in March 2013. The highest amount of transactions, 481, was recorded in October 2015.

Table 2. Descriptive statistics transactions.

	N	Mean	Median	St.Dev.	Min	Max	Skewness	Kurtosis
Price A (€)	150	333390	302541	72787.68	246973	509230	1.16	2.99
Price U (€)	150	275415	260953	42185.28	203712	401009	1.35	3.92
% change A price	149	0.47	-	4.67	-10.77	15.8	-	-
% change U price	149	0.25	-	5.01	-16.85	11.67	-	-
Price A per m2	150	3858.57	3514.25	760.42	3060.57	5939.94	1.22	3.10
Price U per m2	150	2755.5	2660.02	350.44	2238.99	3690.90	1.19	3.52
Transactions A	150	658	635	154.56	393	1139	0.50	2.78
Transactions U	150	297	290	72.94	141	491	0.15	2.58

Observation period is 12.5 years, from January 2007 until June 2019. 'N' is the amount of monthly observations. Price A and Price U reflect the monthly average transaction prices is Amsterdam and Utrecht. The transaction data is winsorized at the 1% and 99% percentiles. The transaction prices are measured in euro. Transactions A and U stand for the monthly number of transactions.

3.3.3 Liquidity measures

If houses stay on the market for a longer period, one would expect that the level of liquidity in the market deteriorates. It takes a longer time before sellers receive their money. Therefore, I use the liquidity measure *days on market* (Zhu et al., 2019). Furthermore, if the spread between the list price and the transaction price increases, this means that the level of liquidity also increases (Jud et al., 1995). Buyers are willing to overbid the list price, meaning that there more money is available. The *spread*

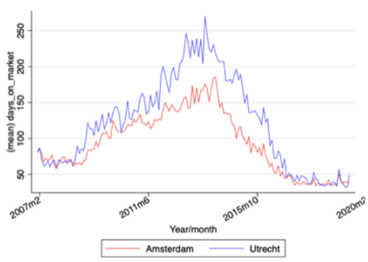

Figure 6. Days on the market.

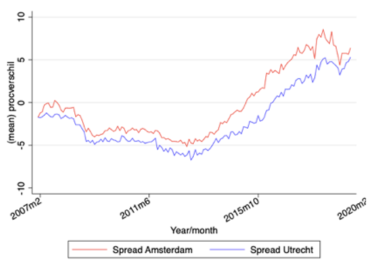

Figure 7. Spread transaction-list price.

Figure 8. Downward adjustment list price.

transaction-list price is therefore an often-used liquidity measure. If sellers need to adjust the list price downwardly, this means that buyers are not willing to pay the list price, implying that less money is available. One would expect that the level of liquidity is relatively low when sellers need to adjust the list price downwardly. I therefore use *downward adjustment list price* as a liquidity measure.

The visual results are reflected in Figures 6, 7 and 8. The descriptive statistics are reflected in Table 3. Over the sample period of January 2007 – June 2019, houses were on average 94 days listed on the market in Amsterdam and 120 days in Utrecht. Figure 6 shows that for both Utrecht and Amsterdam, the listing period started to increase in 2009. The listing period persisted increasing until 2014. After 2014, there is an observable trend downwards. Overall, the listing period is longer in Utrecht than in Amsterdam, suggesting that the housing market in Amsterdam is more liquid that the housing market in Utrecht.

The spread between the transaction price and the list price is, in the whole sample period, negative for both Amsterdam and Utrecht. Figure 7 shows that the spread in Amsterdam is lowest in the period 2012-2013. The lowest recorded monthly average spread in Amsterdam in December 2013. Transaction prices were on average 5.18% lower than the list prices. In May 2018, the highest monthly average spread value is recorded. Transaction prices were 8.57% higher than list prices. Utrecht also shows low spread values in the period 2012-2013. High spread values are recorded in 2018 and 2019. The highest spread value in Utrecht is recorded in June 2019. Monthly average spreads in Amsterdam are always higher than monthly average spreads in Utrecht, which again suggests that the housing market in Amsterdam is more liquid than the housing market in Utrecht.

Figure 8 shows a downward trend in the period 2011-2013. The lowest monthly average value in Amsterdam is -3.49%, meaning that sellers on average adjusted the list price downwardly with 3.49%. This value is observed in March 2013. The lowest value, -6.1%, in Utrecht is measured in January 2013. From 2013 onwards, monthly average adjustments decreased. In February 2018, list prices in Amsterdam were only downwardly adjusted with .17%. In January 2017 list prices in Utrecht were only downwardly adjusted with .18%.

Table 3. Descriptive statistics liquidity measures.

	N	Mean	Median	St.Dev.	Min	Max	Skewness	Kurtosis
Days on market A	150	94	90	42.34	34	185	0.20	1.92
Days on market U	150	120	122	63.81	32	270	0.28	1.93
Transaction-list A	150	-.22	-1.42	4.05	-5.17	8.57	0.72	2.10
Transaction-list U	150	-2.10	-3.42	3.43	-6.74	5.60	0.80	2.38
Price adjustment A	150	-1.34	-1.03	.88	-3.49	-.17	-0.56	2.07
Price adjustment U	150	-1.67	-1.42	1.33	-6.1	-.19	-1.13	3.68

Observation period is 12.5 years, from January 2007 until June 2019. 'N' is the amount of monthly observations. The observations are monthly averages. A stands for Amsterdam and U stands for Utrecht. Days on market are measured by day. 'Transaction-list' is the relative value of the difference between transaction price and list price, divided by the list price. The number is a percentage. Price adjustment values are percentages, calculated by dividing the difference between the new and the old list price with the old list price. The values are relative values.

3.3.4 Normality test

As mentioned earlier, if the skewness value of a particular variable is lower than -1 or higher than +1, this is an indication that the distribution of the variable is skewed. A Kurtosis value higher than 1 indicates that the distribution is too peaked (Hair et al., 2017). The skewness and Kurtosis values of the transaction data (liquidity measures included) are lined up in the last two columns of tables 2 and 3. The transaction prices in both Amsterdam as Utrecht are significantly skewed, and so are the transaction prices per m^2. Also, the downward price adjustment in Utrecht is significantly skewed. Every variable in the sample has a high Kurtosis level. I therefore transform the transaction data to ln-values, similar to the GSVI data. I also use a Shapiro-Wilk test of normality to test whether the variables are normally distributed. When performing the Shapiro-Wilk test of normality, I find that none of the variables is normally distributed.

3.3.5 Augmented Dickey-Fuller test

I apply the same augmented Dickey-Fuller test to the transaction data as I apply to the GSVI data (see section 3.2.6). I conclude that a unit root is present in the time series. Only the variable spread transaction-list price is stationary. Most of the lagged values therefore do not provide information about the rate of change in the future value.

3.4 Data control variables

3.4.1 Introduction

There are factors that impact the composition of the housing market. These factors are for example the population growth and the average rent. A higher population growth leads to more demand in the housing market. An increasing average rent leads to higher house prices, since the cashflow of a particular house increases. In my thesis, I use seven control variables. These control variables are the population growth, the GDP growth, the number of households, the mortgage interest rate, the average rent, the property tax and the consumer price index. An important variable that I could not include, is the yearly development of new houses or additions to the housing stock (Tu et al., 2016). Data about the yearly development of new houses is available at Statistics Netherlands, but only from 2015 onwards.

3.4.2 Population growth

I collect the population growth from Statistics Netherlands. These statistics reflect the total population, the total monthly growth and the relative monthly growth (%). The trend line of the cumulative population growth in the Netherlands, Amsterdam and Utrecht is displayed in figure 9.

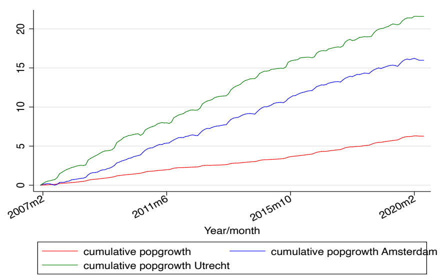

Figure 9. Population Growth.

3.4.3 GDP growth

Data about the GDP (Gross Domestic Product) growth, is gathered from the website of the Organisation for Economic Co-operation and Development (OECD). Data is provided on a quarterly base. The GDP is a real value i.e. it is corrected for price changes and seasonal trends. The provided data reflects the quarterly rate of change (in relation to the previous quarter) in Gross Domestic Product. GDP growth remains, in most of the cases during the observation period, positive. Only in 2009, when the financial crisis had a huge impact on the (world) economy, a negative GDP growth is observed (Economic Co-operation and Development, 2020).

3.4.4 Number of households

Data about the number of households is obtained from the website of Statistics Netherlands. The data is only provided on a yearly base. On average, the number of households in the Netherlands grew with 0.81% per year. In Amsterdam the average growth of the number of households was 1.14% and in Utrecht the average growth of the number of households was 1.44%.

3.4.5 Mortgage interest rate

Data about the mortgage interest rate is obtained from *De Nederlandsche Bank* (DNB), the central bank of the Netherlands. DNB provides monthly mortgage interest rates. The mortgage interest rate shows a downward trend from 2008 onwards. In October 2008, the mortgage interest rate was at its highest level with 4.96%. The interest rate kept, almost linear, decreasing and reached its lowest point in the last month of the observation period: 3.02%. These percentages are a weighted average of all outstanding loans. Thus, they reflect a weighted average of both old and new mortgages. It is also a weighted average of different durations. Some mortgages have a duration between one and five years while others have a duration longer than ten years (De Nederlandsche Bank, 2020).

3.4.6 Average Rent

The average rent level per m^2 is obtained from the independent housing platform 'Pararius'. Pararius provides quarterly data about the average rent level per city per m^2. Quarterly data is only provided from 2010 onwards. Pararius provides data on the average rent in the period 2007-2009, but only on a yearly frequency. I therefore use a yearly average rent level per m^2 in the period 2007-2009.

3.4.7 Consumer Price Index

Data about the consumer price index (CPI) is collected from the website of Statistics Netherlands. The data is observed on a monthly frequency, with 2015 as basis year. The value of a particular observation reflects the increase in CPI in a particular month compared to the same month, one year earlier.

3.5 Abnormal returns

If prices deviate from their fundamental value, abnormal returns can be obtained. In order to calculate the abnormal returns, one needs to subtract the fundamental value from the realized value. A possible method to calculate the fundamental value, is by discounting the imputed rent levels. The fundamental value – what is in the name – can be obtained with the use of fundamentals. Verbruggen et al. (2005) wrote a paper in which they explain the important factors that influence the fundamental value of a house in the Netherlands. One of the main factors that have an impact on the fundamental value is the real income growth. Furthermore, real household capital, tax policies, the real rent level, consumer price index and the stock on the housing market are also important fundamentals. Case and Shiller (1990) calculate quarterly excess return, based on several fundamentals. The model has two main components as a starting point: the annual imputed rent and an estimation of the capital gains earned in the four quarters prior to the observation. I apply a model that relates to the model of Case and Shiller and that includes the fundamental factors as explained by Verbruggen et al. (2005). The model is formulated as follows (Poterba 1992; Levin et al., 2010; Tu et al., 2016):

$$P = \frac{R*(1-r)}{r} \tag{6}$$

$$r = i + \tau + \delta - \pi \tag{7}$$

P is the dependent variable, the house price. R is the cost of renting. I use the yearly average rent per square meter. r is the constant discount factor (Andrle & Plasil, 2019). i is the interest rate that the buyer could have earned in the money market (the opportunity cost of capital). I use the average mortgage rate in the observation period for i, which is 4.3%. τ is the property tax rate. Homeowners pay different taxes which are related to their house. They pay for example a yearly WOZ tax to the local government. WOZ, which means *Waarde Onroerende Zaken* (property valuation), is a taxation of the house by the local government. Besides that, a part of the value of the house is added to the taxable income of the homeowner, meaning that a homeowner needs to pay additional income tax. Furthermore, homeowners pay tax when their house is transferred. An exact yearly tax rate is hard to define based on the different taxes a Dutch homeowner needs to pay. I therefore use the same constant tax rate τ Levin et al. (2010) use in their model: 1.0% per year. δ is the rate of deprecation. The depreciation rate differs among houses, since particular houses lack maintenance while other houses are well maintained (Wilhelmsson, 2008). Harding et al. (2006) use a repeated sales model framework and estimate an average depreciation rate of 2.5% per year. Levin et al. (2010) use a depreciation rate of 2%. I use a depreciation rate of 2.25%. π is the yearly rate of capital gains. I use the average consumer price index in the observation period January 2007 – June 2019 as π, which is around 1.61%. The trend line of the fundamental price and the market price in Amsterdam (both per m^2) looks as follows:

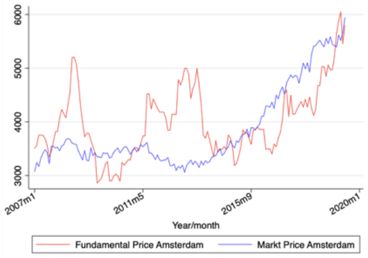

Figure 10. Fundamental and Market Price Amsterdam.

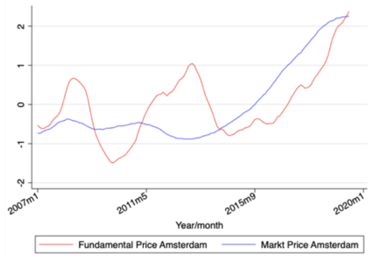

Figure 11. Standardized values Fundamental and Market Prices Amsterdam.

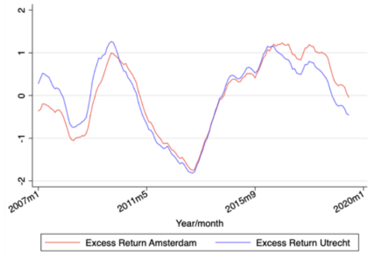

Figure 12. Excess return Amsterdam and Utrecht.

Remarkable is the trend line of the fundamental value, which intuitively would be a smoother line (figure 10). Explanation can be that the fundamental value depends on the rent level, which not directly has to be a fundamental value (Hott & Monnin, 2006). It is measured with noise (Ayuso & Restoy, 2006). Another explanation can be the housing shortage in the Netherlands, which cause an upward trend for rent levels and house prices. Increasing prices therefore do not directly mean that prices are overvalued. Furthermore, Tu, de Haan and Boelhouwer (2016) find that there is a mismatch between house price models and the regulatory environment in the Netherlands. Based on figure 12 one could argue that excess returns are getting closer to zero because of an increase in attention (on Google), which would substantiate that attention leads to an increase in price efficiency. When comparing the standardized values of the fundamental price in Amsterdam with the market price (figure 11), one can observe that the market price exceeded the fundamental price the last six years, while the reverse holds for the four years prior to this surplus.

CHAPTER 4 Methodology

4.1 Introduction

In this chapter I discuss the methodology I use in this research. First, I discuss the Granger causality test. Next I discuss the efficiency model. Thereafter I discuss the correlation models and I conclude with the regression model.

4.2 Granger causality test

In some occasions it is difficult to decide in which direction causality goes. The causality problem often arises when information is slowly incorporated or when possible causal variables are omitted (Granger, 1969). I need to ascertain whether Google search behavior affects the movements in the housing market or whether the movements in the housing market affect Google search behavior. I therefore firstly perform a Granger causality test to find out which time series is useful in predicting another. I test for different twelve lag periods. I apply these tests on the seasonal corrected trend data.

With the Granger causality test, the monthly change in the dependent variable (e.g. house price or liquidity measure) is being regressed on its own lagged value and on the lagged value of the monthly change in an independent variable (e.g. search query). The other side of the test is whether the lagged value of the dependent variable is increasing explanatory power on the current value of the independent variable. The following regression is therefore being estimated (Wiedermann & von Eye, 2016):

$$\Delta y_t = \sum_{i=1}^{m} \alpha_t \Delta GSVI_{j,t-i} + \sum_{i=1}^{n} \beta_t \Delta y_{j,t-i} + u_t \quad (8)$$

$$\Delta GSVI_{j,t} = \sum_{i=1}^{m} \gamma_t \Delta GSVI_{j,t-i} + \sum_{i=1}^{n} \delta_t \Delta y_{j,t-i} + u_t \quad (9)$$

$$\Delta GSVI_{j,t} = \ln(GSVI_{j,t}) - \ln(GSVI_{j,t-1}) \quad (10)$$

$$\Delta y_t = \ln(y_{,t}) - \ln(y_{t-1}) \quad (11)$$

$$\Delta y_t = \frac{y_t - y_{t-1}}{y_{t-1}} \quad (12)$$

Δy_t is the monthly change in the dependent variable. The dependent variables I test are the monthly average house price, volume and excess return and the liquidity measures. The liquidity measures are the downward price adjustments, the listing period and the spread between the transaction price and the list price. $\Delta GSVI_{j,t}$ is the independent variable and reflects the monthly change in the different search queries. u_t is the disturbance term. I use formula 11 for the monthly average transaction prices, transaction volume and the Google search queries. For the liquidity measures, which are percentual values, I use formula 12, since those percentual values can be negative, which cannot be transformed to ln-values. The null hypothesis is that the independent variable $\Delta GSVI_{j,t}$ does not Granger cause the dependent variable Δy_t. Four different outcomes can be distinguished. Firstly, there can be unidirectional causality from $\Delta GSVI_{j,t}$ to Δy_t ($\sum \alpha_i \neq 0$), ($\sum \delta_i = 0$). Secondly, there can be unidirectional causality from Δy_t to $\Delta GSVI_{j,t}$ ($\sum \alpha_i = 0$), ($\sum \delta_i \neq 0$). Thirdly, there can be bilateral causality ($\sum \alpha_i \neq 0$), ($\sum \delta_i \neq 0$). Lastly, there is no Granger causality in both directions ($\sum \alpha_i = 0$), ($\sum \delta_i = 0$).

4.3 Efficiency model

Before I test my regression model, I apply an efficiency test on the data. As already mentioned in the literature review, a price that is efficient, fully reflects all information available. According to Fama, efficiency can empirical be approached in three different ways. Firstly, *strong-form* tests, which test whether investors have monopolistic access to information which is relevant for price formation. Secondly, *semi-strong-form* tests where information includes all obviously publicly available information. Lastly, *weak form* tests, which test whether historical values of prices or returns are correlated with actual prices or returns (Fama, 1970).

In this master thesis I regress the lagged values of the independent variables on the dependent variable to analyze to which extent the data is efficient. Hence, I perform a weak form test. The analysis is time-dependent, meaning that I also test whether efficiency is increasing or decreasing over time. The formula used to test for efficiency:

$$\Delta y_{j,t} = \alpha + \gamma T + \beta \Delta y_{j,t-1} + \epsilon_{j,t} \quad (13)$$

$$\Delta y_{j,t} = \alpha + \gamma T + \beta \Delta GSVI_{j,t} + \epsilon_{j,t} \quad (14)$$

$$\Delta y_{j,t} = \ln(y_{j,t}) - \ln(y_{j,t-1}) \quad (15)$$

$$\Delta GSVI_{j,t} = \ln(GSVI_{j,t}) - \ln(GSVI_{j,t-1}) \quad (16)$$

$$\epsilon_{j,t} = v_{j,t} - u_{j,t} \quad (17)$$

$\Delta y_{j,t} =$ is the change in the dependent variable compared to a month earlier. The independent variables that I test are the change in the monthly average transaction price, the change in monthly average excess return, the change in trading volume and the change in the liquidity measures. j represents a keyword in the search query and t represents the time period (month). So j and t are the cross-section and timeseries combined. γT reflects the seasonal correction. Variables $\Delta GSVI_{j,t}$ represent the monthly change in GSVI. ϵ_t is the error term which is the difference between a normally distributed disturbance $v_{j,t}$, which is an estimation and measurement error, and a one-sided disturbance u_t, which is the technical inefficiency. Technical inefficiency is the difference between the actual output and the potential output (Belotti et al., 2013). The null hypothesis is that $\beta = 0$ and that $\Delta GSVI_{j,t}$ has no predictive power and that the changes in the dependent variables therefore are not significantly correlated with the change in GSVI. The alternative hypothesis is that $\beta \neq 0$ and that $\Delta GSVI_{j,t}$ therefore, has predictive power in forecasting the dependent variable.

4.4 Correlation

With the fifth, sixth and seventh hypothesis I test the correlation between the monthly average transaction price, excess return and trading volume on the one hand and the liquidity measures on the other hand. One would expect that the level of liquidity and the transaction price, excess return and trading volume are positively correlated. The most preferable test to investigate this, is the Pearson's correlation test. In order to correctly apply the Pearson's correlation test, four assumptions must be met (Pearson, 1896). First, the variables that are being tested must be continuous variables, i.e. they must be measurable on a certain scale. This assumption applies on the tested data, since the data in measured in euros or in days. The second assumption is that their exists a linear relationship between the two tested variables. In order to test whether this linear relationship exists, I create scatterplots using Stata and check visually whether a linear relationship exists. The scatterplots show that not every relation reflects linearity. The third assumption is that there should be no significant outliers. Since the data is already winsorized at a 1% and 99% level, outliers are already taken care off. The fourth assumption is that the variables should be normally distributed. I use a Shapiro-Wilk test of normality to test whether the variables are normally distributed. When performing the Shapiro-Wilk test of normality, I find that none of the variables is normally distributed. I therefore cannot use the Pearson's Correlation test.

Alternatively, I use a Spearman test (Spearman, 1904). In contrary to the Pearson Correlation test, variables do not need to be normally distributed (Hauke & Kossowski, 2011). The Spearman rank-order correlation coefficient also measures the strength and the direction between two variables. There are two assumptions that must be met in order to obtain a valid result. Firstly, the variables must be measured on a continuous scale. As already discussed, this is applicable on the to be tested variables. Secondly, the variables must have a monotonic relation, meaning that variables increase together or that one variable is increasing while the other is decreasing. I test this by visually checking the scatterplots of the relations. There is no monotonic relation between the number of deals and the monthly average transaction price (for both cities). For the other scatterplots, there is a monotonic relation.

The Spearman test is necessary to test the fifth, sixth and seventh hypothesis. They are formulated as follows:

<ins>*Hypothesis 5:*</ins>
H_0 *Liquidity is not correlated with the monthly average transaction price*
H_α *Liquidity is positively correlated with the monthly average transaction price*

Hypothesis 6:

H_0 *Liquidity is not correlated with the monthly transaction volume*

H_α *Liquidity is positively correlated with the monthly transaction volume*

Hypothesis 7:

H_0 *Liquidity is not correlated with the monthly excess return*

H_α *Liquidity is positively correlated with the monthly excess return*

With the Spearman test, a Spearman's rho will be obtained. To obtain this rho, the following steps need to be followed. To test the correlation between two variables, there are two series of data needed. Each of the liquidity measures (spread transaction-list price, downward price adjustment and days on market) is alternately being regressed with respectively the monthly average transaction price, transaction volume and excess return. I rank the ln-values, except for the excess returns and the spread between the transaction price and the list price. Next, the pairs of timeseries (the first one is *X* and the second one is *Y*) are ranked from the smallest score to the largest score. The rank list therefore looks as follows $(X_1Y_1)(X_2Y_2)(X_nY_n)$. Next, the difference between the *X* and *Y* rank is calculated, this difference is denoted as d_i. The test statistic looks as follows (Salkind, 2007):

$$r_s = \frac{6\sum d_i^2}{n^3-n} \qquad (18)$$

The null hypothesis is that there is no correlation. Given a 5% significance level, this means that there is no correlation when p > -.05. The alternative hypothesis is that there is a correlation between the two variables. However, the p-value does not directly indicate so much. Spearman's rho reflects the rate of correlation. A value of Spearman's rho can vary between -1 and 1. 1 Means a perfect and positive relationship. -1 means a perfect and negative relationship. Values between 00 – 0.19 supposed to be very weak. Values between .20 and .39 are supposed to be weak. Values between 0.40 and 0.59 are moderate, 0.60 and 0.79 strong. Very strong correlations are values between 0.8 and 1.0.

4.5 Autoregressive model

In order to estimate the relation between the search queries and the housing prices, I use an autoregressive model. I start with a baseline model where I predict the change in the monthly average abnormal return, volume and the monthly average transaction price with the Google search volume indices. Later, I add the control variables. I start testing the first hypothesis:

<u>Hypothesis 1</u>

H_0 *Housing related Google search volume indices changes are not correlated with changes in the Dutch housing transaction prices*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the Dutch housing transaction prices*

For testing this hypothesis, I first analyze the relation between the search queries and the monthly average transaction price. I do this by applying an autoregressive model inspired by Van Veldhuizen et al. (2016). The model looks as follows:

$$\Delta HP_t = \alpha_j + \gamma T + \beta_1 \Delta GSVI_{j,t} + \epsilon_{j,t} \tag{19}$$

The independent variable ΔHP_t represents the monthly average change in house price. α and γT represent the constant term and monthly and yearly fixed effects respectively. j represents a keyword in the search query and t represents the time period (month). ϵ is the error term. Variable $\Delta GSVI_{j,t}$represents the lagged monthly change in GSVI. Every $GSVI_{j,t}$ variable represents a particular search term of the current month. In this model I test whether GSVI has statistical power in forecasting the monthly average transaction price. If the null hypothesis applies, β_1 = 0. The alternative hypothesis states that GSVI has statistical predictive power.

After testing the first model with the search queries, I add the control variables which should have a significant relation with the monthly average transaction price. Case and Shiller (1990) find that an increase in population and income growth explain the growth in house prices. I also add the variables mortgage interest rate and the average rent levels. If the mortgage interest rate decreases, the costs of owning a house decrease. One would expect that the price of a house therefore increases. I also add the average rent level as a control variable. From an investor's perspective, the value of a house is the accumulation of future cash flows. When the rent rises, the value of a property also rises. I put these control variables in vector X. The vector contains the control variables CPI, population growth, number of households, mortgage interest rate and the average rent level. The second regression model with control variables looks as follows:

$$\Delta HP_t = \alpha_j + \gamma T + \beta_1 \Delta GSVI_{j,t} + \delta \Delta X + \epsilon_{j.t} \tag{20}$$

The second hypothesis tests whether the change in Google search indices is related to the change in monthly transaction volume. The hypothesis is formulated as follows:

Hypothesis 2

H_0 *Housing related Google search volume indices changes are not correlated with changes in transaction volume on the Dutch housing market*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the transaction volume on the Dutch housing market*

I test this hypothesis with the following regressions. The intuition of these regression formulas is similar to the regression formulas related to the change in transaction prices. ΔVOL_t is the change in monthly trading volume, which reflects the change in the number of transactions in a particular month. I put the control variables in vector X. The vector contains the control variables CPI, population growth, number of households, mortgage interest rate and the average rent level. If the null hypothesis applies β_1 = 0. The alternative hypothesis states that GSVI has statistical predictive power.

$$\Delta VOL_t = \alpha + \gamma T + \beta_1 \Delta GSVI_{j,t} + \epsilon_{j,t} \tag{21}$$

$$\Delta VOL_t = \alpha_j + \gamma T + \beta_1 \Delta GSVI_{j,t} + \delta \Delta X + \epsilon_{j.t} \tag{22}$$

The third hypothesis tests whether the change in Google search indices is related to the change in monthly excess return. The hypothesis is formulated as follows:

Hypothesis 3

H_0 *Housing related Google search volume indices changes are not correlated with changes in excess return on the Dutch housing market*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the excess return on the Dutch housing market*

I test this hypothesis with the following regressions. Again, the intuition of these formulas is similar to the regression formulas related to the change in transaction prices and the regression formulas related to the change in transaction volume. $\Delta EXCESS_t$ reflects the change in excess return in a particular month. I put the control variables in vector X. The vector contains the control variables CPI, population growth, number of households, mortgage interest rate and the average rent level. As discussed in the data section, excess return is the difference between the market value and the fundamental value. If the null hypothesis applies, β_1 = 0. The alternative hypothesis states that GSVI has statistical predictive power.

$$\Delta EXCESS_t = \alpha + \gamma T + \beta_1 \Delta GSVI_{j,t} + \epsilon_t \quad (23)$$

$$\Delta EXCESS_t = \alpha_j + \gamma T + \beta_1 \Delta GSVI_{j,t} + \delta \Delta X + \epsilon_{j.t} \quad (24)$$

Previous three hypotheses each are being tested with the autoregressive model. With the results of these regression models, the fourth hypothesis can be tested:

<u>*Hypothesis 4*</u>

H_0 *Internet search behaviour does not increase price efficiency on the Dutch housing market*

H_α *Internet search behaviour does increase price efficiency on the Dutch housing market*

So far, the autoregressive model section is only applicable on testing the first four hypotheses. These hypotheses are focused on testing whether analyzing search queries leads to an improvement of price efficiency or whether increasing investor attention leads to more price pressure. I also test whether search queries related to liquidity have an impact on liquidity measures, which are *days on the market*, *spread transaction-list price* and *downward adjustment list price*. With testing this relation, I test the eight and last hypothesis:

<u>*Hypothesis 8:*</u>

H_0 *Liquidity is not correlated with housing related Google search volume indices changes*

H_α *Liquidity is positively correlated with housing related Google search volume indices changes*

For each city, I test a model with the google search queries and an extended model where I add the control variables. The models I test, look as follows:

$$\Delta DOM_t = \alpha + \gamma T + \beta_1 \Delta GSVI_{j,t} + \epsilon_t \quad (25)$$

$$\Delta DOM_t = \alpha_j + \gamma T + \beta_1 \Delta GSVI_{j,t} + \delta \Delta X + \epsilon_{.t} \quad (26)$$

$$\Delta SPREAD_t = \alpha + \gamma T + \beta_1 \Delta GSVI_{j,t} + \epsilon_t \quad (27)$$

$$\Delta SPREAD_t = \alpha_j + \gamma T + \beta_1 \Delta GSVI_{j,t} + \delta \Delta X + \epsilon_t \quad (28)$$

$$\Delta DA_t = \alpha + \gamma T + \beta_1 \Delta GSVI_{j,t} + \epsilon_t \quad (29)$$

$$\Delta DA_t = \alpha_j + \gamma T + \beta_1 \Delta GSVI_{j,t} + \delta \Delta X + \epsilon_t \quad (30)$$

ΔDOM_t is the monthly change of the average listing period. $\Delta SPREAD_t$ is the monthly change in the spread between the transaction price and the list price. ΔDA_t is the monthly change in the downward adjustment of the list price. If the null hypothesis applies, $\beta_1 = 0$. The alternative hypothesis states that GSVI has statistical predictive power.

4.6 Endogeneity

In this master thesis I try to find a causal relation between the search behaviour on Google and the movements on the housing market. Yet, the regression might be prone to some endogeneity biases. Endogeneity points out the situation in which the explanatory variable is correlated with the error term. Three main categories of endogeneity biases can be distinguished: (i) omitted variables, (ii) simultaneity and (iii) measurement errors. The omitted variable bias refers to the situation in which the regression does not include variables that actually are important to explain the variation of the dependent variable. When certain variables are omitted, the regression is prone to unobserved heterogeneity. The simultaneity bias refers to the situation in which the variables simultaneously affect each other. A measurement error refers to the difference between the measured quantity of the independent variable and its true value. Consequently, the correlation between the independent variable and the dependent variable is much weaker (Ullah et al., 2018).

My research might be prone to some of these endogeneity biases. I try to address some of the unobserved heterogeneity. Firstly, the research design is such that I apply the regression models separately for each city (i.e. I do not aggregate the transaction data of both cities). Furthermore, I correct for seasonality and year effects. I also add some control variables to the regression. However, I am aware that the regression still can contain some unobserved heterogeneity. For example, I do not control for the type of house (e.g. terraced house or apartment) or the zip code (some zip codes reflect areas where wealthy people live while other zip codes reflect areas where less fortunate people live).

4.7 Multicollinearity

When more than two predicting variables are applied in the regression, multicollinearity can exist. Multicollinearity is the phenomena of a (perfect) relation among the predicting variables. To a certain extent, it is not a problem if there is a collinear relation among predicting variables. A high level of correlation entails multicollinearity, but the reverse does not directly hold. Multicollinearity does not directly imply correlation. Multicollinearity is about an exact linear dependence. One predicting variable can be predicted by another predicting variable (Alin, 2010).

In the regression models I include multiple predicting variables. I therefore need to test whether the variables are independent enough. I do this by applying the *Variance inflation factor* (VIF) test to the model for multicollinearity. In this test, the correlation matrix of the independent variables is being investigated. The formula looks as follows:

$$\Delta y_j = \beta \Delta X_j + \epsilon_j \tag{31}$$

$$VIF_j = \frac{1}{1-R_j^2} \tag{32}$$

R_j^2 is the determination coefficient of the variation in variable j in relation to the variation in the other independent variables in vector *X*. The larger the VIF coefficient, the larger the variation in relation to the other explanatory variables and therefore the larger the multicollinearity. A VIF coefficient larger than 10 is supposed to imply a large extent of multicollinearity (Alin, 2010). I apply the VIF test to the time series that include multiple explanatory variables, these variables reflect ΔX_j. Δy_j reflects the independent variable. ϵ_j is the error term.

CHAPTER 5 Results

5.1 Introduction

In this chapter I discuss the results. I start with the Granger causality tests. Next, I discuss the correlations between the liquidity measures and the dependent variables. Thereafter I continue with the efficiency tests. After the efficiency test, I continue with the main regression models. Then I discuss the Variance Inflation Index. I end with the robustness checks.

5.2 Granger causality test

I apply Granger causality tests to detect the direction of the causality between the search queries and the market data. I test for different lag periods and I use the standardized values. The results of the Granger causality test, are summarized in table 4. Higher order lags loose predictive power (Wu & Brynjolfsson, 2015). So, I focus primary on the first three lagging periods. Most of the search queries, that I test ('funda', 'huis te koop' 'hypotheek' 'makelaar'), for almost every lagging period, are having a Granger cause effect on the dependent variables. The term 'huis verkopen' is Granger caused by the monthly average transaction price, which might suggest the presence of price pressure: there is more investor attention on the market when more people are selling their house. 'hypotheek' has neither a significant effect on the downward price adjustments nor on the spread between the transaction price and the list price.

Table 4. Granger causality tests.

Granger causality tests												
Search query 'funda'												
Lagging period	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	-10	t-11	t-12
House price A	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
House price U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Search query 'huis te koop'												
Lagging period	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12
House price A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
House price U	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	N
Search query 'huis verkopen'												
Lagging period	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12
House price A	N	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y
House price U	Y	N	Y	N	N	N	N	Y	Y	Y	Y	Y
Search query 'makelaar'												
Lagging period	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12
House price A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
House price U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Search query 'hypotheek afsluiten'												
Lagging period	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12
House price A	N	N	N	N	Y	N	N	N	N	N	N	N
House price U	N	N	N	N	N	N	N	N	N	N	N	N
Search query 'hypotheek'												
Lagging period	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12
House price A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
House price U	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Days on market A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Days on market U	Y	Y	Y	N	N	N	N	N	Y	Y	N	N
Spread A	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N
Spread U	Y	N	N	N	N	Y	N	N	N	N	Y	N
Price adjustment A	N	N	N	N	N	N	N	N	N	N	N	N
Price adjustment U	N	N	N	N	N	N	N	N	Y	Y	Y	Y

In this table Granger causality tests between the search queries and the dependent variables are shown. The search queries are displayed on the horizontal axes while the dependent variables are displayed on the vertical axes. The first five search queries are only tested on the house prices. The last search query is also tested on liquidity measures. The value 'Y' reflects yes and 'N' reflects no. The table only reflects whether the search query is Granger causing the dependent variable. The reverse relation (i.e. the dependent variable Granger causing the search query) is not displayed. P-values are measured at a 5% significance level.

5.3 Correlation

This section focuses on the correlation between the monthly average liquidity on the one hand and the monthly average transaction price, volume and excess return on the other hand. Liquidity is measured in three ways: (i) the spread between the transaction price and the list price, (ii) the downward price adjustment and (iii) the number of days a dwelling is averagely listed.

The Spearman test of the correlation between the monthly average transaction price and the spread between transaction price and the list price in Amsterdam results in a rho of 0.87, which implies a very strong correlation. The test of the correlation between the monthly average transaction price and the downward price adjustment displays a strong correlation with a Spearman's rho of -0.76. The minus sign implies that the level of monthly average transaction prices is increasing when the level of downward adjustment is decreasing. This relation confirms the theory, since a low level of downward adjustment indicates a high liquidity level. Transaction prices should be higher when the level of liquidity is higher, which is the case. So, a *negative* relation between the liquidity measure 'downward price adjustment' and the monthly average transaction price implies a *positive* correlation between the level of liquidity and the monthly average transaction price. The relation between the monthly average transaction price and the average listing period, shows a rho of -0.85, which is a very strong relation. The sign of this relation is negative, which also makes sense according to the theory. When the listing period of a dwelling decreases, this means that the level of liquidity increases. Transaction prices should increase, which is the case. In Utrecht these tests deliver values of respectively 0.73 (strong), -0.76 (strong) and -0.83 (very strong). All tests have a p-value lower than 0.05, meaning that the results are statistically significant.

The relation between the monthly transaction volume and the level of liquidity is less strong. The relation between the transaction volume and the spread transaction-list price in Amsterdam is 0.49, which is moderate. The relation between the monthly average transaction volume and the downward price adjustment is also -0.61, which is a strong relation. The relation between the monthly average transaction volume and the average listing period is -0.28, which is weak. In Utrecht these tests deliver values of respectively 0.62 (strong), -0.62 (strong) and -0.43 (moderate). The p-values are all below 0.05, meaning that the null-hypothesis of no correlation can be rejected.

Lastly, the relation between the monthly average excess return and the level of liquidity is calculated. In Amsterdam, Spearman's rho for the correlation between the monthly average excess return and the spread between list and transaction price is 0.61, which is strong. The rho for the correlation between the monthly average excess return and downward price adjustment is -0.60, what also reflects a strong correlation. Spearman's rho for the relation between the monthly average excess return and the number of days a dwelling is listed is -0.52, which is a moderate correlation. In Utrecht Spearman's rhos are respectively 0.48 (moderate), -0.52 (moderate) and -0.39 (weak). The p-values are all below 0.05, meaning that the null-hypothesis can be rejected.

To conclude, the monthly average transaction price is very strongly correlated with liquidity, in both Amsterdam and Utrecht. The monthly transaction volume is strongly correlated with liquidity, but only for the liquidity measures spread list-transaction price and the downward adjustment. Based on the Amsterdam data, excess returns are strongly correlated with the level of liquidity. Excess returns in Utrecht are moderately/weakly correlated with the level of liquidity.

5.4 Efficiency test

In this section I perform the (weak form) efficiency tests. As already mentioned in the theory-section, efficient prices follow a random walk, implying that prices are not predictable (Fama, 1965). The weak form test shows whether prices (or other dependent variables) can be predicted with the lagged values of the search queries.

The results of the efficiency test between the search queries and the monthly average transaction price in Amsterdam are displayed in table 5. Worth noting for the reader, columns (1) to (5) do not represent five different regressions, but all represent *one* regression. For some search terms, significant results can be found in different lag periods. 'huis verkopen' (to sell a house) has a highly significant correlation with the monthly average transaction price in the first lagging period. 'huis verkopen' is negatively correlated with the monthly average transaction price, which is explainable since this relation indicates that supply is increasing while demand probably is not. Price pressure might be present. 'makelaar' is highly significant twelve months before the measure period. This might indicate that searching for a dwelling starts with looking for a real estate agent, a year before the transaction takes place. A process that starts early before a transaction takes place. Furthermore, 'funda' is significant in the first lagging period and 'huis verkopen' is significant in the fifth lagging period (at a 5% significance level).

The monthly transaction volume in Amsterdam (not presented in a table) is not correlated with the search query 'funda', 'huis te koop', 'huis verkopen' and 'makelaar'. The monthly transaction volume is highly positively correlated with the search term 'huis verkopen' in the first lagging period (at a 1% significance level). The monthly excess return in Amsterdam is positively correlated with the search term 'hypotheek' in the first lagging period (at a 5% significance level).

Table 5. Efficiency test search queries on monthly average transaction price in Amsterdam.

Dependent variable 'monthly average transaction price Amsterdam' Search query:					
	'funda' (1)	'huis te koop' (2)	'hypotheek' (3)	'huis verkopen' (4)	'makelaar' (5)
t-1	0.516**	0.0165	-0.111	-0.115***	-0.114
	(2.268)	(0.0680)	(-1.439)	(-3.784)	(-0.497)
t-2	-0.448	0.298	0.105	-0.0494	-0.0585
	(-1.254)	(0.805)	(0.808)	(-1.248)	(-0.167)
t-3	0.0903	0.152	-0.0527	-0.0394	0.254
	(0.252)	(0.427)	(-0.397)	(-0.973)	(0.679)
t-4	-0.0105	0.323	-0.0482	-0.0689	-0.0142
	(-0.0316)	(0.955)	(-0.362)	(-1.610)	(-0.0397)
t-5	0.374	-0.131	-0.0597	-0.0839**	-0.0207
	(1.167)	(-0.407)	(-0.428)	(-2.083)	(-0.0586)
t-6	-0.278	0.250	0.0460	-0.0757*	0.145
	(-0.827)	(0.772)	(0.346)	(-1.922)	(0.366)
t-7	-0.368	0.475	0.0875	0.00667	-0.100
	(-1.050)	(1.354)	(0.701)	(0.164)	(-0.266)
t-8	0.360	0.0644	-0.0145	0.00134	0.0987
	(1.031)	(0.169)	(-0.123)	(0.0334)	(0.286)
t-9	-0.0367	0.150	0.0381	-0.0319	0.132
	(-0.0995)	(0.375)	(0.324)	(-0.795)	(0.374)
t-10	-0.342	0.143	0.0443	-0.0118	0.188
	(-0.928)	(0.351)	(0.381)	(-0.301)	(0.519)
t-11	0.493	-0.0348	0.0512	-0.0327	-0.320
	(1.380)	(-0.0955)	(0.463)	(-0.814)	(-0.851)
t-12	0.0174	-0.181	-0.0528	-0.0140	0.669***
	(0.0828)	(-0.777)	(-0.646)	(-0.386)	(2.726)
Constant	0.172**				
	(2.086)				
Observations	138	138	138	138	138
Adj. R-squared	1.000	1.000	1.000	1.000	1.000
Year dummies	Y	Y	Y	Y	Y
Month dummies	Y	Y	Y	Y	Y

The table shows the results of the weak form efficiency test. The dependent variable is the monthly average transaction price in Amsterdam. The independent variables are the search queries. All the search queries in columns (1) to (5) are part of one regression. Therefore, the constant term and the adj. R-squared are the same. The y-axis reflects the lags. The independent variables are lagged values for 12 months. Values are standardized and transformed to values in ln. Both the dependent and independent variables are transformed to ln. T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

In table 6 the relation between the lagged search queries and the monthly average transaction price in Utrecht is displayed. 'hypotheek' is significant at a 5% significance level in the first lagging period, which confirms the results of van Veldhuizen, Vogt and Voogt (2016). Just as for the Amsterdam data, 'makelaar' is significant in lagging period twelve. Furthermore, there are significant results for the search query 'funda' in the eight and twelfth lagging period and for 'huis te koop' in the first and last lagging period. The significant results in the eight and twelfth lagging period are probably a coincidence, the results in the first lagging period, however, are important.

The monthly transaction volume in Utrecht is not correlated with the lagged values of the search terms 'funda' and 'huis te koop'. It is highly significant for the lagged search query 'hypotheek' in the first lagging period, again a confirmation of the results found by van Veldhuizen et al. (2016). It is not significant for the search term 'huis verkopen'. It is significantly correlated for the search term ‘makelaar’ in the first lagging period at a 5% significance level. The monthly excess return is significantly correlated with the search term ‘funda’ in the second lagging period and in the tenth lagging period. It is also significant for the search term ‘huis te koop’ in the first and tenth lagging period. Furthermore, the monthly excess return is significantly correlated with the search term ‘hypotheek’ in the first and eight lagging period (again, in line with van Veldhuizen et al., 2016).

Table 6. Efficiency test search queries on monthly average transaction price in Utrecht.

Dependent variable ‘monthly average transaction price Utrecht' Search query:					
	‘funda’ (1)	‘huis te koop’ (2)	‘hypotheek’ (3)	‘huis verkopen’ (4)	‘makelaar’ (5)
t-1	0.456	0.763*	-0.302**	-0.0532	-0.203
	(1.263)	(1.978)	(-2.466)	(-1.103)	(-0.558)
t-2	-0.403	-0.115	0.0134	-0.0508	0.126
	(-0.711)	(-0.197)	(0.0653)	(-0.809)	(0.226)
t-3	0.663	-0.556	-0.106	0.00670	0.0277
	(1.165)	(-0.987)	(-0.505)	(0.104)	(0.0466)
t-4	-0.493	0.767	-0.198	-0.00866	0.586
	(-0.934)	(1.430)	(-0.935)	(-0.128)	(1.033)
t-5	0.261	0.370	0.115	-0.101	-0.684
	(0.513)	(0.728)	(0.521)	(-1.585)	(-1.219)
t-6	-0.159	-0.0133	0.0749	-0.0921	0.619
	(-0.298)	(-0.0259)	(0.355)	(-1.474)	(0.984)
t-7	-0.667	0.801	0.147	-0.00180	-0.203
	(-1.200)	(1.439)	(0.743)	(-0.0279)	(-0.340)
t-8	1.178**	-0.235	-0.321*	0.00901	-0.0348
	(2.130)	(-0.388)	(-1.710)	(0.141)	(-0.0637)
t-9	-0.326	0.0530	0.116	-0.0558	0.457
	(-0.556)	(0.0838)	(0.622)	(-0.877)	(0.816)
t-10	-0.496	0.839	0.0411	-0.00703	-0.201
	(-0.849)	(1.294)	(0.223)	(-0.113)	(-0.351)
t-11	0.157	-0.0446	0.0423	-0.0190	0.00350
	(0.277)	(-0.0771)	(0.241)	(-0.299)	(0.00587)
t-12	0.651*	-0.817**	-0.131	-0.0843	0.920**
	(1.958)	(-2.207)	(-1.010)	(-1.469)	(2.363)
Constant	0.172				
	(1.319)				
Observations	138	138	138	138	138
Adj. R-squared	0.999	0.999	0.999	0.999	0.999
Year dummies	Y	Y	Y	Y	Y
Month dummies	Y	Y	Y	Y	Y

The table shows the results of the weak form efficiency test. The dependent variable is the monthly average transaction price in Utrecht. The independent variables are the search queries. All the search queries in columns (1) to (5) are part of one regression. Therefore, the constant term and the adj. R-squared are the same. The y-axis reflects the lags. The independent variables are lagged values for 12 months. Values are standardized and transformed to ln. Both the dependent as independent variables are in ln. T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.5 Regression model

5.5.1 Monthly average transaction price

Table 7 shows the results of the regression model for both Amsterdam in columns (1) and (2) and Utrecht in columns (3) and (4). Columns (1) and (3) show the relation between the search queries and the dependent variable ‘monthly average transaction price’. The control variables are added in columns

(2) and (4). In column (1) I test the relation between the change in search queries and the change in the monthly average transaction value in Amsterdam. Besides search query ‘hypotheek’, every search query has a highly significant impact on the transaction price in Amsterdam. The search terms ‘funda’, ‘huis te koop’ and ‘makelaar’ are positively correlated with the monthly average transaction price. The search term ‘huis verkopen’ (to sell a house) is negatively correlated with the monthly average transaction price. Apparently, the monthly average transaction value is decreasing when more people are trying to sell their house, which intuitively makes sense. Column (3) reflects the results of the regression of the search queries on the monthly average transaction price in Utrecht. For every search query, except for ‘huis verkopen’, the relation is significant at a 1% significance level. Also, in Utrecht, ‘huis verkopen’ is negatively correlated with the monthly average transaction price. The same applies for the search query ‘hypotheek’.

Columns (2) and (4) reflect the results of the regression model when the control variables are included. The explanatory power of Google search volume indices decreases when other explanatory variables are added. However, some of the search queries are still significantly and positively (‘huis verkopen’ excepted) correlated with the monthly average transaction price. The results show that an 1% increase in the search queries ‘funda’ and ‘huis te koop’ in Amsterdam leads to an increase in the future average house price in Amsterdam of 0.156% and 0.625% respectively. Given a monthly average transaction price of € 333,390 in Amsterdam, this means that the future average transaction price will rise with € 520.09 when the search query ‘funda’ rises with 1% and € 2,083.69 after an increase of 1% of the search query ‘huis te koop’. An 1% increase in search queries ‘huis te koop’ and ‘hypotheek’ leads to a change in future average house prices in Utrecht of 0.351% and -0.218% respectively. Given a monthly average transaction price of € 275,415 this means that an increase of 1% by the search query ‘huis te koop’ leads to an increase of € 966.71 of the future average transaction price. An increase of 1% by the search query ‘hypotheek’ leads to a decrease of the future average transaction price of € 600.40.

Table 7. Monthly average transaction value Amsterdam and Utrecht.

Dependent variable: monthly average transaction price	Amsterdam		Utrecht	
	(1)	(2)	(3)	(4)
Search term 'funda'	0.350***	0.156*	0.561***	0.0361
	-6.460	-1.848	-8.753	(0.488)
Search term 'huis te koop'	1.003***	0.625***	1.052***	0.351**
	(11.45)	-2.743	(10.13)	-2.446
Search term 'hypotheek'	-0.0412	0.0588	-0.316***	-0.218***
	(-0.934)	-1.408	(-6.043)	(-6.270)
Search term 'huis verkopen'	-0.143***	0.0589	-0.0962*	0.0397
	(-3.074)	-1.354	(-1.748)	(0.832)
Search term 'makelaar'	0.544***	-0.295	0.784***	0.0867
	-7.232	(-1.600)	-8.796	(0.649)
CPI		-0.778***		-0.811***
		(-3.694)		(-3.811)
Population Amsterdam/Utrecht		0.792		1.332***
		(0.975)		-2.966
Mortgage interest rate		-1.505***		-1.144***
		(-6.845)		(-7.977)
Average rent Amsterdam/Utrecht		-0.0606		0.470***
		(-0.736)		-7.752
Number of households		-1.447***		-1.369***
		(-3.189)		(-3.731)
Constant	0.0266	0.0985	0.0123	0.226*
	(0.172)	(0.816)	(0.0672)	-1.881
Observations	150	150	150	150
Adj. R-squared	0.992	0.996	0.989	0.996
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. The dependent variable monthly average transaction price in Amsterdam are displayed in columns (1) and (2) and for Utrecht in columns (3) and (4). The transaction prices, search queries and control variables are standardized values and transformed to ln. The same applies for the independent variables which are all standardized ln-values. The control variables are the CPI, the mortgage interest rate, the average rent in Amsterdam (column (2)) and Utrecht (column (4)), the number of households and the population in Amsterdam (column (2)) and Utrecht (column (4)). The observation period is from January 2007 to June 2019. T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.5.2 Monthly transaction volume

Table 8 displays the results of the autoregressive model in which the search queries are regressed on the monthly transaction volume. Again, columns (1) and (2) display the results related to the transactions in Amsterdam while columns (3) and (4) display the results related to the transactions in Utrecht.

In column (1) I test the relation between the search queries and the monthly transaction volume in Amsterdam. For every search query, ‘huis verkopen’ excluded, applies that the impact on the monthly transaction volume is significant. In Utrecht (column (3)), every search query is significantly related to the monthly transaction volume. Except for ‘funda’, every search query is positively correlated with the monthly transaction volume.

The control variables are added in columns (2) and (4). All the control variables have explanatory power. The results of the regression related to the housing market in Amsterdam differ from the results related to the housing market in Utrecht. After adding the control variables, only the search queries ‘huis verkopen’ and ‘makelaar’ are significant while in Utrecht ‘funda’ and ‘huis te koop’ also have explanatory power. The search term ‘huis verkopen’ does not have a significant relation with the search queries without controls, while it does have a significant relation (or a more significant relation) when the control variables are added. Furthermore, the relation between the monthly transaction volume and the search query ‘huis verkopen’ is negative in Amsterdam while it is positive in Utrecht. ‘funda’ is not negatively correlated anymore, but positively (in Utrecht). It seems that the control variables do not have a big impact on the significance of the other search queries in Utrecht, which are all positively and significantly correlated with the monthly transaction volume.

Table 8. Monthly transaction volume Amsterdam and Utrecht.

Dependent variable: monthly average transaction volume				
	Amsterdam		**Utrecht**	
	(1)	(2)	(3)	(4)
Search term 'funda'	-0.468***	-0.0919	-0.318***	0.578***
	(-4.527)	(-0.506)	(-3.437)	(5.110)
Search term 'huis te koop'	0.612***	0.559	0.885***	0.818***
	(3.663)	(1.138)	(5.918)	(3.722)
Search term 'hypotheek'	0.271***	0.0795	0.185**	0.0676
	(3.221)	(0.883)	(2.462)	(1.270)
Search term 'huis verkopen'	0.0872	-0.238**	0.155*	0.183**
	(0.985)	(-2.543)	(1.956)	(2.501)
Search term 'makelaar'	0.310**	1.092***	0.687***	1.381***
	(2.158)	(2.746)	(5.350)	(6.754)
CPI		-1.564***		-2.792***
		(-3.445)		(-8.572)
Population Amsterdam/Utrecht		7.111***		7.210***
		(4.060)		(10.49)
Mortgage interest rate		1.424***		1.610***
		(3.004)		(7.329)
Average rent Amsterdam/Utrecht		0.754***		-0.286***
		(4.251)		(-3.076)
Number of households		-4.315***		-2.712***
		(-4.409)		(-4.827)
Constant	0.145	0.369	0.128	0.233
	(0.494)	(1.416)	(0.485)	(1.267)
Observations	150	150	150	150
Adj. R-squared	0.953	0.967	0.963	0.983
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. The dependent variable monthly average transaction volume in Amsterdam is displayed in columns (1) and (2) and for Utrecht it is displayed in columns (3) and (4). The transaction volume, search queries and control variables are standardized values and transformed to ln. The same applies for the independent variables which are all standardized ln-values. The control variables are the CPI, the mortgage interest rate, the average rent in Amsterdam (column (2)) and Utrecht (column (4)), the number of households and the population in Amsterdam (column (2)) and Utrecht (column (4)). The observation period is from January 2007 to June 2019. T-statistics in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

5.5.3 Monthly average excess return

Table 9 shows the results of the regression of the search queries on the monthly change in excess returns in the Amsterdam and the Utrecht housing market. Columns (1) and (2) reflect the results related to the excess return in Amsterdam while columns (3) and (4) display the results related to the excess return in Utrecht.

In column (1) the relation between the search queries and the excess return in Amsterdam are tested. Only the search queries 'funda' and 'huis te koop' have a significant impact (at a 1% and a 10% significance level) on the excess returns. In Utrecht, only the search query 'hypotheek' has a significant impact on the monthly average excess return (column (3)).

When adding the control variables, the search term 'funda' still has significant power in explaining the monthly average excess return in Amsterdam, at a 1% significance level. The relation remains negative. In Utrecht, the search query 'hypotheek' still has explanatory power at a 5% significance level. The results of Amsterdam and Utrecht are not similar, which makes it difficult to generalize the data to a national level.

Table 9. Monthly average excess return Amsterdam and Utrecht.

Dependent variable: monthly average excess return	**Amsterdam**		**Utrecht**	
	(1)	(2)	(3)	(4)
Search term 'funda'	-0.572***	-0.515***	-0.00436	-0.179
	(-3.586)	(-2.812)	(-0.0330)	(-1.390)
Search term 'huis te koop'	0.435*	0.346	0.204	0.218
	(1.815)	-1.325	(1.028)	-1.234
Search term 'hypotheek'	-0.190	-0.206	0.373***	0.317**
	(-1.140)	(-1.157)	(2.712)	-2.501
Search term 'huis verkopen'	0.193	0.100	0.261	0.162
	(0.974)	(0.492)	(1.588)	-1.083
Search term 'makelaar'	-0.00299	0.228	-0.206	-0.0737
	(-0.0121)	(0.819)	(-1.013)	(-0.389)
CPI		10.46*		-0.137
		-1.699		(-1.180)
Population Amsterdam/Utrecht		4.025		-0.0333
		-1.024		(-0.341)
Mortgage interest rate		-17.74***		0.235**
		(-2.710)		-2.517
Average rent Amsterdam/Utrecht		-0.773		0.349***
		(-0.175)		-4.889
Number of Households		49.87*		-0.297**
		-1.713		(-2.352)
Constant	0.0379	-10.40*	-0.122**	-0.366***
	(0.649)	(-1.700)	(-2.518)	(-3.819)
Observations	150	150	150	150
Adj. R-squared	0.616	0.638	0.530	0.632
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. The dependent variable monthly average transaction excess return in Amsterdam in columns (1) and (2) and Utrecht in columns (3) and (4). The variables represent the rate of change. The control variables are the CPI, the mortgage interest rate, the average rent in Amsterdam (column (2)) and Utrecht (column (4)), the number of households and the population in Amsterdam (column (2)) and Utrecht (column (4)). The observation period is from January 2007 to June 2019. T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.5.4 Liquidity measures

Table 10 shows the results of the regression of the change in search queries on the liquidity measure *days on market*. The table contains both the results related to the housing market in Amsterdam and

Utrecht. Columns (1) and (3) show the results of the regression of the search queries on the liquidity measure. In both cities, most of the search queries are significantly correlated with the liquidity measure. In columns (2) and (4), the control variables are added. In Amsterdam, four search queries are highly significant at a 1% significance level. In Utrecht, three search queries have a significant relationship with the monthly average listing period at a 1% significance level. The relation is negative, which is opposite to the relation between the search queries and the monthly average transaction value and the transaction volume. The results make sense, since there is a negative correlation between the listing period and the transaction value and transaction volume (in line with the findings of testing hypotheses 5 and 6). The sign of the relation between the search queries and the liquidity measure should therefore be opposite. When the control variables are added, still four search queries have a significant relation with the liquidity measure in Amsterdam. The same applies to the Utrecht data, where three search queries remain significantly.

Table 10. Liquidity measure 'days on market'.

Independent variable: liquidity measure 'days on market'				
	Amsterdam		Utrecht	
	(1)	(2)	(3)	(4)
Search term 'funda'	-0.496***	-0.188*	-0.536***	-0.0638
	(-6.768)	(-1.878)	(-7.565)	(-0.802)
Search term 'huis te koop'	-0.315***	-0.0874	-0.615***	-0.489***
	(-2.654)	(-0.322)	(-5.367)	(-3.163)
Search term 'hypotheek'	-0.199***	-0.238***	0.0158	-0.101***
	(-3.326)	(-4.788)	(0.274)	(-2.711)
Search term 'huis verkopen'	0.0799	-0.144***	0.0767	-0.0426
	(1.271)	(-2.789)	-1.263	(-0.828)
Search term 'makelaar'	-0.516***	0.938***	-0.860***	0.636***
	(-5.067)	-4.274	(-8.739)	(4.426)
CPI		1.808***		1.536***
		-7.219		(6.708)
Population Amsterdam/Utrecht		0.0909		0.154
		(0.0940)		(0.318)
Mortgage interest rate		1.855***		1.680***
		-7.091		(10.88)
Average Rent Amsterdam/Utrecht		-0.168*		-0.349***
		(-1.712)		(-5.345)
Number of households		0.563		0.866**
		-1.043		(2.191)
Constant	0.103	0.0913	-0.0811	-0.0671
	(0.494)	(0.636)	(-0.401)	(-0.519)
Observations	150	150	149	149
Adj. R-squared	0.986	0.994	0.987	0.995
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. The dependent variable is the change in the liquidity measure 'days on market' in a particular month. This liquidity measure reflects the average days a particular house is on the market e.g. the listing period. The dependent variable is a standardized value and transformed in ln. The same applies for the independent variable which is also standardized ln-values. The table contains both the liquidity measure for Amsterdam in columns (1-2) and Utrecht in Columns (3-4). Control variables are added in columns (2) and (4). T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

In table 11 the relation between the search queries and the liquidity measure *spread transaction-list price* is displayed. The spread is a value which is negative in periods when the level of liquidity is low (since people are not willing to pay the original list price) and positive in periods when the level of liquidity is high. In Amsterdam, only the search query 'huis te koop' is significantly correlated with the liquidity measure. In Utrecht, also the search query 'funda' is significantly correlated with the liquidity measure. After adding the control variables, the coefficients remain significantly.

Table 11. Liquidity measure 'spread transaction price - list price'.

Independent variable: liquidity measure 'spread transaction price - list price'				
	Amsterdam		**Utrecht**	
	(1)	(2)	(3)	(4)
Search term 'funda'	0.00689	-0.00622	0.518***	0.433**
	(0.0745)	(-0.0722)	(3.444)	(2.452)
Search term 'huis te koop'	0.234*	0.263**	-0.492**	-0.547**
	(1.687)	(2.051)	(-2.177)	(-2.289)
Search term 'hypotheek'	0.0184	-0.0155	-0.152	-0.127
	(0.191)	(-0.171)	(-0.967)	(-0.771)
Search term 'huis verkopen'	0.117	0.0464	-0.139	-0.240
	(1.016)	(0.434)	(-0.741)	(-1.204)
Search term 'makelaar'	-0.0799	-0.179	-0.105	0.0101
	(-0.560)	(-1.298)	(-0.451)	(0.0397)
CPI		-0.291		0.304
		(-1.458)		(0.727)
Population Amsterdam/Utrecht		-1.681**		-1.567
		(-2.493)		(-1.350)
Mortgage interest rate		-0.0534		-0.182
		(-0.357)		(-0.853)
Average Rent Amsterdam/Utrecht		-0.0257		0.0476
		(-0.335)		(0.353)
Number of households		1.689***		0.948
		(2.704)		(1.029)
Constant	0.926***	0.255	-0.0308	-0.280
	(27.40)	(1.470)	(-0.560)	(-0.926)
Observations	150	150	149	149
Adj. R-squared	0.895	0.912	0.617	0.612
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. The dependent variable is the change in the liquidity measure 'spread list-transaction price' in a particular month. This liquidity measure reflects the difference between the list and transaction price. The dependent variable is a standardized value and is calculated by subtracting the value from the lagged value and divided by the lagged value. The same applies for the independent variable which is also standardized. The table contains both the liquidity measure for Amsterdam in columns (1-2) and Utrecht in columns (3-4). Control variables are added in columns (2) and (4). T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

The relation between the search queries and the liquidity measure *downward price adjustment* is displayed in table 12. There will always be some downward price adjustments, since there will always be sellers that list their house for a higher price than potential buyers are willing to pay for it. However, the downward price adjustment is higher in periods when the level of liquidity is low. When the level of liquidity is low, more sellers need to adjust the list prices downwardly since there are not enough potential homebuyers. In Amsterdam, three search queries have a significant impact on the liquidity measure. The relation is negative, what confirms the theory: the level of liquidity is high when there is more attention. In Utrecht, four search queries have a significant impact on the liquidity measure. When adding the control variables, even five search terms have a significant impact on the liquidity measure in Utrecht. In Amsterdam, three search queries have a significant impact on the liquidity measure.

Table 12. Liquidity measure 'downward price adjustment'.

Independent variable: liquidity measure 'downward price adjustment'				
	Amsterdam		**Utrecht**	
	(1)	(2)	(3)	(4)
Search term 'funda'	-0.0237	-0.494***	-0.0217	-0.210**
	(-0.300)	(-4.237)	(-0.304)	(-2.197)
Search term 'huis te koop'	-0.740***	-0.197	-0.997***	-1.173***
	(-5.784)	(-0.625)	(-8.614)	(-4.538)
Search term 'hypotheek'	-0.310***	-0.294***	-0.129**	-0.189***
	(-4.816)	(-5.091)	(-2.213)	(-3.999)
Search term 'huis verkopen'	-0.0273	0.0911	-0.112*	-0.119**
	(-0.403)	(1.516)	(-1.820)	(-2.416)
Search term 'makelaar'	-0.648***	-0.517**	-0.846***	0.381*
	(-5.901)	(-2.026)	(-8.505)	(1.822)
CPI		3.218***		2.859***
		(11.05)		(11.97)
Population Amsterdam/Utrecht		-6.423***		-2.849***
		(-5.718)		(-3.092)
Mortgage interest rate		-0.576*		0.977***
		(-1.896)		(3.916)
Average Rent Amsterdam/Utrecht		-0.260**		-0.0818
		(-2.285)		(-0.876)
Number of households		2.808***		2.297***
		(4.475)		(4.464)
Constant	-0.240	-0.290*	-0.294	-0.203
	(-1.065)	(-1.738)	(-1.439)	(-1.484)
Observations	150	150	149	149
Adj. R-squared	0.982	0.991	0.986	0.994
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. The dependent variable is the change in the liquidity measure 'downward price adjustment' in a particular month. This liquidity measure reflects the percentage house prices on average are downwardly adjusted. The dependent variable is a standardized value and transformed to ln. The same applies for the independent variable which is also a standardized ln-value. The table contains both the liquidity measure for Amsterdam in columns (1-2) and Utrecht in columns (3-4). Control variables are added in columns (2) and (4). T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.6 Multicollinearity

A variance inflation factor (VIF) test is used to test for multicollinearity. The VIF test is applied to the time series that include multiple explanatory variables. These are the regressions that include the change in the monthly average transaction price, transaction volume and excess return as dependent variables and where the search queries ('funda', 'huis te koop', 'hypotheek' 'huis verkopen' and 'makelaar') are the independent variables. Table 13 shows the results of the VIF tests. Column (1) shows the correlation among the search queries. This correlation is the same for every regression (change in monthly average transaction price, volume and excess return). There are no extremely high (all values are below 10) multicollinearity values, meaning that multicollinearity is not present among the explanatory variables.

Table 13. Variance inflation factor test search queries.

VIF test	
	(1)
'funda'	1.08
'huis te koop'	1.26
'hypotheek'	1.45
'huis verkopen'	1.75
'makelaar'	1.75
mean VIF	1.46

The table shows the results of the variance inflation factor test among the independent variables. Column (1) shows the correlation among the search queries. The values are measured in ln and are standardized. They are corrected for seasonality.

5.7 Robustness checks

5.7.1 Introduction

In this section I apply two robustness checks on the dataset. Firstly, I test the difference between the periods during and after the financial crisis. I define the crisis period as the period from the moment when prices in the sample started to fall to the moment when the prices started to rise again. Prices started to fall from April 2008 and prices started to rise again from March 2013. I only apply this robustness check on the relation between the monthly average transaction price and the monthly transaction volume and the search queries in Amsterdam.

Another robustness check is the spread between the transaction value and the WOZ value. As explained earlier, the WOZ value is the taxation of the house by the local government. When the transaction value is higher than the WOZ value, this means that the house is actually overpriced. The transaction value is a good proxy for the value of the average mortgage. In periods in which the level of liquidity is high, the spread between the transaction price and the WOZ value should be higher. If the search queries are correlated with the overvaluation, this can imply the presence of price pressure.

5.7.2 In- and outside the financial crisis

Table 14 shows the relation between the search queries and the monthly average transaction price in Amsterdam, during and after the financial crisis. The crisis period is the period between May 2008 and February 2013. Columns (1) and (2) represent the period during the crisis and columns (3) and (4) represent the period after the crisis. Columns (2) and (4) include the control variables. After adding the control variables, four search queries have a significant impact on the monthly average transaction price, during the financial crisis. Three of them are significant at a 1% significance level. In the period after the financial crisis, only one search query ‘hypotheek’ has a significant impact. This result suggests that search queries are better in explaining the monthly average transaction price in volatile periods.

I find similar results when I test the relation between the search queries and the monthly average transaction volume in Amsterdam in the period during and after the financial crisis. Four search queries have a significant impact on the monthly average transaction volume during the financial crisis while there is no significant relation between search queries and the monthly average transaction volume after the financial crisis. This result also suggests that search queries have more explanatory power during a financial crisis compared to less volatile periods. These results confirm the findings by Drake et al. (2012), stating that abnormal search volume is higher for higher idiosyncratic volatility.

Table 14. Results regression monthly average transaction value during and after financial crisis.

Dependent variable: Monthly average transaction price Amsterdam				
	During crisis		**After crisis**	
	(1)	(2)	(3)	(4)
Search term 'funda'	-0.269***	-0.620***	-0.181	-0.128
	(-2.692)	(-4.898)	(-1.453)	(-0.913)
Search term 'huis te koop'	-0.0633	0.189	-0.329*	-0.269
	(-0.368)	(1.084)	(-1.729)	(-1.414)
Search term 'hypotheek'	0.324**	0.213*	0.217*	0.286*
	(2.609)	(1.985)	(1.693)	(1.993)
Search term 'huis verkopen'	-0.169	-0.502***	-0.208	-0.231
	(-1.291)	(-3.800)	(-1.298)	(-1.437)
Search term 'makelaar'	0.373**	0.428***	0.288	0.238
	(2.259)	(2.702)	(1.596)	(1.188)
CPI		0.371***		-0.167
		(2.988)		(-1.315)
Population Amsterdam		-0.147		-0.0386
		(-1.235)		(-0.365)
Mortgage interest rate		0.557***		-0.147**
		(4.045)		(-2.076)
Average Rent Amsterdam		-0.0571		-0.108
		(-0.795)		(-1.076)
Number of Households		0.282*		-0.0621
		(1.681)		(-0.504)
Constant	-0.206***	-0.605***	0.0240	0.0214
	(-5.531)	(-5.658)	(0.718)	(0.579)
Observations	59	59	75	75
Adj. R-squared	0.489	0.647	0.504	0.525
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. Columns (1) and (2) represent the financial crisis period between May 2008 and March 2013. Columns (3) and (4) represent the period after the financial crisis period. The dependent variable monthly average transaction price in Amsterdam is a standardized value and transformed to ln. The same applies for the independent variables which are all standardized ln-values. The control variables are the CPI, the mortgage interest rate, the average rent in Amsterdam, the number of households and the population in Amsterdam. T-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.7.3 Spread Transaction value - WOZ value

Table 15 reflects the results of the regression of the search queries on the dependent variable, which is the difference between the transaction value and the WOZ value. Columns (1) and (3) show the relation between the search queries and the difference between the transaction value and the WOZ value. In

Utrecht, every search query is highly significant at a 1% significance level. In Amsterdam there are significant and positive relations for every search query except for 'funda'. After adding the control variables, still three search queries have a significant relation with the spread between the transaction value and the WOZ value. In Utrecht, four search queries remain having significant explanatory power. These results confirm earlier findings that search queries significantly impact the level of liquidity and that search queries are positively correlated with liquidity. Apparently, search queries are also positively correlated with 'overvaluation' (since the difference between the transaction value and the WOZ value is an indication of overvaluation). Attention is for almost every search query highly correlated with overvaluation. These findings therefore suggest the presence of price pressure.

Table 15. Results regression spread transaction value – WOZ value.

Independent variable: spread transaction value - WOZ value				
	Amsterdam		**Utrecht**	
	(1)	(2)	(3)	(4)
Search term 'funda'	0.103	0.486***	0.445***	0.616***
	(1.325)	(4.001)	(4.316)	-4.575
Search term 'huis te koop'	1.035***	1.432***	0.960***	1.434***
	(8.252)	(4.357)	(5.756)	-3.938
Search term 'hypotheek'	0.116*	0.138**	-0.301***	-0.465***
	(1.835)	(2.294)	(-3.584)	(-6.971)
Search term 'huis verkopen'	0.210***	0.0465	0.388***	0.165**
	(3.166)	(0.742)	(4.388)	-2.375
Search term 'makelaar'	0.649***	0.0119	0.797***	-0.201
	(6.030)	(0.0447)	(5.567)	(-0.683)
CPI		-2.542***		-3.701***
		(-8.371)		(-11.00)
Population Amsterdam/Utrecht		5.772***		7.579***
		(4.927)		-5.838
Mortgage interest rate		-0.377		-0.626*
		(-1.188)		(-1.780)
Average Rent Amsterdam/Utrecht		0.246**		1.014***
		(2.073)		-7.705
Number of households		-4.988***		-6.487***
		(-7.621)		(-8.943)
Constant	0.556**	0.570***	0.571*	0.738***
	(2.515)	(3.276)	(1.942)	-3.823
Observations	150	150	149	149
Adj. R-squared	0.982	0.990	0.970	0.988
Controls	No	Yes	No	Yes
Year dummies	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes

The table shows the results of the autoregressive model. The dependent variable is the spread between the transaction value and the WOZ value. The transaction value is a monthly average transaction value. The WOZ value is yearly average value. The regression is done with standardized values in ln which are corrected for seasonality. Columns (1) and (2) represent the data about the housing market in Amsterdam. Columns (3) and (4) represent the data about the housing market in Utrecht. Control variables are added in columns (2) and (4). T-statistics in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

CHAPTER 6 Discussion and interpretations

Various researchers already find a significant relation between investor attention on the internet (e.g. Google search queries, number of clicks on Funda) and the housing market. For example, Van Veldhuizen Vogt and Voogt (2016) find a significant relation between the Google search query 'hypotheek' and the housing market (using a dataset from Statistics Netherlands). Van Dijk and Francke (2018) find a significant relation between the number of clicks on Funda on the one hand and the level of liquidity and prices on the other hand. Berache and Wintoki (2013) find that abnormal Google search intensity can predict future house price changes. In my research, I find similar results.

I find a significant relation at a 1% significance level between the search queries 'funda', 'huis te koop', 'huis verkopen' and 'makelaar'. 'funda' 'huis te koop' and 'makelaar' are positively correlated with the monthly average transaction price, meaning that the monthly average transaction price rises when there is more attention on the internet. The relation between the search query 'huis verkopen' and the monthly average transaction price is negative, meaning that the transaction price decreases when there are more potential sellers trying to sell their house. In Utrecht, I find a significant relation between all the search queries and the monthly average transaction price. After adding the control variables, search terms 'funda' and 'huis te koop' in Amsterdam and 'huis te koop' and 'hypotheek' in Utrecht still have explanatory power in explaining the variance of the monthly average transaction price. The results show that an 1% increase in search queries 'funda' and 'huis te koop' in Amsterdam leads to an increase in future house prices in Amsterdam of 0.156% and 0.625% respectively. An 1% increase in search queries 'huis te koop' and 'hypotheek' leads to a change in future house prices in Utrecht of 0.351% and -0.218% respectively.

From the efficiency test I can detect a significant relation between the monthly average transaction price in Amsterdam and the search term 'funda' in t-1. I also find a significant relationship with the monthly average transaction price and the search term 'huis verkopen' in t-1, t-5 and t-6. The same applies for the search query 'makelaar' in t-12. In Utrecht I also find various significant relations between the monthly average transaction price and the search queries. The Granger causality test shows in most of the cases a Granger causal relation from the search query to the monthly average transaction price.

<u>*Hypothesis 1*</u>

H_0 *Housing related Google search volume indices changes are not correlated with changes in the Dutch housing transaction prices*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the Dutch housing transaction prices*

The findings are more or less in line with the alternative hypothesis of hypothesis 1, stating that housing related Google search volume indices are positively correlated with changes in the Dutch housing transaction prices. This means that changes in the transaction prices can be predicted by the search queries.

There is a significant relation between the search queries ‘funda’, ‘huis te koop’, ‘hypotheek’ and ‘makelaar’ and the monthly transaction volume in Amsterdam. In Utrecht, every search query has a significant relation with the monthly transaction volume. Remarkable is the negative sign in front of the search term ‘funda’, meaning that the transaction volume increases when less people search for the most popular housing website. This might be caused by the fact that the website funda.nl is not only used by people who buy a house, but also by people who sell a house. In periods of economic deterioration, less people want to buy a house and more people want to sell their house and therefore use Funda. More activity on Funda leads to a lower transaction volume. Another explanation can be the tightness in the housing market (Nijskens & Lohuis, 2019). In this case, a lot of people use Funda but because of the tightness in the housing market, the monthly transaction volume decreases. The search terms ‘huis te koop’ and ‘makelaar’ have a positive and significant relation with the monthly transaction volume in both Amsterdam and Utrecht. So, when more people start the process of buying a house, or at least show interest in buying a house, the transaction volume increases. The search query related to liquidity, ‘hypotheek’ is also positively correlated with the monthly transaction volume.

When adding the control variables, the search queries ‘huis verkopen’ and ‘makelaar’ have a significant relation with the monthly transaction volume in Amsterdam. In Utrecht, every search query except for ‘hypotheek’ has significant power in explaining the variance in the monthly transaction volume. Interesting is the change in sign in front of the coefficient. Adding the control variables leads to a positive relation between the search query ‘funda’ and the monthly transaction volume in Utrecht. The monthly transaction volume has especially a strong relation with the control variables ‘CPI’, ‘population’ and the ‘mortgage interest rate’. So, including these variables, the relation can better be explained. After adding the control variables there is still a positive and significant relation between the search queries ‘funda’, ‘huis te koop’ and ‘makelaar’ and the monthly transaction volume.

The efficiency tests show no relation between the lagged values of the search queries ‘funda’, ‘huis te koop’, ‘huis verkopen’ and ‘makelaar’ and the monthly transaction volume in Amsterdam. The monthly transaction volume in Amsterdam is negatively correlated with the search term ‘huis verkopen’ in the seventh lagging period (at a 1% significance level) and positively correlated with the search query ‘hypotheek’ in the first lagging period. It is also significantly correlated with the search term ‘makelaar’ in the first lagging period at a 5% significance level.

Hypothesis 2

H_0 *Housing related Google search volume indices changes are not correlated with changes in transaction volume on the Dutch housing market*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the transaction volume on the Dutch housing market*

Without the control variables, the monthly transaction volume in both Amsterdam and Utrecht is increasing in the Google search indices 'funda', 'huis te koop' 'hypotheek' and 'makelaar'. The alternative hypothesis of hypothesis 2 is therefore applicable. When adding the control variables, the relation loses significance in Amsterdam. In Utrecht, however, the search queries still have a significant impact on the monthly transaction volume.

Next, I discuss the relation between the monthly average excess return and the search queries. I find a negative and significant relation between the search query 'funda' and the monthly average excess return in Amsterdam. I do find a positive relation between the search query 'hypotheek' and the monthly average excess return in Utrecht, at a significance level of 1%. The negative relation can be explained as follows: excess return can both reflect a positive and a negative value. Since I find negative excess returns in the sample period, a negative relation between the investor attention and the excess return seems logical. More investor attention decreases the negative abnormal returns (i.e. the absolute value of the excess returns). Adding the control variables, does not impact the results.

Looking at the efficiency test, lagged values of 'hypotheek' are correlated with the monthly average excess return at a 5% significance level in Amsterdam. In Utrecht, lagged values of every search query significantly impact the relation with the monthly average excess return. So, excess return can be explained by lagged values of the search queries. The results indicate that potential house buyers start the process of buying a house months before they actually buy a house.

Hypothesis 3

H_0 *Housing related Google search volume indices changes are not correlated with changes in excess return on the Dutch housing market*

H_α *Housing related Google search volume indices changes are positively correlated with changes in the excess return on the Dutch housing market*

When looking at the relation between the Google search queries and the monthly average excess returns in Amsterdam in the current period, there are no significant relations. In Utrecht, the search terms 'funda' and 'makelaar' are significantly correlated with the monthly average excess return. After including the control variables and the lagged value of the monthly average excess returns, only the relation with the search term 'makelaar' remains significant. The alternative hypothesis of hypothesis 3 therefore only applies for the search term 'makelaar' in Utrecht.

The results show that lagged values of either the monthly average transaction price, monthly transaction volume and monthly average excess return significantly impact the current transaction price, transaction volume and excess return. A random walk pattern is therefore not present. Prices are not efficient. Does the inclusion of search queries lead to an increase in price efficiency? Or does it strengthen the price pressure effect? The negative relation between the search query 'makelaar' and the monthly average excess return in Utrecht suggests that excess return is decreasing when attention is increasing. This implies that prices in Utrecht deviate less from their intrinsic value when investor attention is increasing and that more information is processed, which might be evidence for the statement that Google search behaviour increases price efficiency (Fama, 1970). An increase of price efficiency can also be measured by the speed information is processed (Drake et al., 2012). This, however, is not examined in this master thesis. Unfortunately, I do not find other significant results to strengthen these findings. A positive relation between investor attention and monthly trading volume, indicates the presence of price pressure (Barber and Odean, 2008). Since I find a significant relation between four search queries and the monthly trading volume in Amsterdam and Utrecht, these findings suggest that the search queries strengthen the price pressure effect. The same applies for the relation between the search queries and the monthly average transaction price. The more people are paying attention to the search queries, the more the monthly average transaction price increases.

The robustness checks also suggest the presence of price pressure. Making a distinction between the relation between search queries during and after the financial crisis, leads to the conclusion that the search queries have a stronger relation with the monthly average transaction price and monthly transaction volume during the crisis compared to after the crisis. Investor attention has a larger impact on the average transaction price during periods of high volatility and uncertainty (Drake et al., 2012). Another robustness check is the spread between the transaction value and the WOZ value. The search queries are highly significantly and positively correlated with this spread. Investor attention is high when prices are more overvalued (Drake et al., 2012). Herd behaviour might leads to inflation of prices. Fear of not getting a chance to buy a house, may lead to generously overbidding the list price (Shiller, 2016). This confirms the price pressure theory.

Hypothesis 4

H_0 *Internet search behaviour does not increase price efficiency on the Dutch housing market*

H_α *Internet search behaviour does increase price efficiency on the Dutch housing market*

Given the results, investor attention overall leads to an increase of the price pressure effect. There is little evidence that more attention leads to an increase in price efficiency. The null hypothesis of hypothesis 4 is therefore applicable.

The following four hypotheses are dedicated to the liquidity. First, I test the extent in which liquidity is correlated with the monthly average transaction price, volume and excess return. I need to test this relation first, before I test hypothesis 8, which elaborates on whether liquidity is correlated with the Google search queries. I test the correlation between the liquidity on the one hand and the monthly average transaction price, transaction volume and excess return on the other hand, with a Spearman test. Three liquidity measures represent 'liquidity': (i) the spread between the list price and the transaction price, (ii) the number of days a particular dwelling is on the market and (iii) the downward price adjustment of the list price.

Hypothesis 5:
H_0 *Liquidity is not correlated with the monthly average transaction price*
H_α *Liquidity is positively correlated with the monthly average transaction price*

Hypothesis 6:
H_0 *Liquidity is not correlated with the monthly transaction volume*
H_α *Liquidity is positively correlated with the monthly transaction volume*

Hypothesis 7:
H_0 *Liquidity is not correlated with the monthly excess return*
H_α *Liquidity is positively correlated with the monthly excess return*

The Spearman's tests show that the relation between the liquidity and the average transaction price in both Amsterdam and Utrecht is strong/very strong. Liquidity is positively correlated with the monthly average transaction price. The alternative hypothesis of hypothesis 5 applies. Liquidity is moderately/strongly and positively correlated with the monthly transaction volume in Amsterdam and Utrecht. The alternative hypothesis of hypothesis 6 therefore applies. Liquidity is moderately and positively correlated with the excess return in Amsterdam and Utrecht. The alternative hypothesis of hypothesis 7 therefore applies. Knowing that the liquidity is positively correlated with the monthly average transaction price, the monthly transaction volume and the monthly average excess return, one would suggest that the Google search queries that have a significant relation with the monthly average transaction price, transaction volume and excess return, also might have a significant relation with the liquidity measures.

Google search queries have a significant relation with the liquidity measure 'days on market' in Amsterdam. After adding the control variables, still four search queries are significantly correlated with the monthly average listing period. In Utrecht, three search queries have a significant relation with the liquidity measure *days on market*. After adding the control variables, still three Google search queries have a significant impact on the liquidity measure. The liquidity measure 'spread transaction-list price' is less significantly correlated with the search queries. Still, 'huis te koop' has a significant impact at a 5% significance level in Amsterdam and Utrecht and 'funda' is significant at a 5% significance level in Utrecht. Three search queries are significantly correlated with the liquidity measure 'downward price adjustment' in Amsterdam and four search queries are significantly correlated with this liquidity measure in Utrecht.

<u>*Hypothesis 8:*</u> H_0 *Liquidity is not correlated with housing related Google search volume indices changes* H_α *Liquidity is positively correlated with housing related Google search volume indices changes*

From the results I can conclude that liquidity is positively correlated with the Google search volume indices. The alternative hypothesis of hypothesis 8 applies. Knowing that liquidity is positively correlated with the monthly average transaction price, the volume and the excess return means that analyzing the relation between liquidity and investor attention can be useful in determining which direction the housing market will go.

This research is applied on data of two large cities in the Netherlands: Amsterdam and Utrecht. The results between these two cities differ sometimes and are not always consistent. This can be explained by the heterogeneity among buyers (Qiu et al., 2020). For example, in Amsterdam, there are more foreign investors who are not from the Netherlands (De Nederlandsche Bank, 2019). The results of this master thesis cannot directly be generalized to the Netherlands as a whole. Specifically: results about the relation between particular search queries and the dependent variable cannot directly be used for a generalization. However, the fact that movements in the housing market can be predicted by the use of search queries, can be generalized.

CHAPTER 7 Conclusion

The central question in this research is whether internet search behaviour is a good predictor of movements in the Dutch housing market. To answer this question, I test different hypotheses. I use Google search volume indices as a proxy for investor attention on the internet and I use transaction data about the housing market in Utrecht and Amsterdam. I use a sample period of January 2007 - June 2019. Firstly, I find that Google search queries are positively correlated with average transaction prices in both Amsterdam and Utrecht. For example, an 1% increase in the search query 'huis te koop' leads to an increase of 0.156% of the future average house price in Amsterdam and an increase of 0.351% in Utrecht (after adding the control variables to the regression), corresponding to an increase of the future average house price of € 520.09 in Amsterdam and € 966.71 in Utrecht. I also find that some lagged values of these search queries are significantly correlated with the monthly average transaction price, suggesting that the process of buying a house starts with searching on the internet, months before the transaction takes place. These findings confirm the findings of van Veldhuizen et al. (2016), van Dijk & Francke (2018) and Berache &Wintoki (2013).

Next, I find that Google search volume indices are positively correlated with the transaction volume in both Amsterdam and Utrecht. For example, an increase of 1% of the search query 'makelaar' leads to an increase of 1.092% of the future monthly average transaction volume in Amsterdam and 1.381% in Utrecht, corresponding to seven more monthly transactions in Amsterdam and four more monthly transactions in Utrecht. Apparently, sellers will only sell their house when the housing market is subject to positive attention (Genesove & Mayer, 2001). Furthermore, I find that Google search volume indices are significantly correlated with the excess return in both Amsterdam and Utrecht, which confirms the findings of Da et al. (2011). Some lagged Google search queries are significantly correlated with the trading volume and excess return in Amsterdam. Based on these findings I conclude that a random walk is not always present and that prices are not efficient (Fama, 1970).

The results suggest the presence of price pressure. The robustness checks confirm this suggestion: Google search queries are more significantly correlated with the transaction price and the transaction volume during the financial crisis compared to after the financial crisis. There is more investor attention during a more volatile period. Furthermore, I find that Google search queries are highly positively correlated with the spread between the transaction value and the WOZ value, suggesting that more investor attention leads to more overvaluation of transaction prices. This confirms the statement that the housing market is prone to investor sentiment, since the capital gains are not justified by the intrinsic value (Baker & Wurgler, 2007). Herd behaviour might be present (Shiller, 2005; Shiller, 2016). The price pressure effect might be intensified by the housing shortage in the Netherlands (Boelhouwer, 2005) in combination with the fear of not being able to buy a house (Kishore, 2006; Farlow, 2004b; Shiller, 2016). The findings of the robustness checks are in line with Drake et al.

(2012), who find that the abnormal search volume is higher when spreads are higher and when idiosyncratic volatility is higher.

Also, I find a significant relation between Google search behaviour and liquidity. First, I find that liquidity has a very strong correlation with the monthly average transaction price and that the liquidity has a moderate/strong relation with the monthly average transaction price and excess return. Next, I find that Google search volume indices are significantly correlated with the liquidity measures, which is in line with van Dijk & Francke (2018). The findings suggest that there is more liquidity when investor attention is high. Knowing that liquidity is positively correlated with the monthly average transaction price, volume and excess return, this means that analyzing the relation between the liquidity and investor attention can be useful in determining which direction the housing market will go.

All these findings together, I can answer the research question by concluding that internet search behaviour is a good predictor for movements in the housing market. The findings of this research are useful for both retail and professional investors. The keywords I use in this research, are – in most of the cases – positively and significantly correlated with the transaction price and are – in most of the cases – Granger causing the transaction price. This means that (retail) investors can predict movements in the housing market with a free and user-friendly tool: Google Trends. in a market that is known for its inefficiency, this tool can therefore be very useful.

This research has its limitations. In general, search queries *do* have predictive power. However, results about an individual search query in a particular city do not directly imply a generalized significance for this particular search query. Furthermore, Tu, de Haan and Boelhouwer (2016) show that there exists a mismatch between house price models and the regulatory environment in the Netherlands. Besides that, imputed rent levels are measured with noise (Ayuso & Restoy, 2006). This makes the calculation of the excess returns less reliable. Lastly, the master thesis is only applied on the housing market in Utrecht and Amsterdam. Further research with transaction data on the whole of the Netherlands is recommendable.

Also, from an econometric point of view the research has its limitations. The results might be driven by unobserved heterogeneities, as already mentioned in the endogeneity section. For example, the regulatory environment (Tu et al., 2016) and the stock market (Kakes & Van den End, 2004) are not included in the regression. I was not able to get proper data about the yearly construction rate (or additions to the housing stock), which is an important factor of the housing supply. I furthermore do not make a distinction between the type of house and the zip code a particular transaction took place.

REFERENCES

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 370 374.

Andrle, M., & Plasil, M. (2019). *Assessing House Prices with Prudential and Valuation Measures* (IMF Working Paper No. 19/59).

Ayuso, J., & Restoy, F. (2006). House prices and rents. An equilibrium asset pricing approach. *Journal of Emperical Finance, 13*(3), 371-388.

Baker, M., & Wurgler J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645-1680.

Barber, B.M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies, 21*(2), 785 818.

Barberis, N., Schleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics, 49*(3), 307-343.

Belotti, F., Daidone, S., & Ilardi, G. (2013). Stochastic frontier analysis using Stata. *The Stata Journal, 13*(4), 719-758.

Ben-Rephael, A., Da, Z., & Israelsen, R.D. (2017). It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. *The Review of Financial Studies, 30*(9), 3009-3047.

Beracha, E., & Wintoki, E.B. (2013). Forecasting Residential Real Estate Price Changes from Online Search Activity, *Journal of Real Estate Research, 35*(3), 283-312.

Boelhouwer, P.J. (2005). The incomplete privatization of the Dutch housing market: Exploding house prices versus falling house-building output. *Journal of Housing and the Built Environment, 20*(4), 363-378.

Boelhouwer, P.J., & Hoekstra, J. (2009). Towards a Better Balance on the Dutch Housing Market? Analysis and Policy Propositions. *European Journal of Housing Policy, 9*(4), 457-475.

Braun, N. (2016). Google search volume sentiment and its impact on REIT market movements. *Journal of Property Investment & Finance, 34*(3), 249-262.

Brown, L.A., & Moore, E.G. (1970). The Intra-Urban Migration Process: A Perspective. Geografiska Annaler. *Series B, Human Geography, 52*(1), 1-13.

Case, K.E., Quigley, J.M., & Shiller, R.J. (2003). Home-buyers, Housing and the Macroeconomy. Paper presented at *the Conference on Asset Prices and Monetary Policy*.

Case, K.E., & Shiller R.J. (1990). Forecasting Prices and Excess Return in the Housing Market. *Real Estate Economics*, *18*(3), 253-273.

Centraal Bureau voor de Statistiek. (2020). *Bestaande koopwoningen; verkoopprijzen prijsindex 2015=100*. Retrieved from https://www.cbs.nl/nl nl/cijfers/detail/83906NED?q=cpi%20wonen.

Conclusr. (2014). Makelaardijmonitor 2014: Benchmark huizensites. Conclusr Research Report.

Da, Z, Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance, 61*(5), p. 1461 1499.

Daniel, K., Hirshleifer, D. & Subrahmanyam, A. (1998). Investor Psychology and Security Market Under- and Overreactions. *The Journal of Finance,* 53(6), 1839-1885.

DeBondt, W. (1995). Real Estate Cycles and Animal Spirits. In J. Pagliari (Ed.), *The Real Estate Portfolio Handbook*. Chicago: Richard Irwin.

DeBondt, W., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance, 40*(3), 793 805.

Dellavigna, S., & Pollet, J.M. (2009). *Investor Inattention, Firm Reaction, and Friday Earnings Announcements* (NBER Working Paper No. 11683).

De Nederlandse Bank. (2020). *Deposito's en leningen van MFI's aan huishoudens, rentepercentages, gecorrigeerd voor breuken (Maand). https://statistiek.dnb.nl/downloads/index.aspx#/details/deposito-s-en-leningen-van-mfi-s-aan huishoudens-rentepercentages-gecorrigeerd-voor-breuken-maand/dataset/efba2d4e-fb53 49a8-a1fe-d5ee3263e14c/resource/8d3ccc86-8396-43b8-a18b-5ba293f01c1d.*

De Nederlandsche Bank. (2017). *Overzicht Financiële Stabiliteit.* Retrieved from https://www.dnb.nl/binaries/OFS_Najaar17%20WEB_tcm46-363954.pdf.

De Nederlandsche Bank. (2019). *Amsterdam housing market resembles Paris and Frankfurt more closely than rest of Netherlands.* Retrieved from https://www.dnb.nl/en/news/news-and archive/Nieuws2019/dnb384913.jsp#.

Dieleman, F.M. (2001). Modelling residential mobility; a review of recent trends in research. *Journal of Housing and the Built Environment*, *16*(3)*,* 303-329.

Van Dijk, D., & Francke, M. (2015). *Internet search behavior, liquidity and prices in the housing market* (DNB Working Paper No. 481).

Drake, M.S., Roulstone, D.T., & Thornock, J.R. (2012). Evidence from Google Searches Around Earnings Announcements. *Journal of Accounting Research, 50*(4), 1001-1040.

Fama, E.F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal*, *21*(5), 55-59.

Fama, E.F. (1970). Efficient Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417.

Fama, E.F. (1998). Market Efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics 49*(1), 283-306.

Farlow, A. (2004a). UK house prices: a critical assessment. Paper presented at *the Credit Suisse First Boston Housing Market Conference.*

Farlow, A. (2004b). The UK Housing Market: Bubbles and Buyers. Paper presented at *the Credit Suisse First Boston Housing Market Conference.*

Ferguson, N. (2008). *The Ascent of Money*. NY: Penguin Press.

Gallimore, P., & Gray, A. (2002). The role of investor sentiment in property investment decisions. *Journal of Property Research*, *19*(2), 111-120.

Gallimore, P., Hansz, J.A., & Gray, A. (2000). Decision making in small property companies. *Journal of Property Investment & Finance*, *18*(6), 602-612.

Genesove, D., & Mayer, C. (2001). Loss Aversion and Seller Behavior: Evidence from the Housing Market. *The Quarterly Journal of Economics, 116*(4), 1233-1260.

Glaeser, E., J., Gyourko, & Saiz, A. Housing Supply and Housing Bubbles. *Journal of Urban Economics, 64*(2), 198-217.

Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica, 37*(3), 424-439.

Van Haaren, J. (2020). *Real Estate Economics* [PDF slides]. Retrieved from https://canvas.eur.nl/courses/27433/files/20058899?module_item_id=409694

Hair, J.F. Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2017). *A primer on Partial Least Squares Structural Modeling (PLS-SEM).* CA Sage, Thousand Oaks.

Harding, J.P., Rosenthal, S.S., & Sirmans C.F. (2006). Depreciation of housing capital, maintenance, and house price inflation: estimates from a repeat sales model. *Journal of Urban Economics, 61*(2), 193-217.

Hauke, J., & Kossowski, T. (2011). Comparison of Pearson's and Spearman's Correlation
Coefficients on the Same Sets of Data. *Quaestiones Geographicae, 30*(2)*,* 87-93.

Hidalgo, M.C., & Hernandez, B. (2001). Place attachment: Conceptual and Empirical Questions. *Journal of Environmental Psychology, 31*(3), 273-281.

Hott, C. & Monnin, P. (2006). *Fundamental Real Estate Prices: An Emperical Estimation with International Data* (NCCR FINRISK Working Paper No. 356).

Hirschleifer, D., Lim, S.S., & Teoh, S.H. Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *The Journal of Finance, 44*(5), 2289-2325.

Hu, D., Lou, X., Xu, Z., Meng, N., Xie, Q., Zhang, M., Zou, Y., Liu, J., Sun, P.G. & Wang, F. (2020). More Effective Strategies are Required to Strengthen Public Awareness of COVID-19: Evidence from Google Trends. *Journal of Global Health, 10*(1), 1-12.

Hull, J.C. (2018). *Risk Management and Financial Institutions*. John Wiley & Sons, Inc., New Jersey.

International Monetary Fund. (2008). *World Economic Outlook, April 2008: Housing and the Business Cycle.* Retrieved from https://www.imf.org/en/Publications/WEO/Issues/2016/12/31/World Economic-Outlook-April-2008-Housing-and-the-Business-Cycle-20359.

Jarsulic, M. (2010). *Anatomy of a Financial Crisis. A Real Estate Bubble, Runaway Credit Markets, and Regulatory Failure*. Palgrave Macmillan, New York.

Joseph, K., Wintoki, M.B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting, 27*(4)*,* 1116-1127.

Jud, G.D., Winkler, D.T., & Kissling, G.E. (1995). Price Spreads and Residential Housing Market Liquidity. *Journal of Real Estate Finance and Economics, 11*(3), 251-260.

Kahneman, D. (1973). *Attention and Effort*. Prentice Hall: Englewood Cliffs.

Kakes, J., & Van den End, J.W. (2004). Do stock prices affect house prices? Evidence for the Netherlands. *Applied Economics Letters, 11*(12), 741-744.

Kishore, R. (2006). Theory of Behavioural Finance and its Application to Property Market: A Change to Paradigm. Paper presented at *the Twelfth Annual Pacific Rim Real Estate Society Conference*.

Kranendonk, H., & Verbruggen, J. (2008). *Is de huizenprijs in Nederland overgewaardeerd?* (CPB Memorandum No. 81).

Levin, E.J., Montagnoli, A., & Pryce, G. Mark-to-market and house asset valuation. An initial attempt at extending the Poterba model using the term structure of real forward interest rates. *International Journal of Housing Markets and Analysis, 4*(2), 172-179.

Levitt, S.D., & Dubner, S.J. (2005). *Freakonomics: a rogue economist explores the hidden side of everything*. New York: William Morrow.

Levy, D., & Murphy, L., & Lee, C.K.C. (2008). Influences and Emotions: Exploring Family Decision making Processes when Buying a House. *Housing Studies, 23*(2), 271-289.

Malmendier, U., & Tate, G. (2005). CEO Overconfidence and Corporate Investment. *The Journal of Finance*, *60*(6), 2661-2700.

Manzo, L.C. (2003). Beyond house and haven: towards a revisioning of emotional relationship with places. *Journal of Environmental Psychology, 23*(1), 47-61.

Merton, R. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *Journal of Finance, 42*(3), 483-510.

Meshcheryakov, A. (2018). Using Online Search Queries in Real Estate Research with an Empirical Example of Arson Forecast. *Journal of Real Estate Literature, 26*(2), 331-361.

Moat, S.H., Preis, T. & Stanley, H.E. (2013). Quantifying Trading Behavior in Financial Market Using Google Trends. *Scientific Reports, 3*(1684), 1-6.

Morawski, J. (2008). *Investment Decisions on Illiquid Assets. A Search Theoretical Approach to Real Estate Liquidity*. Wiesbaden: GWV Fachverlage GmbH.

Mulder, C.H., & Hooimeijer, P. (1999). Residential relocations in the life course. In L.J.G. Wissen & P.A. Dykstra (Ed.), *Population Issues. An Interdisciplinary Focus*. New York: Kluwer.

Nenov, P.T., Larsen, E.R., & Sommervoll, D.E. (2015). Thick-market Effects, Housing Heterogeneity, and the Determinants of Transaction Seasonaility. *The Economic Journal, 126(*12), 2402-2423.

Nijskens, R., & Lohuis, M. (2019). The Housing Market in Major Dutch Cities. In: Nijskens, R., Lohuis, M., Hilbers, P., & Heeringa, W. (Ed.) *Hot Property.* Springer, Cham.

Organisation for Economic Coorperation and Development. (2020). *Household disposable income*. https://www.cbs.nl/nl-nl/cijfers/detail/83906NED?q=cpi%20wonen.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society Ser. A, 187*(1), 253-318.

Poterba, J.M. (1992). *Taxation and housing: old questions, new answers* (NBER Working Paper No. 3963).

Qui, L., Tu, Y., & Zhao, D. (2020). Information asymmetry and anchoring in the housing market: a stochastic frontier approach. *Journal of Housing and the Built Environment 35*, 573-591.

Quigley, J.M. (1999). Real Estate Prices and Economic Cycles. *International Real Estate Review*, *2*(1), 1-20.

Salkind, N.J. (2007). *Encyclopedia of Measurement and Statistics*. Sage Publications, Inc., Thousand Oaks.

Salzman, D., & Zwinkels, R.C.J. (2017). Behavioural Real Estate. *Journal of Real Estate Literature, 25*(1), 77-106.

Savills. (2019). *Utrecht - Where to invest(next)?*. Retrieved from https://pdf.euro.savills.co.uk/ the-netherlands/commercial---dutch-other/utrecht-city-special-2019.pdf.

Van der Schaar, J. (1987). *Groei en bloei van het Nederlandse Volkshuisvestingsbeleid*. Delftse University Press, Delft.

Shiller, J.R. (2007). *Understanding Recent Trends in House Prices and Home Ownerships* (NBER Working Paper 13553).

Shiller, J.R. (2016). *Irrational Exuberance*. Princeton University Press, Princeton.

Sirgy, M., Grzeskowiak, S., & Su, C. (2005). Explaining housing preference and choice: The role of self-congruity and functional congruity. *Journal of Housing and the Built Environment, 20*(4), 329-347.

Sornette, D., & Woodard, R. (2010). Financial Bubbles, Real Estate Bubbles, Derivative Bubbles, and the Financial and Economic Crisis. In: Takayasu M., Watanabe T., Takayasu H. (Ed.) *Econophysics Approaches to Large-Scale Business Data and Financial Crisis*. Springer, Tokyo.

Spearman, C.E. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*(1), 72-101.

Statista. (2020a, April 27). *Global digital population as of April 2020.* Retrieved from https://www.statista.com/statistics/617136/digital-population-worldwide/.

Statista. (2020b, June 8). *Household indebtedness to gross disposable income ratio in Europe as of the 2nd quarter 2019, by country*. Retrieved from https://www.statista.com/statistics/1073593/household-debt-ratio-europe-by-country/

Taltavull de la Paz, P., & White, M. (2016). The sources of house price change: identifying liquidity shocks to the housing market. *Journal of European Real Estate, 9*(1), 98-120.

Tetlock, P.C., Griffin, J., Hahn, R.W., Han, B., Kumar, A.H., Murray, T.A., Parsons, C., & Starks, L.T. (2007). *Does Liquidity Affect Securities Market Efficiency* (Working Paper).

Tu, Q., de Haan, J., & Boelhouwer, P. (2018). House prices and long-term equilibrium in the regulated market of the Netherlands. *Housing Studies, 33*(4), 408-432.

Tu, Q, de Haan, J., & Boelhouwer, P. (2016). The mismatch between conventional house price modeling and regulated markets: insights from the Netherlands. *Journal of Housing and the Built Environment, 32*(3), 599-619.

Thwaites, G., & Wood, R. (2003). The measurement of house prices. *Bank of England. Quarterly Bulletin, 43*(1), 38-46.

Ullah, S., Akhtar, P., & Zaefarian, G. (2018). Dealing with Endogeneity Bias: The Generalized Method of Moments (GMM) for Panel Data. *Industrial Marketing Management, 71*(2), 69-78.

Van Veldhuizen, S., Vogt, B., & Voogt, B. (2016). Internet searches and transactions on the Dutch housing market. *Applied Economics Letters, 23*(18), 1321-1324.

Verbruggen, J., Kranendonk, H., van Leuvensteijn, M., & Toet, M. (2006). *Welke factoren bepalen de ontwikkeling van de huizenprijzen in Nederland*? Retrieved from https://www.cpb.nl/sites/default/files/publicaties/download/welke-factoren-bepalen-de ontwikkeling-van-de-huizenprijs-nederland.pdf

Wilhelmsson, M. (2008). House price depreciation rates and level of maintenance. *Journal of Housing Economics, 17*(1)*,* 88-101.

Wong, S.K., Yiu C.Y., & Chau, K.W. (2011). Liquidity and Information Asymmetry in the Real Estate Market. *Real Estate Economics, 45*(1)*,* 49-62.

Wu, L., & Brynjolfsson. (2015). The Future of Prediction. How Google Searches Foreshadow Housing Prices and Sales. In Goldfarb, A., Greenstein, S.M., & Tucker, C.E. (Ed.). *Economic Analysis of the Digital Economy*. University of Chicago Press, Chicago.

Wu, J., & Deng, Y. (2015). Intercity Information Diffusion and Price Discovery in Housing Markets: Evidence from Google Searches. *The Journal of Real Estate Finance and Economics*, *50*(3), 289-306.

Zhu, H., Xiong, H., Tang, F., Ge, Y., Chen, E., & Fu, Y. (2019). Days on Market: Measuring Liquidity in Real Estate Markets. Paper presented at *the 22nd ACM SIGKDD International Conference*.