NTNU
Norwegian University of
Science and Technology

Can Google Search be Used as a Housing Bubble Indicator?

- a US 2006/07 Bubble Case Study

Ole Martin Eidjord

Industrial Economics and Technology Management
Submission date: June 2017
Supervisor: Sjur Westgaard, IØT
Co-supervisor: Are Oust, NTNU Handelshøyskolen

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Problem Description

In this paper, we investigate whether Google search query data can be used to operationalize point 2, 3, 4, 5 and 6 in Shiller`s (2010) checklist for asset pricing bubbles. We investigate the following key questions: can search terms, related to housing, discover the states experiencing a housing bubble during the U.S. housing market crash in 06/07? Can search volume levels for the same queries predict the housing trend and the house prices at state level? Can the inclusion of Google searches improve the prediction of price models? Do Google search have higher predictive power than the well-established Consumer Confidence Index?

Assignment was given on January 12, 2017

Supervisor: Associate Professor Are Oust

Preface

This paper, TIØ4900 Financial Engineering, Master`s Thesis, represents the final assignment of the Master of Science in Finance program through the Department of Industrial Economics and Technology Management at the University of Science and Technology in Trondheim, Norway. The paper is a continuation of TIØ4550 Financial Engineering, Specialization Project that I wrote together with Harald Eskerud autumn 2016.

I would like to direct massive thanks to Are Oust who was my supervisor and a tremendous help for me! Autumn 2016, Are Oust gave me and Harald Eskerud the idea of using Google to operationalize Shiller`s checklist points for asset pricing bubbles. Are have been a motivating supporter, in addition to giving professional support. I am also grateful for valuable inputs from Aras Kj who helped me choose suitable models and interpret the result.

Trondheim, June 8, 2017

Ole Martin Eidjord

Abstract

The aim of this paper is to operationalize five out of seven points in Shiller's (2010) asset-pricing bubble checklist, using Google search. We start with 204 housing related queries and reduce them to 20 based on their correlation with the house prices in identified bubble states, from the 2006/07 US housing market crash. Next, we test the search terms and four self-created indexes ability to indicate the states experiencing a bubble, based on differences in search volume during and after the bubble period. We find Google search for *Housing Bubble* to perform best in wide number of tests, and conclude that it can be a strong bubble indicator. Google search for *Real Estate Agent* displayed the most predictive power for the house prices, of all the queries and indexes tested, globally in the US. Constructing a simple linear model using only Google search for *Real Estate Agent* and a one period lag of the dependent variable, produced good in-sample prediction results at state level, in both the short and long run. Including the Google searches in a baseline error correction model, improved all points of criteria. The adjusted coefficient of determination increases for both the short and long run and the speed of adjustment is higher and more significant. Substituting Google searches with the well-established Consumer Confidence Index yielded worse result for all assessments. Due to their huge impact on the economy and the difficulty of discovering them, housing bubble indicators are of interest for academic purposes and policy makers such as banks, governments, and asset managers.

Keywords: Google Trends, Housing, Cointegration, Housing Bubble, Real Estate Agent

Sammendrag

Hensikten med denne oppgaven er å bruke Google søk for å operasjonalisere fem av syv punkter i Shiller (2010) sin sjekkliste for bobler. Vi starter med 204 boligrelaterte søkeord og reduserer dem til 20 ved å teste korrelasjonen til boligprisen for de statene som opplevde en boligboble under den amerikanske boligkollapsen i 2006/07. Videre tester vi om søkeordene og fire selvopprettede indekser kan indikere stater som opplevde en boble, basert på forskjeller i søkevolum under og etter boligboblen. Vi finner at Google søk for *Housing Bubble* presterer best i et stort antall tester og kan fungere som en sterk bobleindikator. Google søk for *Real Estate Agent* har størst forklaringsevne på boligprisene, av søkeordene og indeksene som ble testet, globalt i USA. Ved å konstruere en enkel linear modell som kun bruker Google søk for *Real Estate Agent* og en lag av den avhengige variabelen gir gode, kortsiktige og langsiktige, prediktive resultater på statsnivå. Ved å inkludere Google søkene i en standard error correction modell blir alle vurderingspunktene forbedret. Forklaringsgraden til modellen øker på både kort og lang sikt samt den beveger seg raskere mot den langsiktige likevekten. Ved å substituere Google søkene med en veletablert indeks for forbrukeroptimisme reduseres modellen på alle punkter. Boligbobler har enorme økonomiske konsekvenser og er vanskelig å oppdage før de sprekker. Indikatorer er derfor av stor interesse for akademiske formål og beslutningstakere som banker, regjeringer og kapitalforvaltere.

Table of Contents

1 Introduction

Asset-pricing bubbles have been the cause for some of history's biggest economic downturns. Housing bubbles, in particular, have a massive impact on the economy and tend to have longer-term effects than other types of bubbles. This can be explained by wealth and amplification effects. The housing comprises the majority of many households' wealth, and the wealth effect on consumption is significant and apparently larger than the wealth effect of financial assets (see e.g. Case et al. (2001); Benjamin et al. (2004); Campbell and Cocco, 2004). Also, amplification mechanisms play a significant role. Spillover effects from a housing bubble can, in particular, be major due to the large share of housing debt in bank portfolios. Amplification mechanisms that arise during financial crises can be either direct, i.e. caused by direct contractual links, or indirect, i.e. caused by spillovers or externalities that are due to common exposure or the endogenous response of various market participants (Brunnermeier and Oehmke, 2012). Due to their huge impact, housing bubble indicators are of interest for academic purposes, for policy makers such as banks, governments, investors, Insurance companies and asset managers, for homeowners and the public.

Despite their significant impact housing price bubbles are notoriously difficult to discover. Therefore, it is of great interest to find tools that can help discover housing bubbles and improve the prediction of house prices. We use Google Trends data to operationalize five of Shiller's (2010) checklist points for asset pricing bubbles. Through Google search volume level for queries related to housing, we measure animal spirits as described by Akerlof and Shiller (2009). They argue that human psychology has a significant impact on economic decisions and the aggregate economy, and are highly critical of the assumption of rational decision-making. In contrast to the efficient-market hypothesis, (see e.g. Fama, 2014). Akerlof and Shiller (2009) further argue that confidence is the most important animal spirit in determining behaviour and that it plays a major role in the business cycle. This is because confidence, or the lack of it, has a big impact on people's purchase decisions. Put simply, when people are confident they spend more, which fuel the economy; lack of confidence makes people withdraw and sell, slowing down the economy. Feedback loop theory, as described by Shiller (2005) in Irrational Exuberance, is closely related to the confidence multiplier. In general, feedback loop theory argues that initial price increases lead to stories about price increases and people making money.

The stories and the effects of the initial price increases feed back into more stories and higher prices through increased investor demand. Thus, the total effect of increased confidence and price increases is much greater than the initial effect and stimulus caused by it.

Development in information technology and the widespread use of search engines enables a new way of predicting the future (see e.g., Kuruzovich et al. (2008); Horrigan (2008); Choi and Varian, 2009). Forecasting has traditionally relied on statistical information gathered by the government and private companies. Reports based on this information are published with lags of several months and quarters for certain parts of the economy such as the housing market. According to the National Association of Realtors (2016), a typical homeowner takes three months to buy but engage with agents earlier in the process and 83% of all home respondents frequently use the internet to search for their home. Thus, millions of persons are at all time searching through search engines, looking for a home, leaving behind economic intentions about their future economic behaviour. Pentland (2010) found Google searches to precede purchase decisions and in many cases to be a more "honest signal" of actual interests and preferences because no bargaining, gaming, or strategic signalling is involved, in contrast to many market-based transactions or other types of data gathering such as surveys. Google is by far the biggest search engine and has since the beginning of 2004 published indexed search volumes at Google Trend. This information is free and easily available for different geographic regions such as country, state and metro level. Google Trends data have become increasingly popular in (financial) econometrics in recent years (see e.g. Bijl et al. 2015; Preis et al. 2010 and 2013). Wu and Brynjolfson (2009; 2015) find evidence that queries submitted to Google`s search engine are correlated with both the volume of housing sales as well as a house price index – specifically the Case-Shiller index – released by the Federal Housing Agency. They further found that search queries can reveal the current housing trend, but Google search is especially well suited for predicting the future unit sales of housing.

Analysing the U.S. housing market, we find that four states experienced a *real* bubble during the 06/07 housing market crash. Several states experienced a major increase followed by a significant decrease in the house prices. We define the six successor states, sorted after the largest decline in housing prices, as *minor* bubble states. These bubble states, along with the ten states that experienced the smallest price decrease, are used as benchmark states in an in-sample bubble identification test. Based on our review of asset pricing bubble literature, we identify 204 search terms related to housing bubbles and the real estate market and reduces these to twenty queries by testing for correlation between the house prices in the identified bubble states. Next, we propose a housing bubble identification approach based on the differences in Google Search Volume Index, henceforth GSVI, levels in the housing bubble period compared to a non-bubble period. We find that GSVI for Housing Bubble and Real Estate Agent performs best of the single search terms in the in-sample prediction and that they also outperform the self-created indexes consisting of the average GSVI for different search terms. Furthermore, GSVI for the two queries is both highly correlated with the Housing Price Index, henceforth HPI. GSVI for Housing Bubble performs especially well on finding a global housing bubble for the United States and indication of bubbles at state level. Taking predictive abilities, simplicity and robustness into account, GSVI for Housing Bubble is considered the best candidate as a housing bubble indicator. When optimising the result about finding all bubble states, GSVI for Housing Bubble indicates all bubble states and erroneously indicates bubbles in only one non-bubble state. Changing the objective to not erroneously detecting non-bubble states as bubbles, GSVI for Housing Bubble indicates bubbles in all four real bubble states and four out of six minor bubble states.

Predicting the house prices globally in the U.S. with GSVI for Housing Bubble, Real Estate Agent and the best performing index, we found GSVI for Real Estate Agent to give the best results. GSVI for Real Estate Agent displays the highest correlation with HPI, especially for the non-bubble period. The correlation between them is largest when we use lagged values for the Google searches, implying Real Estate Agent is leading the house prices. Furthermore, we find the two time-series to be cointegrated, and there is a long run effect running from GSVI for Real Estate Agent to HPI. This effect is strongest in the states experiencing a real bubble, somewhat less for the states experiencing a minor bubble and the least significant for the non-bubble states. GSVI for Real Estate Agent show good in-sample predictive abilities at the state level, using simple linear models including only GSVI, and lead the house prices during both the bubble and non-bubble period.

Including GSVI for Real Estate Agent in our Baseline error correction model for the house prices, improved all points of criteria. The adjusted coefficient of determination increases for both the short and long run and the speed of adjustment is higher and more significant. Substituting Google searches with the well-established Consumer Confidence Index yielded worse result for all assessments. The results are valid for the *real*, *minor* and non-bubble states. In addition to the thirty states not defined as either bubble nor non-bubble states. Based on the results found in this paper, we conclude that GSVI for Housing Bubble can be a strong housing bubble indicator while GSVI for Real Estate Agent can predict the housing trend and be included in price models to improve their predictive abilities at state levels.

The rest of the paper is organised as follows. First, we present our literature review before deriving a baseline model for house prices, which includes Google searches. Next, we present our data in section 4, followed by our empirical approach in section 5. Our results comes in section 6, discussion in section 7 and we present our conclusions in section 8.

2 Literature Review

2.1 Asset-pricing Bubble Theory

The term asset pricing bubbles is commonly used in economics, media and everyday speech. There are, however, issues related to the term. First is the very existence of asset-price bubbles. Fama (2014) rejects bubbles on empirical grounds by referring to the lack of reliable evidence that price declines are predictable. Others claim that bubbles certainly exist and that they are a psychological phenomenon (see, e.g. Shiller, 2005; 2010). Shiller (2005) offers the following normative definition of a bubble:

“A situation in which news of price increases spurs investor enthusiasm which spreads by psychological contagion from person to person, in the process amplifying stories that might justify the price increase and bringing in a larger and larger class of investors, who, despite doubts about the real value of the investment, are drawn to it partly through envy of others’ successes and partly through a gambler’s excitement.”

Furthermore, there are many different definitions of asset bubbles and Stiglitz (1990) offers perhaps the most famous normative description:

“A bubble exists if the price of an asset is high today only because investors believe it will be high tomorrow, and ”fundamental” factors do not seem to justify such increases”

In the case of houses, however, it is difficult to determine the fundamental value. Furthermore, several studies (see, e.g. Lind, 2009) argue that Stiglitz’ definition is inadequate, as it only refers to the price increase aspect of a housing bubble, and not the subsequent fall in prices. Lind (2009) offers a descriptive bubble definition, where:

“There is a bubble if the real price of an asset first increases dramatically over a period of several months or years and then almost immediately falls dramatically.”

For our purposes, the definition presented by Lind (2009) seems most fitting, as it relies only on the time series of housing prices.

2.2 The U.S. Housing Market, Bubble Indictors and Shiller’s List

According to Hardaway (2011), the 06/07 U.S. housing bubble is the greatest ever asset-pricing bubble. The collapse of the housing market and subsequent sub-prime mortgage crisis triggered one of the most significant economic downturns in history and affected virtually every corner of the world economy. In 2008, it had already triggered record wealth destruction on a global basis, because most banks and financial institutions in the U.S. and Europe held hundreds of billions of dollars’ worth of rotten subprime mortgage-backed securities. The economic downturn caused many other businesses in various industries to either go bankrupt or seek financial assistance.

The most important aspect of bubble indicators is their predictive abilities (Lind, 2009). This means that indicators must give strong indications that a dramatic increase in housing prices will be followed quickly by a dramatic decrease. Lind (2009) argues that bubbles cannot be explained by a single factor, but are the result of the interaction between different factors. Accordingly, a set of housing bubble indicators that combined provide strong indications of impending dramatic price decreases is required. Housing bubble indicators are of interest for academic purposes, for media, for policy makers such as banks, governments, investors, Insurance companies and asset managers, for homeowners and the population in general. A good set of bubble indicators can be used to both model risk and raise investor awareness of the risk associated with their positions, and help investors and asset managers rebalance portfolios to both achieve returns and avoid losses.

Shiller (2005; 2009) argues that asset-pricing bubbles are rooted to a great extent in human psychology. Due to their psychological nature, Shiller argues that asset pricing bubbles can be diagnosed with a checklist, similar to those used by psychologists to diagnose mental illnesses (Shiller, 2010). Shiller’s checklist points, published in the New York Times are:

1. Sharp increases in the price of an asset like real estate or shares
2. Great public excitement about said increases
3. An accompanying media frenzy
4. Stories of people earning much money, causing envy among people who are not
5. Growing interest in asset class among the general public
6. ”New era” theories to justify unprecedented price increases
7. A decline in lending standards

The issue with points 2, 3, 4, 5 and 6 on the list is that they are difficult to measure. Earlier attempts to measure the points include Case and Shiller (2004), who use newspaper articles related to housing to try to measure the extent of housing-related media frenzy.

2.3 Animal Spirits and Rational Exuberance

The asset pricing literature often distinguish between rational and irrational bubbles. Rational price bubbles exist when the price of an asset exceeds the asset’s fundamental value (see e.g. Engsted, 2014), as per Stiglitz’ definition. The bubble element in housing prices is driven by investor expectations. As the rational bubble is driven by investor expectations, investors are aware that a bubble exists. Investors can exploit the overpricing and expected future overpricing, ”riding the bubble”. It follows from the efficient markets theory that in efficient markets, bubbles and bubble indicators do not and cannot exist (Lind, 2009). Several empirical studies and well-renowned economists challenge, however, the assumption of rational decision making and the efficient markets theory (see, e.g. Jones, 2015; Kahneman, 2011).

Points 2, 3, 4, 5 and 6 in Shiller’s list presented in Section 2.2 are highly related to animal spirits, described by Akerlof and Shiller (2009). The authors argue that human psychology has a significant impact on economic decisions and the aggregate economy, and are highly critical of the assumption of rational decision making. In the first part of their book, Akerlof and Shiller treat five animal spirits. The spirits are confidence, fairness, corruption and bad faith, money illusion and stories. They argue for the importance of animal spirits by discussing, among eight key questions, why real estate markets go through cycles.

Akerlof and Shiller argue that all five animal spirits were clearly visible during the boom period of the 06/07 U.S. housing bubble. The authors argue that confidence is the most important animal spirit in determining behaviour and that it plays a major role in the business cycle. This is because of confidence, or the lack of it, has a big impact on people's purchase decisions. Put simply, when people are confident they spend more, which fuel the economy; lack of confidence makes people withdraw and sell, slowing down the economy. To illustrate the importance of confidence the authors describe the confidence multiplier, which is the same type of multiplier as the consumption, investment and government multipliers originally described by Keynes (1936). The confidence multiplier differs from the others because it cannot be measured as directly. Its properties and effects are nevertheless similar to those of the other multipliers. Each unit of money spent, because of increased confidence, will become income for other businesses and their employees, which they then spend. This will feed back into the economy as further increased confidence and income, round by round. The opposite happens when there is a negative change in confidence. Thus, the total effect of increased confidence is much greater than the initial effect and stimulus caused by it. This can be further explained by the feedback loop theory, as described by Shiller (2005) in Irrational Exuberance. The Feedback loop is tightly related to the confidence multiplier. In general, feedback loop theory argues that initial price increases lead to more price increases as the effects of the initial price increases feedback into yet higher prices through increased investor demand. Shiller (2005) describes a change in investor confidence, with increased confidence in the real estate market at the start of the U.S. housing boom. Confidence in real estate then grew during the housing boom, which fueled the dramatic increases in housing prices.

Shiller (2005) describes 12 different precipitating factors for the U.S. housing bubble. Of the 12 factors, the most relevant to our study is (arguably) the capitalist explosion and the ownership society. The ownership society refers to the increased desire to own rather than rent. According to Shiller (2005), owning homes became more and more important to people in the years before and during the housing boom. Thus, there was a general increase in demand for housing, which caused an initial price increase. This initial price increase was amplified through the feedback loop and fueled the dramatic housing price increase.

2.4 Measuring Interest with Google Trends Data

Research has shown that online behaviors can be used to reveal consumer`s intention and predict purchase outcomes (e.g., Kuruzovich et al. 2008). One of the earlier papers to use web search to forecast economic statistics was Ettredge et al. (2005), which examined the U.S. unemployment rate. In the recent years, Google Trends data have become increasingly popular in (financial) econometrics (see e.g. Bijl et al. 2015; Wu and Brynjolfson, 2009; 2015). Google Trend is one of several data sources that can be used to measure public's interest. Additional sources include other search engines (such as, e.g. Bling, Yahoo and Ask) and social media such as Facebook, Instagram and Twitter activity.

Using the internet as a research tool, consumers can find critical information to make purchase decisions (Horrigan 2008; Brynjolfson, Hu, and Rahman 2013). As the web becomes ubiquitous, more shoppers are using the Internet to gather product information and refine their purchasing choices, especially for products that require a high level of financial commitment, such as buying a home. According to the National Association of Realtors (NAR) (2016), a typical homeowner takes three months to buy but engage with agents earlier in the process and 83% of all home respondents frequently use the internet to search for their home. Thus, millions of persons are at all time searching through search engines, looking for a home, leaving behind economic intentions about their future economic behaviour. There have been several studies on whether Google searches can forecast financial markets.

Pentland (2010) found Google searches to precede purchase decisions and in many cases to be a more "honest signal" of actual interests and preferences because no bargaining, gaming, or strategic signalling is involved, in contrast to many market-based transactions or other types of data gathering such as surveys.

Wu and Brynjolfson (2009; 2015) find evidence that queries submitted to Google`s search engine are correlated with both the volume of housing sales as well as a house price index – specifically the Case-Shiller index – released by the Federal Housing Agency. They further found that search queries can reveal the current housing trend, but Google search is especially well suited for predicting the future unit sales of housing. Constructing a simple linear model, which includes Google searches, Wu and Brynjolfson (2009) predicted future housing sales and compared their results with NAR. They found their own prediction results to outperform the prediction released by NAR with 21.3 percent. These results have persisted over time (see e.g. Wu and Brynjolfson (2015)).

Choi and Varian (2009; 2011) uses Google Trends data to forecast near-term values of economic indicators such as automobile sales, unemployment claims, travel destinations planning, and consumer confidence. They have found that queries can be useful leading indicators for subsequent customer purchases in situations where consumers start planning purchases significantly in advance of their actual purchase decision. Further, they found that simple seasonal AR models that include relevant Google Trends variables tend to outperform models that exclude these predictors by 5% to 20%.

Preis et al. (2010; 2013) analyzed Google search queries for terms related to the financial market. The study found that the Google search volume reflected the current state of the stock market and that the search volume may predict future trends. Bijl et al. (2015) investigated the predictive power of Google search volume on stock returns. They found quarterly searches to be positively related to excess returns without reversal. They further examined a trading strategy and found that there is economic value in including Google search statistics in forecasting models.

Others are more skeptical to the use of web searches in predictions and have found less positive results. Goel et al. (2010) describe some of the limitations of web search data. They point of that, search data is easy to acquire and it is often helpful in making forecasts, but may not provide dramatic increases in predictability. Damien and Ahmed (2013) investigate previously results that Google search volume can predict future financial index returns but find that strategies based on financial related queries do not outperform strategies based on unrelated search terms.

3 Deriving a Baseline Housing Price Model

3.1 Deriving a Baseline Housing Price Model Based on Fundamental Factors

House prices are a result of the supply and demand for housing. The supply of houses are relatively stable in the short run, due to the time it takes to build new homes and that completion of new homes is low compared to the total housing stock. House prices will therefore mainly fluctuate due to short-term changes in demand. In the long run, housing stock will adapt to the demand. A long run model for house prices should, therefore, include explanatory variables for changes in housing value such as building and land costs and the cost of new homes.

The demand for housing consists of two components, the demand for residential purposes and the demand for housing as pure investment objects. It is reasonable to assume the former component to cover most of the housing demand, and we will, therefore, focus on this.

Households can consume housing services by either owning or renting a home. In deriving a baseline housing price model, we start with the following aggregated demand function derived by Jacobsen and Naug (2004):

$$H^D = f\left(\frac{C^O}{CPI}, \frac{C^O}{C^R}, Y, X\right), \qquad f < 0, \qquad (1)$$

Were

H^D = housing demand
C^O = total cost of living in a owned house
CPI = consumer price index
C^R = total cost of living in a rented home
Y = real disposable personal income
X = vector of non-fundamental factors affecting the housing demand
f_i = the derivative of $f(*)$ with respect to i

Equation (1) shows that demand for housing increases as the income increases and decreases if the cost of owning goes up compared to renting or compared to the consumer price index. The vector X contains factors that captures households expectation of future income and overheads. Expectations of future income and overheads are important, mainly due to three factors. First, housing is a lasting consumer good, secondly, home purchases are normally the greatest investment throughout lifetime and thirdly, most households finance a significant part of the purchase with a mortgage when buying their first home or advances in the housing market. The content of X will be further discussed later on.

The living cost of a homeowner, C^O, measures the benefits the owner must relinquish. The real living costs for a homeowner can be defined as:

$$\frac{C^O}{CPI} = \frac{P^H}{CPI} C^L = \frac{P^H}{CPI} [i(1-\tau) - E\pi - (E\pi^{PH} - E\pi)], \qquad (2)$$

Were
C^O = total cost of living in a owned house
CPI = consumer price index
C^L = living costs divided by amount invested in housing
P^H = the price of an average home
i = interest rate
τ = tax shield on capital income and expenditures
$E\pi$ = expected inflation
$E\pi^{P^H}$ = expected growth in P^H

The expression $[i(1-\tau) - E\pi]$ is the real interest rate after tax deductions and measures the real cost of interest rates from the mortgage and the potential real capital income, from interest rates, if private equity were placed in the bank instead of housing. Increased interest rates leads to increased mortgage costs and higher expected returns from placing money in the bank, leading to increased costs of living. The expression $\left(E\pi^{P^H} - E\pi\right)$ shows the expected real increase in house prices. The expected home equity goes up if $\left(E\pi^{P^H} - E\pi\right)$ increases, meaning that the real cost of owning decreases. This leads to benefits of owning compared to renting a home, increasing the demand for housing. Equation (2) can be simplified to:

$$\frac{C^O}{CPI} = \frac{P^H}{CPI} C^L = \frac{P^H}{CPI} \left[i(1-\tau) - E\pi^{P^H}\right], \qquad (2*)$$

The variable C^L is now the nominal interest rate after tax deduction and less the expected increase in nominal house prices. Equation (1) and (2) describes the expected demand for housing with regard to living purposes. The variables in (1) and (2) will affect the demand for housing as a pure investment object in addition to the demand from living purposes. It is reasonable to assume that this demand, as for others, increases with the income. If rent costs increases compared to housing prices, it becomes more attractive to invest in housing for rent. This leads to higher demand for housing. Correspondingly will lower interest rates/or higher $E\pi^{P^H}$ make it more beneficial to invest in property compared to placing private equity in bank. This pushes the demand curve for housing as an investment object upwards.

The housing supply is, as discussed above, relatively stable in the short run. The house price, P^H, ensures demand for housing equals the supply. We substitute (2) into (1) and resolve with regard to P^H. In addition, we use a logarithmic function:

$$lnP^H = \beta_1 lnCPI + (1-\beta_1)lnC^O + \ \beta_2 lnY + \beta_3 lnC^L + \beta_4\, ln\, H^V + \beta_5 \mathrm{g(X)}, \qquad (3)$$

Where

H^V = total housing value
C^O = total cost of living in a owned house
CPI = consumer price index
C^L = living costs divided by amount invested in housing
P^H = the price of an average home
Y = Disposable Personal Income
X = vector of non-fundamental factors affecting the housing demand
β_i = is the corresponding coefficient for the respective variable

Further, we define the disposable personal income by:

$$Y = \frac{YN}{CPI^{\alpha_1} C^{L^{\alpha_2}} P^{H^{\alpha_3}}}, \qquad \alpha_1 + \alpha_2 + \alpha_3 = 1, \quad \alpha_1 < \beta_1, \qquad \alpha_2 < \beta_2, \qquad (4)$$

Where YN = nominal disposable personal income

Equation (4) takes into account that higher housing prices lead to reduced purchasing power in the housing market. Solving (3) and (4) with regard to P^H yields:

$$lnP^H = \varphi_1 lnCPI + \varphi_2 lnC^O + \ \varphi_3 lnYN_t + \varphi_4 lnC^L{}_t + \varphi_5 \mathrm{ln}H^V + \varphi_6 \mathrm{g(X)} + \varepsilon_t, \qquad (5)$$

Were

$$\varphi_1 = (\beta_1 - \beta_2\alpha_1)/\gamma$$
$$\varphi_2 = (1 - \beta_1 - \beta_2\alpha_1)/\gamma$$
$$\varphi_3 = \beta_2/\gamma$$
$$\varphi_4 = \beta_3/\gamma$$
$$\varphi_5 = \beta_4/\gamma$$
$$\varphi_6 = \beta_5/\gamma$$
$$\gamma = (1 - \beta_2\alpha_3)$$

t is the period and ε_t is a stochastic residual, which captures effects of non-fundamental relationships. We see that $lnCPI$ and lnC^O disappear from (5) when $(\beta_1 - \beta_2\alpha_1) = (1 - \beta_1 - \beta_2\alpha_1) = 0$. This happens when the income elasticity, β_2, in (3) becomes greater than one.

The variable $lnC^L{}_t$ describes expected real price growth from period t to period $t+1$. This is an unobservable measure. We expect the future price expectations to depend on the observable, fundamental, right-hand sided variables in (5), the real price growth in period $t-1$ and a bubble factor B_t capturing psychological and other non-fundamental effects that can affect the price expectations. This leads to the following context:

$$\begin{aligned} lnP^H{}_t &= h^V(\text{fundamental factors})_t + \theta(real\ price\ growt)_{t-1} + B_t + \varepsilon_t \qquad (6) \\ &= h(\text{fundamental factors})_t + (deviation\ from\ fundamental\ value)_t \\ &= (\text{fundamental value})_t + (deviation\ from\ fundamental\ value)_t \end{aligned}$$

In equation (6), the house prices can deviate from their fundamental value when $\theta \neq 0$ or if the factors B_t and ε_t deviates from zero. If the deviation from fundamentals are both positive and significant, there may be an asset-pricing bubble in the real estate market. Such a bubble may develop from rising house prices as a result of change in fundamental conditions or a shift in price expectations $(B_t > 0)$. If $\theta > 0$, which can be reasonable, house price increases will fuel up under higher expectation higher prices. It will then become relatively more beneficial to own compared to renting. This leads to increased housing demand and prices. Due to this, expectations rises even more and house prices are pushed further up. This is closely related to the feedback loop theory as described by Shiller (2005). Such a process may lead to house prices deviates far from its fundamental values. However, it is reasonable to assume $\theta > 1$ so that the process dies over time. We will next discuss the fundamental factors of (6) before we derive an expression for the $(deviation\ from\ fundamental\ value)_t$ with respect to the bubble factor, B_t, in section 3.2.

In addition to changes in non-fundamentals, house prices can also fluctuate due to changes in fundamentals, e.g. such as changes in interest rates. Fluctuations may be reinforced by changes in supply. As discussed above, increased demand for housings will only lead to a short-term increase in house prices. The price increase leads to the construction of more housing.

β_i

This will over time, push down the prices, and the effect is amplified if demand has decreased at the time the new housings are finished. Household's expectations will also contribute to fluctuations in house prices. Lowered interest rates normally lead to expectations of (higher) price increases. Therefore, it becomes rational to accelerate planned housing purchases. This may lead to relative sharp housing price increases in the short-term and before falling back to an equilibrium in the long run.

We have argued that the demand for housing depend on households expectations for income. Due to expectations of future price increases also affects demand; households will emphasise the expected growth in income for other households as well. Developments in the job market are important for how households view their own and others economic future. Increased unemployment leads to expectations of lower wage increases and increased uncertainty of future income and solvency. The population also affects the demand for houses. The more the population increases, the more, the higher demand for housing. Increased population will only shift the house prices upwards if demand is not met by the supply of new accommodations. As earlier discussed, the increased demand will not affect prices in the long run but will have short-term effects. We will test the effect of both unemployment rate and population in our further analysis.

Based on the discussion above we include a one period lag of the dependent variable and the new, short-term, variables into equation (6). This yields the following baseline model for house prices explained by fundamental factors:

$$lnHPI_{s,t} = \alpha + \beta_1 lnHPI_{s,t-1} + \beta_2 lnDPI_t + \beta_3 lnHPA_{s,t} + \beta_4 \ln\left(1 + UR_{s,t}\right) \quad (7)$$
$$+ \beta_5 \ln(1 + IR_t) + \beta_6 lnPO_{s,t}$$

Where

$HPI_{s,t}$ = The House Price Index for state s, at time t
DPI_t = Disposable Personal Income at time t
$HPA_{s,t}$ = Housing Permits Authorized for state s, at time t
$UR_{s,t}$ = Unemployment Rate for state s, at time t
IR_t = Interest Rate at time t
$PO_{s,t}$ = Population in state s, at time t
β_i = Is the corresponding coefficient for the respective variable

3.2 Including Non-fundamentals in the Baseline Model Using Google Search

The house prices naturally fluctuate above and below the long-run equilibrium. When prices are above equilibrium, we call it overpricing and underpricing when the prices are below the long-run trend. Occasionally, with long time intervals in between, natural overpricing lead to a housing bubble, where the house prices deviates too much from the long run equilibrium trend to be explained by changes in short-term factors such as interest rate, unemployment rate, population, etc. DiPasquale and Wheaton (1994) say that:

"Indeed, it appears to be normal for housing prices to deviate from the fundamental value or equilibrium price since housing markets clear gradually rather than quickly in a short run."

To model the non-fundamental factors in the house prices we start with an expression of the bubble factor derived in the section above:

$$HPI_{s,t} - F_{s,t} = B_{s,t} \tag{8}$$

Where $F_{s,t}$ is the fundamental value of the Housing Price Index in state s, at time t , and $B_{s,t}$ is a possible bubble element in the $HPI_{s,t}$ in state s, at time t. The fundamental value $F_{s,t}$ can be divided into the following:

$$F_{s,t} = HPI"_{s,t} + S_{s,t} \tag{9}$$

Where $HPI"_{s,t}$ is the house price if it had followed the long-term fundamental value and $S_{s,t}$ is the cycle element of the house price. An eventual overpricing or underpricing of the house prices can be modeled:

$$HPI_{s,t} - HPI"_{s,t} = B_{s,t} + S_{s,t} \tag{10}$$

We believe that Google Trends can be used to measure animal spirit as described by Akerlof and Shiller (2009) and rational exuberance, as described by Shiller (2005), among the population at the state level. By including Google Search Volume Index ($GSVI_{w,s,t}$) for search terms related to housing bubbles and the housing market in general, we hope to explain the bubble factor, $B_{s,t}$, and thereby improve the model. In addition to being indexed by s and t, GSVI is also indexed by w for the different search terms. Including Google searches in equation (10) yields:

$$HPI_{s,t} = \alpha + \beta_1 HPI_{s,t-1} + \beta_2 DPI_t + \beta_3 HPA_{s,t} + \beta_4 UR_{s,t} + \beta_5 IR_t + \beta_6 PO_{s,t} + +\beta_7 GSVI_{w,s,t} \tag{11}$$

Where

$HPI_{s,t}$ = The House Price Index for state s, at time t
DPI_t = Disposable Personal Income at time t
$HPA_{s,t}$ = Housing Permits Authorized for state s, at time t
$UR_{s,t}$ = Unemployment Rate for state s, at time t
IR_t = Interest Rate at time t
$PO_{s,t}$ = Population in state s, at time t
β_i = Is the corresponding coefficient for the respective variable
$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t

4 Data

All data used in this paper, except for Google searches, are downloaded directly from NTNU Handelshøyskolen database DataStream. The data, as relevant to, are adjusted for seasonality effects using the Centered Moving Average (CMA) method as described in Appendix B and adjusted for inflation using the consumer price index (CPI) obtained from the Federal Reserve Bank of St. Louis.

Google Search Volume Index data are only available from Q1 2004, and the Housing Price Index data were only available for Q3 2016 when we started. Therefore, all data used in this master thesis are running from Q1 2004 until Q3 2016. The Housing Price Index at the state level is published quarterly, and we have therefore downloaded and converted the rest of the data into quarterly time-series. Last, all data are transformed into logarithmic form after the other adjustments.

4.1 House Prices

We use the quarterly, all-transactions Housing Price Index (henceforth referred to as HPI) published by the Federal Housing Finance Agency (FHFA) as a housing market indicator. The all-transactions HPI is a broad measure of the development of house prices for each geographic area (i.e. state or district). The prices are estimated using repeated observations of housing values for individual single-family residential properties on which at least two mortgages were originated and subsequently purchased by either Freddie Mac or Fannie Mae.

4.2 Real Disposable Personal Income

We have downloaded quarterly Real Disposable Personal Income, adjusted for both inflation and seasonal effects, from FRED, Federal Reserve Bank of St. Louis. Disposable Personal Income is the amount of money that households have available for spending and saving after income tax deduction. The data are not specific at the state level. We first tried Real Personal Income, downloaded from the same place and treated for the same effects, which is at the state level, but it led to collinearity and autocorrelation in the baseline model. Substituting Personal Income with Disposable Personal Income removed the collinearity and reduce the autocorrelation in the error term.

4.3 Housing Permits authorised

The housing permits authorised (HPA) is a proxy for the change in the housing stock, by signalling the number of new homes going to be built. The data are downloaded from U.S. Bureau of the Census, New Private Housing Units authorised by Building Permits, retrieved from FRED, Federal Reserve Bank of St. Louis. All numbers are in 1000 before transforming it into logarithm form. We have converted the data from monthly to quarterly by taking the average of three and three months.

4.4 Unemployment Rate

We have downloaded monthly, seasonally adjusted, unemployment rate data from United States Department of Labor, Bureau of Unemployment Statistics. The data are at the state level, and we have converted them into quarterly data by taking the average of three and three months. Before transforming the Unemployment Rate into logarithmic form, we included the value of one to avoid negative numbers.

4.5 Interest Rate

As variable for the interest rate, we use the US yield 10-years Treasury Note. The data are downloaded from the 10-year Treasury, which has become the security most frequently quoted when discussing the performance of the U.S. government bond market and is used to convey the market's take on longer-term macroeconomic expectations. Since the 1970s, the 10 Year Treasury Note and the 30 years fixed mortgage have had a very tight correlation. Before transforming the Interest Rate into logarithmic form, we included the value of one to avoid negative numbers.

4.6 Population

Data of the population are downloaded from the United States Census Bureau. The data at state level in the U.S. are only published yearly, and to get the data at state level, we have used cubic-spline interpolation to convert it into quarterly data. Due to autocorrelation in the baseline, error correction, model for house prices, when including population, we made a dummy. The dummy variable equals one if the population growth is significantly higher (1.5 times higher) than average and zero otherwise. This reduced the autocorrelation.

4.7 Google Search Volume Index

Google Trends data can be used to measure relative interest for a search term. A Google Search Volume Index (henceforth referred to as GSVI) level of 100 implies that this is the point in time the total searches for a term made up the biggest proportion of all Google Searches. Thus it reflects the point in time when the relative interest in a search term was highest. All other GSVI values are relative to the maximum. High values indicate that interest for the search term is high, while low values indicate low interest in the search term. An important aspect of the construction of the GSVI is that the total number of searches at some point in time must be above a threshold set by Google for the GSVI to be published. We have not been able to find the exact threshold. Nevertheless we find it reasonable to interpret GSVI=0 as very low interest in the search term if the specific state has a relatively high population and disregard the result if the population, in the specific state, is relatively low. The data set we use covers twenty different search terms, in addition to four self-created indexes, for the fifty United States

4.8 Consumer Confidence Index

The Conference Board Consumer Confidence Index (CCI) is a barometer of the health of the U.S. economy from the perspective of the consumer. The Index is based on consumers' perceptions of current business and employment conditions, as well as their expectations for six months hence regarding business conditions, employment, and income. The Consumer Confidence Index and its related series are among the earliest sets of economic indicators available each month and are closely watched as leading indicators for the U.S. economy. It was started in 1967 as a mail survey conducted every two months. Since 1977, the Consumer Confidence Index have been published monthly and the concept, definitions and questions have stayed consistent. The CCI is indexed for the calendar year of 1985 and is then used as a benchmark. The data are Seasonally adjusted with the U.S. Census X-12 seasonal adjustment. We have converted the monthly data into quarterly by taking the average of three and three months and then taking the natural logarithm of it. The data are only available at the Country level and not for each specific state.

4.9 List of all variables

#	Variable Name	Abbreviations	Available at	Data are adjusted for
1	Housing Price Index	$HPI_{s,t}$	State Level	Seasonality & Inflation
2	Disposable Personal Income	DPI_t	Country	Seasonality & Inflation
3	Housing Permits Authorised	$HPA_{s,t}$	State Level	Seasonality effects
4	Unemployment Rate	$UR_{s,t}$	State Level	Seasonality effects
5	Interest Rate	IR	Country	
6	Population	$PO_{s,t}$	State Level	Dummy of Population
7	Google Search Volume Index	$GSVI_{w,s,t}$	State Level	Seasonality effects
8	Consumer Confidence Index	CCI_t	Country	Seasonality effects

Table 4-1: The table display the eight variables, which are used in the different error correction models (ECM) throughout this paper.

5 Empirical Approach

5.1 Bubble Identification and Ranking

We first use Harding & Pagan's (2002) algorithm to identify housing price peaks and troughs in the different states, with q=j=6 (Bracke, 2013). We then use the peak with the highest value and corresponding date (quarter/year) in our calculations and find the housing price three and five years before the peak to calculate the changes. Then we find the trough with the lowest housing price value after the peak and use this in the calculation of price fall, as per the bubble definition. We identify bubble states and rank all states by the total price decrease. As we want to compare bubble states to non-bubble states, we include the same number of non-bubble states as identified bubble states as benchmark states. The non-bubble states selected are the ones that experienced the smallest price decrease, if any.

- After sorting all the fifty states according to their total price fall and looking at their previous price increase, we found four states standing out from the rest. Nevada, Arizona, Florida and California. See Appendix F.

- To compare the effects in the states that experienced a real housing bubble with those that experienced a large correction, we choose the following six states according to their total price fall and the ten states that experienced the least correction in house prices during the housing bubble in 06/07. See Appendix C to view the list of all fifty states sorted after their total price fall from peak to trough.

5.2 Selection of Search Terms

The first step in testing GSVI as a housing bubble indicator is to identify potential search terms. Our belief is that potential indicator search terms are related to both rational and irrational bubbles and irrational exuberance, as presented in Section 2.3 and real estate in general. We want to identify search terms it would be natural for American investors and potential home buyers to search for, and thus try to put ourselves in the shoes of investors, potential home buyers and people living in prosperous economic times. Search terms related to rational and irrational bubbles are search terms we believe it would be natural for investors to search for, both for gaining general information about possible investments and specific information about a possible housing bubble. Terms related to irrational bubbles are search terms we believe it would be natural to search for in times great interest and confidence in the housing market and economic confidence in general. For people actively looking to buy a home search terms directly related to the housing market would be natural to "Google". Using this approach, we identify 204 search terms. See Appendix B for the full list of search terms. Testing the correlation among each of the 204 search terms and the Housing Price index for the identified bubble states found from 5.1, we reduce the number of search terms by removing those with low correlation in the bubble period. After screening the 204 search terms, we end up with 20 different queries related to the housing market and housing bubbles reflecting the characteristics we look for. The twenty search terms are presented in Table 5-1.

Google Search Queries Related to Housing Bubbles and the Housing Market			
Apartment	Home	Lending	Real Estate Bubble
Broker	Home Equity	Mortgage	Real Estate Investment
Bubble	Housing Bubble	Real Estate	Real Estate Listings
Construction	Housing Market	Real Estate Agent	Realtor
Flat	Investment	Real Estate Broker	Rent

Table 5-1: The table presents the search terms that passed our initial inclusion criterion. These are queries displaying a relatively high correlation with the house prices in the identified bubble states and we believe the interest for them will increase in times of great economic confidence.

5.3 Testing GSVI as Housing Bubble Indicator

After the search term identification presented in 5.2 the next step in testing the GSVI for the different search queries as housing bubble indicators is to assess which search terms have search volume level developments that match Shiller's checklist points. To do this, we propose a red flag test based on differences in search volume levels during the housing bubble period compared to the time after. The Red Flag test results determine which search terms to include in the index. The period for the housing bubble are defined as follows:

- BP = Q1.2004 until Q4.2008

Because Google Trends data are only available from 2004, we cannot measure differences in interests and animal spirits from before the housing boom started. We can only assume that some period after the housing bubble is representative of a non-bubble period. We use the following period as a proxy for a non-bubble period:

- NBP = Q1.2009 until Q3.2016.

We find it reasonable to define NBP as a non-bubble period for several reasons. By setting the start of the normal period quite long after the housing bubble burst, we should avoid potential noise in the data. Secondly, a study conducted by Chen et al. (2012) indicates that the crisis was easing in 2009. Furthermore, by observing the Case-Shiller Home Price Index, we find that the housing prices, in general, started to level out after Q4 2008. Therefore, we find it reasonable to assume that NBP reflects a non-bubble period where interests are at normal levels.

5.3.1 Google Search Volume Index Performance Tests for Specific Search Terms

To assess the twenty search terms in-sample predictive abilities we propose some Red Flag tests. The tests use the average of $GSVI_{w,s}$ in the non-bubble period as benchmark. If the $GSVI_{w,s}$ is above M times the average level for the non-bubble period, it is flagged. The $GSVI_w$ should ideally flag a bubble in all bubble states, and zero of the non-bubble states. We list test names with short descriptions below. Figure 5-1 illustrates the general principle of the tests.

Test Name	Test Description
One in a row	Checks if $GSVI_{w,s,t}$ is M times higher than normal average in at least one quarter
Two in a row	Checks if $GSVI_{w,s,t}$ is M times higher than normal average in two consecutive quarters
Three in a row	Checks if $GSVI_{w,s,t}$is M times higher than normal average in three consecutive quarters
Eight in a row	Checks if $GSVI_{w,s,t}$is M times higher than normal average in eight consecutive quarters

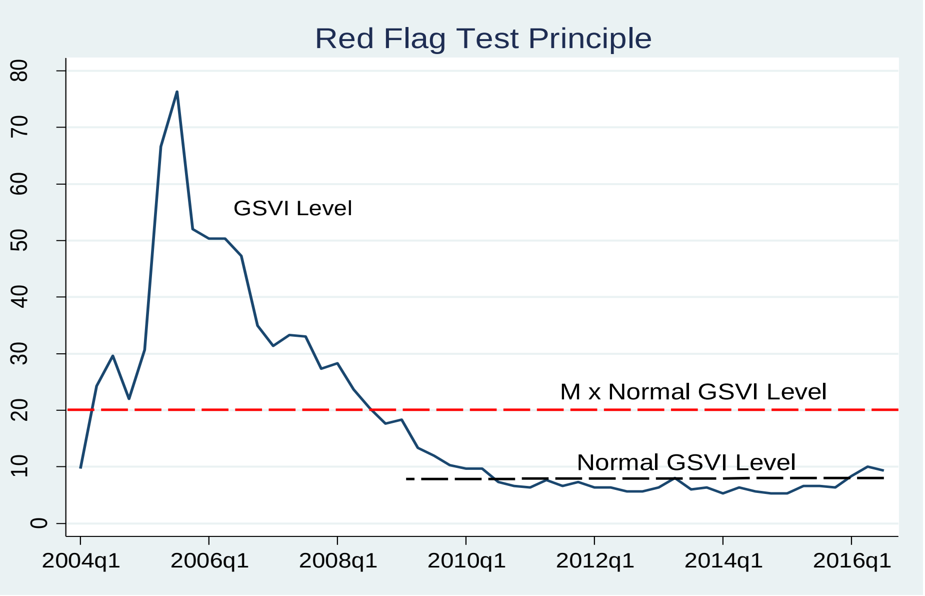

Figure 5-1: The figure illustrates the test principle. The vertical axis represents the value of the Google Search Volume Index (GSVI), with time on the horizontal axis. The black line represents the average value of the GSVIw,s during the normal period, which is defined to run from Q1 2009 to Q3 2016. The red line represents M times the average level during the normal period.

We test with multiples $M = [1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.5, 5, 7, 10]$

The ”1 in a row test” flags a state as a bubble state if GSVIw,s,t is M times higher than normal in at least one quarter during the bubble period. The ”2 in a row test” flags a state as a bubble state if GSVIw,s,t is M times higher than normal for at least two consecutive quarters.

The ”3 in a row test” flags a state as a bubble state if GSVIw,s,t is M times higher than normal for at least three consecutive quarters. The ”8 in a row test” flags a state as a bubble state if GSVIw,s,t is M times higher than normal for at least eight consecutive quarters. The purpose of making the tests stricter, either by increasing M or the required number of subsequent periods with high GSVIw,s,t levels, is primarily to avoid flagging the non-bubble states as bubbles, thereby improving the GSVIw,s,t performance.

5.3.2 GSVI Performance Tests of Indexes of the Best Performing Search Terms

To try improving the in-sample bubble identification of single search terms, we construct indexes of the average GSVI. By including the average GSVI of several search terms, we hope to construct an index that is both more robust and captures a wider part of the interest. From the bubble identification result in the upper part of Table 6-1, we now construct four different housing bubble indexes:

- Average GSVI of all twenty search terms, henceforth Index20
- Average GSVI of the twelve best performing search terms , henceforth Index12
- Average GSVI of the six best performing search terms, henceforth Index6
- Average GSVI of the three best performing search, henceforth Index3

The four housing bubble indexes are now tested the same way as the individual search terms were tested in section 4.2.1 to see if the bubble identification results could be improved. The results are displayed in the lower part of Table 6-1 in the result section.

5.4 Test Performance

The performance of the different GSVI for the specific search terms and indexes can be measured by the number of errors. We have two types of errors:

- Type I-error: $GSVI_{w,s}$does not flag bubble state as bubble
- Type II-error: $GSVI_{w,s}$flags non-bubble state as bubble

Type I-errors have a ”sub-error”, which is that the GSVIw,s does not flag a real bubble. If the GSVIw,s is not able to detect a real bubble state this is more problematic than if the GSVIw,s does not detect a minor bubble. Based on this we make a point system. Three points are given for detecting a real bubble state, one point is given for detecting a minor bubble state, and three points are deducted for wrongly detecting a non-bubble state as a bubble state. We conduct four different tests and rank the search terms according to their total points given. The results are shown in Table 6-1 in the results section.

Good GSVIw predictive ability is valuable for investors, both institutional and private, and in asset management. Investors may incur great losses if they do not liquidise long positions before the bubble bursts. This is particularly important in portfolios with large proportions of total wealth in real estate, as is the case for many households. In addition to effects through direct contractual links in real estate come effects on positions in real-estate related securities and financial securities in general. The 2006/07 U.S. housing bubble illustrates the potential impact of housing bubbles on financial markets and the economy as a whole. It is clear that the importance of good GSVIw predictive abilities extends far beyond the real estate market. Alternatively, the GSVIw can help investors justify short positions. Good housing bubble indicators should, therefore, be of great interest for investors, both for managing positions in real estate and positions in other financial securities and also whether to go long or short.
Conversely, type II-errors can also cause great problems for investors and asset managers. Erroneous bubble indication may lead to unnecessary and inefficient portfolio rebalancing, which leads to costs in the form of unrealized returns. Type II-errors are very problematic from a policy view. Taylor (2015) describes fears during 2006/07 housing bubble that an intervention may have even greater negative consequences than the burst of the bubble. This underlines the importance of good indicator predictive abilities, and that type II-errors should be avoided.

5.5 Testing for Short and Long Run Effects from GSVI for the Best performing Search Terms on the House Prices Globally in the U.S.

Based on the in-sample prediction result in Table 6-1, we further analyse the causality between the two best performing GSVI and index, namely Housing Bubble, Real Estate Agent and Index12, and the Housing Price Index. It is noteworthy that the two search terms are very different from each other. Housing Bubble is directly associated with real estate bubbles and rational bubbles, while Real Estate Agent is a more common term related to housing in general.

To further analyse which of the three are most suited, we find the correlation between GSVI for Housing bubble and HPI, and the correlation between Real Estate Agent and HPI, and the correlation between Index12 and HPI for the whole period, the bubble period and the non-bubble period. The correlation results are shown in Table 6-3 in the results section. We also analyse when the two search terms and Index12, peaked and troughed compared to the house prices in the real, minor and non-bubble states. The results are presented in Table 6-2 in the results section.

We want to test which of the two search terms and Index12 can best explain the house prices in the short and long run globally in the United States. First, we test GSVI for Housing Bubble, Real Estate Agent and Index12, and HPI, in level, using the Dickey-Fuller Generalized Least Square method with one lag. It has been standard procedure to use the augmented Dickey-Fuller (1979) and Phillips-Perron tests to determine whether a series possesses a unit root, but today there exist tests with better statistical properties as shown by Elliott, Rothenberg and Stock (1996). In this paper, we will be using the Dickey-Fuller Generalized Least Square to test for unit roots in the time series. We find that the four time-series are non-stationary for the United States. Next, we transform the variables into first difference and perform the same test again. The results show that all four variables are now stationary at a one percent significance level. See Appendix D for the full test results.

After determining the variables are integrated of the first order, we test for cointegration among the variables using the Johansen test method. We find there exist one or more cointegrating relationship among them. According to Wooldridge (A Modern Approach, 2012), when two variables y_t and x_t are both $I(1)$ and cointegrated, we can first run a linear regression of the HPI with the variables in levels and interpret the results as long-run effects.

Thereafter we run the regression on the first differenced variables, including the error term from the previously model, creating an Error Correction Model (ECM). Now we can interpret the results from the ECM as short run effects and the coefficient of the error term, also called the error correction term, as the speed of adjustment.

Combining the use of OLS regression on variables at levels with the ECM to test for both short and long run relationship between HPI and GSVI, compared to e.g. vector error correction models (VECM) which are common in housing related literature, have several advantages. First is the interpretation of the results. The results from this method are easier to interpret, especially when having a model with several variables with more than one cointegrating relationship. This would have become increasingly problematic later on when testing for short and long run causalities in the three baseline models, for each of the 50 states, which includes seven variables. Secondly, VEC models demand the same amount of lags on all variables. This is not suitable when only testing the effect from GSVI with different lags on house prices. Note: We develop a VEC model, which we tested using both GSVI for Real Estate Agent and Index12, separately and together, at state level. The full result from this model is not included in this paper but we will briefly discuss our findings in the result section.
The general regression model used to model the long-run effect from GSVI for Housing Bubble and Real Estate Agent on the Housing Price Index are shown in (12). $\beta_i = 0$ for the variables not included in the specific test.

$$HPI_t = \alpha + \beta_1 HPI_{t-1} + \beta_2 GSVI_{HB,t} + \beta_3 GSVI_{HB,t-2} + \beta_4 GSVI_{REA,t} + \beta_5 GSVI_{REA,t-2} + \beta_6 GSVI_{Index12,t} + \beta_7 GSVI_{Index12,t-2} \tag{12}$$

Where
$HPI_{s,t}$ = The House Price Index for state s, at time t
$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t

The general regression model used to find the short run effect from GSVI for Housing Bubble, Real Estate Agent and Index, on the Housing Price Index and the speed of adjustment are shown in (13). $\beta_i = 0$ for the variables not included in the specific test.

$$\Delta HPT_t = \alpha + \beta_1 \Delta HPI_{t-1} + \beta_2 \Delta GSVI_{HB,t} + \beta_3 \Delta GSVI_{HB,t-2} + \beta_4 \Delta GSVI_{REA,t} \qquad (13)$$
$$+ \beta_5 \Delta GSVI_{REA,t-2} + \beta_6 \Delta GSVI_{Index12,t} + \beta_7 \Delta GSVI_{Index12,t-2} + \gamma \epsilon_{HPI,t-1}$$

Where

$HPI_{s,t}$ $=$ The House Price Index for state s, at time t
$GSVI_{w,s,t}$ $=$ Google Search Volume Index for search term w, in state s, at time t
$\epsilon_{HPI,t-1}$ = The error correction term

We start regressing the house prices using only GSVI for Housing Bubble, next we only use GSVI for Real Estate Agent and last we use Index12. Regressing the house prices with only one variable gives a good indication of both its short and long run effects. In addition to how much it alone can explain the house prices. Next, we regress the house prices using GSVI for Housing Bubble and different lags of it, then GSVI for Real Estate Agent with different lags before we do the same for Index12.

By including several lags of the independent variable, we want to find whether this improves the model's in-sample prediction results. After testing GSVI for the two search terms and Index12 independently, we include both of the search terms to find whether it can further improve the result and if so, by how much. This will give indications of whether the two search terms captures different information and thereby improves the in-sample prediction results. Finally, we include a one period lag of the house prices in the different regression models. We expect this to improve the model, in both the short and long run. By including a one period lag of the dependent variable, we want to find how the explanatory power of the Google searches change and whether the results are coinciding with which search terms/Index gave the best results alone. See Appendix J to view all the specific models used to regress the house prices globally in the United States.

5.6 Testing for Short and Long Run Effects from GSVI for Real Estate Agent to the House Prices for all 50 States

Finding GSVI for Real Estate Agent best explaining the house prices globally in the U.S. in the previous section, we now want to test its explanatory power on the house price in each of the 50 states. As before, we start by testing GSVI for Real Estate Agent and HPI, in level, using the Dickey-Fuller Generalized Least Square method with one lag. We find that the two time-series are non-stationary for all 50 states. Next, we transform the variables into first difference and perform the same test again. The results show that the Housing Price Index is stationary in 49 out of 50 states and that GSVI for Real Estate Agent is stationary in 45 out of 50 states. In most states, the time series are stationary at a one percent significance level. See Appendix D for the full test results.

Next, we test for cointegration, again using the Johansen test, between the HPI and GSVI for Real Estate Agent and find that the two time-series are cointegrated in 45 out of 50 states. See Appendix E for the full test results. Due to the existence of cointegration, we first run a linear regression of the HPI with the variables in levels and interpret the results as long-run effects. Next, we run the regression on the first differenced variables, including the error term from the previously model, creating an Error Correction Model (ECM). Now we interpret the results from the ECM as short-run effects and the coefficient of the error term as the speed of adjustment.

The general regression model used to model the long-run effect from GSVI for Real Estate Agent on the Housing Price Index are shown in (14). $\beta_i = 0$ for the variables not included in the specific test.

$$\bar{H}\bar{P}\bar{I}_{s,t} = \alpha + \beta_1 HPI_{s,t-1} + \beta_2 GSVI_{REA,s,t} + \beta_3 GSVI_{REA,s,t-1} + \beta_4 GSVI_{REA,s,t-2} \quad (14)$$

Where

$HPI_{s,t}$ = The House Price Index for state s, at time t

$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t

The general error correction model used to model the short run effect from GSVI for Real Estate Agent on the Housing Price Index, and the speed of adjustment are shown in (15). $\beta_i = 0$ for the variables not included in the specific test.

$$\Delta \widetilde{HPI}_{s,t} = \alpha + \beta_1 \Delta HPI_{s,t-1} + \beta_2 \Delta GSVI_{REA,s,t} + \beta_3 \Delta GSVI_{REA,s,t-1} + \beta_4 * \Delta GSVI_{REA,s,t-2} + \gamma \epsilon_{HPI,s,t-1} \quad (15)$$

Where

$HPI_{s,t}$ = The House Price Index for state s, at time t

$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t

$\epsilon_{HPI,t-1}$ = The error correction term

We performs the same regressions as in the previously section except we do on state level for all 50 states instead of the country as a whole and we only use GSVI for Real Estate Agent. Regressing the house prices on the state level will show how Google search performs in the states that experienced a bubble compared to those who did not. When moving from country to state level the total amount of Google searches will be lower and we assume the quality of the data reduced. Thus, we expect GSVI to have higher explanatory power on the house prices in states with a large population compared to states with a low population. We start regressing the house prices using only GSVI for Real Estate Agent. Next, we try adding different lags of GSVI for Real Estate Agent, finding that more than two lags seldom improve the model. Last, we regress the house prices using a one period lag of the house prices and GSVI for Real Estate Agent without any lags. Due to the inclusion of one period lag of the dependent variable, we expect the last model to have better in-sample predictive abilities. We want to find how this simple model performs compared to the baseline models, developed in the literature section, and therefore, calculates the mean absolute error (MAE) for both $\widetilde{HPI}_{s,t}$ and $\Delta \widetilde{HPI}_{s,t}$ using equation (16) and (17).

$$MAE_{\bar{H}\bar{P}\bar{I}_{s,t}} = \frac{1}{\mathrm{N}}\sum_{t=1}^{N}\left|\frac{HPI_{s,t} - \bar{H}\bar{P}\bar{I}_{s,t}}{HPI_{s,t}}\right| \tag{16}$$

and

$$MAE_{\overline{\Delta H}\bar{P}\bar{I}_{s,t}} = \frac{1}{\mathrm{N}}\sum_{t=1}^{N}\left|\frac{\Delta HPI_{s,t} - \overline{\Delta H}\bar{P}\bar{I}_{s,t}}{\Delta HPI_{s,t}}\right| \tag{17}$$

Where

$HPI_{s,t}$ = The House Price Index for state s, at time t

$\bar{H}\bar{P}\bar{I}_{s,t}$ = The predicted values of the House Price Index for state s, at time t

The results are shown together with the results from the baseline models in Table 6-7 in the results section. See Appendix J to view all the specific models used to regress the house prices for each of the 50 United States.

5.7 Testing Whether GSVI for Real Estate Agent Improves the Baseline Housing Price Model

Finding the short and long run dynamics between GSVI for Real Estate Agent and HPI in the previous section, we now want to find whether Google searches can improve the baseline model. First, we test the time-series for all 50 states of the remaining variables, in the baseline model, for stationarity. We find the time series to be non-stationary at level but stationary after transforming them into first differenced. See Appendix D for the full results from the stationarity tests. Next, we test for cointegration among the variables, using the Johansen test method, and find that there exist one or more cointegrating relationship in all 50 states with a 5% significance level. See Appendix E for the full test results.

Due to the existence of cointegration, we first run a linear regression of the HPI with the variables in levels and interpret the results as long-run effects. Next, we run the regression on the first differenced variables, including the error term from the previously model, creating an Error Correction Model (ECM). Now we interpret the results from the ECM as short-run effects and the coefficient of the error term as the speed of adjustment.

The general regression model used to model the long-run effect of the independent variables on the Housing Price Index are shown in (18). $\beta_i = 0$ for the variables not included in the specific test.

$$HPI_{s,t} = \alpha + \beta_1 HPI_{s,t-1} + \beta_2 UR_{s,t} + \beta_3 PO_{s,t} + \beta_4 DPI_t + \beta_5 IR_t \\ + \beta_6 HPA_{s,t} + \beta_7 GSVI_{REA,s,t} + \beta_8 CCI_t \quad (18)$$

Where

$HPI_{s,t}$ = The House Price Index for state s, at time t
DPI_t = Disposable Personal Income at time t
$HPA_{s,t}$ = Housing Permits Authorized for state s, at time t
$UR_{s,t}$ = Unemployment Rate for state s, at time t
IR_t = Interest Rate at time t
$PO_{s,t}$ = Population in state s, at time t
β_i = Is the corresponding coefficient for the respective variable
$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t
CCI_t = The Consumer Confidence Index at time t

The general error correction model used to model the short run effect of the independent variables on the Housing Price Index and the speed of adjustment are shown in (19). $\beta_i = 0$ for the variables not included in the specific test.

$$\Delta\overline{HPI}_{s,t} = \alpha + \beta_1\Delta HPI_{s,t-1} + \beta_2\Delta UR_{s,t} + \beta_3\Delta PO_{s,t} + \beta_4\Delta DPI_t + \beta_5\Delta IR_t + \beta_6\Delta HPA_{s,t} + \beta_7\Delta GSVI_{REA,s,t} + \beta_8\Delta CCI_t + \gamma\epsilon_{HPI,s,t-1} \quad (19)$$

Where

$HPI_{s,t}$ = The House Price Index for state s, at time t
DPI_t = Disposable Personal Income at time t
$HPA_{s,t}$ = Housing Permits Authorized for state s, at time t
$UR_{s,t}$ = Unemployment Rate for state s, at time t
IR_t = Interest Rate at time t
$PO_{s,t}$ = Population in state s, at time t
β_i = Is the corresponding coefficient for the respective variable
$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t
CCI_t = The Consumer Confidence Index
$\epsilon_{HPI,t-1}$ = The error correction term

First, we regress the house prices without including GSVI nor the Consumer Confidence Index (CCI), setting β_7 and β_8 equal to zero. Thus, finding how the baseline, error correction, model performs in both the short and long run in all 50 states. Then, we calculate the MAE of the in-sample prediction error of both $\overline{HPI}_{s,t}$ and $\Delta\overline{HPI}_{s,t}$ using equation (16) and (17) from section 5.6. Next, we include GSVI for Real Estate Agent by removing the requirement of β_7 being equal to zero, to test whether Google searches improve the baseline model. Last, we substitute the GSVI with CCI, setting $\beta_7 = 0$ again and removing the requirement of β_8 being equal to zero. Including CCI instead of GSVI in the baseline model allows us test how well GSVI performs compared to a well-established indicator of consumer confidence. See Appendix J to view the three specific baseline models used to regress the house prices for each of the 50 states.

6 Results

6.1 Results from the In-sample Bubble Identification Tests

In the table below, we present the ranking and result of the twenty single search terms and four self -created indexes based on their in-sample predictive ability to identify bubble states.

Rank	**Search Term**	**1 in a row**	**2 in a row**	**3 in a row**	**8 in a row**	**Total Result**
1	Housing Bubble	16	16	16	16	64
2	Real Estate Agent	14	15	16	14	61
3	Real Estate	14	13	15	13	57
4	Housing Market	13	12	12	14	51
5	Realtors	10	10	13	17	50
6	Real Estate Listings	13	13	13	9	48
7	Mortgage	11	11	11	13	46
8	Investment	8	7	11	8	34
9	Real Estate Broker	9	9	9	6	33
10	Real Estate Bubble	8	8	8	8	32
11	Broker	4	5	14	8	31
12	Home equity	3	3	10	8	24
13	Lending	5	6	7	4	22
14	Real Estate Investment	3	0	3	7	13
15	Property	6	3	1	0	10
16	Apartment	2	0	1	1	4
17	Construction	0	0	0	3	3
18	Bubble	1	0	0	0	1
19	Rent	1	0	0	0	1
20	Flat	0	0	0	0	0
Rank	**Average GSVI of the**	**1 in a row**	**2 in a row**	**3 in a row**	**8 in a row**	**Total Result**
1	12 Best Performing ST	15	15	15	12	57
2	6 Best Performing ST	15	13	13	14	55
3	20 Best Performing ST	15	12	12	13	52
4	3 Best Performing ST	12	12	12	11	47

Table 6-1: The table shows the results of the four flag tests, "1, 2, 3 and 8 in a row", and the total result for each of the twenty search terms in addition to four self-created indexes. The search terms/indexes are given 3 points for correctly indicating a real bubble state, 1 point for correctly indicating a minor bubble state and 3 points are deducted for wrongly indicating a non-bubble state as a bubble state. Total results are the sum from the four tests. "# in a row" flags a state as a bubble state if GSVI for the specific search query is above a constant M times the GSVI level during the non-bubble period for # consecutive quarters, where $\# = \{1,2,3 \text{ and } 8\}$.

Table 6-1 shows the ranking and score from four different, in-sample prediction, tests based on identifying the states that experienced a bubble for the twenty single search terms and the four self-created indexes. To rank the different search terms and indexes we created a point system where each query is given three points for correctly identifying a *real* bubble state, one point for correctly identifying a *minor* bubble state and three points are deducted for erroneously identifying a non-bubble state. The maximum number of points a query may receive in each of the four tests are; three points for each of the four bubble states, one point for each of the six minor bubble states, equaling a maximum of eighteen points. We illustrate this through an example, e.g. Housing Bubble has received sixteen points in all four tests for correctly including all four real bubble states, four out of six minor bubble states and zero non-bubble states.

From the results in Table 6-1, we see that GSVI for the two best performing queries, namely Housing Bubble and Real Estate Agent, outperforms the self-created indexes. We created four different indexes consisting of the average GSVI for the twenty, twelve, six and three single best-performing search terms to improve the robustness and the level of information captured. Viewing the results, we see that this is not the case. From the full test results in Appendix G, we find that the top two single search terms, in addition to getting the highest test score, are displaying more robustness by performing rather well on a wide range of M values. Taking predictive ability, robustness and simplicity into account, GSVI for single search terms seems most fitting as housing bubble indicators. The search term Housing Bubble seems particularly suitable as a bubble indicator as it performed best on all four tests. An advantage of using single queries, such as Housing Bubble and Real Estate Agent over indexes, is that they can be combined and hence increase the robustness and level of market information captured by the bubble indicator. Also, GSVI for single search terms is easier to download and compute.

6.2 Peaks and Troughs for the Search Terms Compared to the House Price

The time difference between peaks and troughs in GSVI for Housing Bubble, Real Estate Agent and a self-created index (Index12) against the Housing Price Index are displayed in the table below. The two listed search terms and the index are the ones that outperformed the other single search queries and indexes shown in Table 6-2. In addition to the ranking results, we include the average number of quarters, ΔQ, from the peak and trough in GSVI level for the search terms and index compared to the top in HPI for the real, minor and non-bubble states. A positive ΔQ indicates that the GSVI for the respective search term and index leads the HPI, and conversely for negative values.

ΔTime	**Housing Bubble - HPI**		**Real Estate Agent - HPI**		**Index12 - HPI**	
State	**ΔQ Top**	**ΔQ Trough**	**ΔQ Top**	**ΔQ Trough**	**ΔQ Top**	**ΔQ Trough**
Nevada	1.00	-6.00	1.00	-6.00	2	-10
Arizona	5.00	1.00	9.00	-6.00	5	-13
Florida	5.00	-7.00	8.00	3.00	6	-2
California	3.00	-3.00	5.00	5.00	4	5
Average RBS	**3.50**	**-3.75**	**5.75**	**-1.00**	**4.25**	**-5**
Maryland	5.00	4.00	9.00	6.00	6	-7
Oregon	7.00	-14.00	11.00	-9.00	7	-14
Washington	4.00	-6.00	5.00	2.00	7	-6
New Jersey	5.00	1.00	11.00	8.00	5	-8
Connecticut	2.00	0.00	8.00	18.00	2	5
Virginia	5.00	-11.00	8.00	7.00	5	-11
Average MBS	**4.67**	**-4.33**	**8.67**	**5.33**	**5.3**	**-6.8**
Kansas	N/A	N/A	4.00	-8.00	5	-8
Nebraska	N/A	N/A	5.00	4.00	-1	-12
Wyoming	N/A	N/A	11.00	6.00	10	-15
Louisiana	N/A	N/A	12.00	2.00	6	-7
Alaska	N/A	N/A	11.00	12.00	4	-10
Texas	3.00	-6.00	10.00	5.00	8	-7
Iowa	N/A	N/A	1.00	-18.00	-1	-7
South Dakota	N/A	N/A	12.00	15.00	-2	-9
Oklahoma	N/A	N/A	12.00	-22.00	8	-25
North Dakota	N/A	N/A	9.00	-5.00	7	-25
Average NBS	**3.00**	**-6.00**	**8.70**	**-0.90**	**4.4**	**-12.5**

Table 6-2: The table show number of quarters, ΔQ, that Google Search Volume Index (GSVI) for Housing Bubble, Real Estate Agent and a self-created index (Index12) peaked and troughed before the Housing Price Index (HPI) peaked and troughed for the real, minor and non-bubble states. A positive value for ΔQ indicates that the GSVI for the respective queries peaked/troughed before the HPI peaked/troughed and vice versa. N/A means there are missing GSVI data for the respective state.

From the result in Table 6-2, we see GSVI for both search terms and Index12 peaks before the house prices, on average, for the real, minor and non-bubble states. We further find GSVI for Real Estate Agent to peak before Housing Bubble and Index12 for all three state groups and seems to be leading during the bubble period. When looking at the troughs, we find the house prices to be slightly leading the GSVI for Real Estate Agent in the real and non-bubble states, but the variance from state to state is too large to conclude anything. In the group of minor bubble states, GSVI for Real Estate Agent leads the house prices by more than five quarters. GSVI for Housing Bubble reaches the troughs roughly four quarters after the house prices for the real and minor bubble state groups, while Index12 is lagging several periods more compared to the house prices. GSVI for Housing Bubble is not recorded/published by Google in nine out of the ten non-bubble states due to search volume levels being under a minimum threshold. We interpret the low search volume levels in two ways; first, low interest in the housing market and housing bubbles, which is understandable for states that did not experience a sharp increase in house prices and high level of animal spirits. Second, several of the non-bubble states have a relatively low population, which diminishes the quality of the data and are prone to low search volumes for specific queries such as Housing Bubble.

6.3 Correlation between Google Search Volume Levels and the House Prices

In the table below, we present the correlation between the house prices and GSVI for the most promising search terms and index for different periods, namely Housing Bubble, Real Estate Agent and Index12.

Correlation	**Housing Bubble - HPI**			**Real Estate Agent - HPI**			**Index12 - HPI**		
State Name	**WP**	**BP**	**NP**	**WP**	**BP**	**NP**	**WP**	**BP**	**NP**
Nevada	0.486	0.301	-0.11	0.874	0.794	0.326	0.78	0.923	-0.695
Arizona	0.855	0.701	0.397	0.846	0.902	-0.048	0.73	0.886	-0.718
Florida	0.887	0.776	0.478	0.957	0.955	0.955	0.84	0.922	-0.489
California	0.925	0.938	0.793	0.963	0.968	0.898	0.78	0.920	-0.465
Ave RBS	**0.788**	**0.679**	**0.390**	**0.910**	**0.905**	**0.533**	**0.78**	**0.913**	**-0.592**
Maryland	0.919	0.638	0.633	0.940	0.820	0.736	0.88	0.787	0.182
Oregon	0.697	0.308	-0.22	0.620	0.118	-0.624	0.67	0.750	-0.540
Washington	0.766	0.385	0.617	0.817	0.573	0.433	0.64	0.497	-0.416
New Jersey	0.939	0.686	0.407	0.884	0.746	0.576	0.89	0.797	0.478
Virginia	0.854	0.479	-0.41	0.860	0.921	0.721	0.81	0.834	-0.585
Connecticut	0.723	0.577	0.366	0.880	0.873	-0.285	0.91	0.815	0.722
Ave MBS	**0.816**	**0.512**	**0.231**	**0.833**	**0.675**	**0.260**	**0.80**	**0.747**	**-0.027**
Kansas	N/A	N/A	N/A	0.752	0.729	-0.012	0.66	0.436	-0.296
Nebraska	N/A	N/A	N/A	0.705	0.658	0.444	0.44	0.232	-0.507
Wyoming	N/A	N/A	N/A	0.544	0.647	0.493	0.19	0.231	-0.505
Louisiana	N/A	N/A	N/A	0.675	0.532	-0.140	0.54	0.321	-0.263
Alaska	N/A	N/A	N/A	0.628	0.332	0.141	0.34	0.152	-0.409
Texas	0.045	0.114	0.464	0.271	0.060	0.848	0.25	0.073	-0.576
Iowa	N/A	N/A	N/A	0.753	0.591	0.178	0.62	0.175	-0.212
South Dakota	N/A	N/A	N/A	0.360	0.198	0.350	-0.37	0.123	-0.695
Oklahoma	N/A	N/A	N/A	0.563	0.468	-0.488	0.36	-0.14	-0.402
North Dakota	N/A	N/A	N/A	0.307	-0.09	0.616	-0.23	0.338	-0.726
Average NBS	**N/A**	**N/A**	**N/A**	**0.556**	**0.412**	**0.243**	**0.28**	**0.202**	**-0.459**

Table 6-3: The table shows the correlation between: Google Search Volume Index (GSVI) for Housing Bubble and the Housing Price Index (HPI), GSVI for Real Estate Agent and HPI, GSVI for Index12 and HPI. The correlation is displayed for the whole period (WP), Q1 2004 – Q3 2016, the bubble period (BP), Q1 2004 – Q2 2010, and the normal period (NP), Q3 2010 – Q3 2016. The correlation is calculated for the states defined as real bubble states (RBS), minor bubble states (MBS) and non-bubble states (NBS). Also, the average for each of the three groups is calculated.. N/A means there are missing GSVI data for the respective state.

Table 6-3 display the correlation between GSVI for Housing Bubble, Real Estate Agent, and Index12 against the house prices in the bubble period, Q1 2004 – Q2 2010, the normal period, Q3 2010 – Q3 2016, and the whole period, Q1 2004 – Q3 2016. GSVI for both search terms and the index displays significantly higher correlation during the bubble period than the non-bubble period. In general, the results show higher correlation for Real Estate Agent, then Housing Bubble, for the whole period, the bubble period and the non-bubble period. For the real and minor bubble states during the bubble period, Index12 displays even higher correlation than Real Estate Agent, 91.3% and 74.7% respectively. For the non-bubble period Index12, show negative correlation to the housing prices for all three state groups.

GSVI for Real Estate Agent shows highest correlation in the real bubble states with an average of 91%. In the states defined as minor bubble states, we see that the average correlation is slightly lower at 83.4% and in the non-bubble states even less with 55.6%. In general, for the three groups, the correlation is higher for lagged values of the Google Searches. This indicates that GSVI for Real Estate Agent is leading the Housing Price Index.

GSVI for Housing Bubble display slightly higher correlation in the minor bubble states, 81.6%, compared to the real bubble states, 78.8%. GSVI for Housing Bubble is not recorded/published by Google in nine out of the ten non-bubble states due to search volume levels being under a minimum threshold. We interpret this in the same way as in section 6.2. Comparing GSVI for Housing Bubble with Real Estate Agent and Index12, we find the former and latter to require fewer lags to reach the highest correlation with the house prices. This indicates that Real Estate Agent is leading the house prices more than Housing Bubble and Index12 is leading the house prices.

6.4 ECM Results for the United States

In the table below, we display the results from the regression of the house prices at level for assessment of the long-run effects from Google searches and the result from the error correction model to assess the short-run effects and the speed of adjustment from Google searches for the whole of the United States.

Short and Long Run Effects from GSVI on the House Prices for the U.S.

Model Variables	Long Run Effects			Speed of Adjustment		Short Run Effects		
	LR C	**P>Z**	**LR R^2**	**SA C**	**P>Z**	**SR C**	**P>Z**	**SR R^2**
HB	0.120	0.000	0.804	0.019	0.645	0.073	0.000	0.314
REA	0.486	0.000	0.896	-0.294	0.000	0.180	0.139	0.567
Index12	0.265	0.000	0.752	-0.046	0.198	0.175	0.013	0.176
HB + L2.HB	0.195 -0.079	0.000 0.000	0.848	0.015	0.704	0.052 0.042	0.010 0.004	0.327
REA + L2.REA	0.051 0.453	0.492 0.000	0.964	-0.298	0.005	0.252 0.228	0.012 0.002	0.593
Index12 + L2.Index12	0.296 -0.016	0.006 0.873	0.782	-0.026	0.437	0.186 0.129	0.05 0.013	0.238
REA + HB	0.325 0.053	0.000 0.000	0.956	-0.190	0.026	0.279 0.043	0.010 0.000	0.516
L.HPI + HB	1.102 -0.017	0.000 0.004	0.974	-0.261	0.371	0.692 0.038	0.056 0.036	0.442
L.HPI + REA	0.711 0.156	0.000 0.000	0.988	-0.814	0.001	0.832 0.162	0.000 0.103	0.651
L.HPI + Index12	0.928 0.02	0 0.165	0.973	-0.929	0.021	1.441 0.124	0.001 0.092	0.483
L.HPI + REA + HB	0.732 0.153 -0.002	0.000 0.000 0.616	0.988	-0.972	0.001	0.991 0.158 -0.022	0.000 0.109 0.150	0.656

Table 6-4: The Table shows the result of an error correction model (ECM) regressing the Housing Price Index (HPI) using only Google Search Volume Index (GSVI) for Housing Bubble (HB), Real Estate Agent (REA) and a self-created index (index12) consisting of the twelve best-performing search terms. L2 in front of a variable stands for a two period lag of the respective variable. LR R^2 is the long run coefficient of determinations, SR R^2 is the short-run coefficient of determination, SA C is the coefficient for the speed of adjustment, and P>Z is the probability that the respective coefficient is significant.

The results in Table 6-4 shows GSVI for Real Estate Agent performs significantly better than both Housing Bubble and Index12, which was found in section 6.1 to be the best performing index, at all points in both the long and short-term for all the models in the United States. Only the models including GSVI for Real Estate Agent have significant values for the speed of adjustment, which means there are cointegration and long run effect running from GSVI for only Real Estate Agent to the House Price Index (HPI). Index12 display some signs of a long run relationship but this is not significant at a ten percent level. GSVI for Housing Bubble is not cointegrated with the HPI and thus, do not explain the house prices in the long run. Housing Bubble is not an everyday term, and we expect search volume levels for it to be relatively low except for in bubble phases as outlined by Aliber and Kindleberger (2005). Therefore, we find it as no surprise that GSVI for Housing Bubble and the house prices are not cointegrated. We did expect the GSVI for Index12 to be cointegrated with the house prices and to perform better, but taking the average GSVI of several queries seems to diminish the information captured.

In the short run, both GSVI for Housing Bubble and Index12 display explanatory power on the house prices. When including GSVI for both Housing Bubble and Real Estate Agent, we find the results to be similar to those produced using only GSVI for Real Estate Agent. Substituting Housing Bubble with a two period lag of Real Estate Agent yields improved results. This indicates that inclusion of GSVI for Housing Bubble does not capture more of the market information than Real Estate Agent do alone.

Real Estate Agent shows good predictive results, explaining the house prices in both the short and long run. We also see that the speed of adjustment is relatively high for all models. When only including GSVI for Real Estate Agent, without any lags, to explain the house prices, we see the long run coefficient is 48.6%, and the long run coefficient of determinations (R^2) is 89.6%. The speed of adjustment is -29.4%, the short-run coefficient is 18%, and the short-run coefficient of determinations is 56.7%. The r-squared values are high for both the short and long run effects. The speed of adjustment is 29.4%, meaning that every period/quarter the error correction term will move by 29.4% towards the long run equilibrium between GSVI for Real Estate Agent and HPI. Taking into account that lags of the dependent variable is not included shows the explanatory power of GSVI for Real Estate Agent on the HPI. When including a two period lag of GSVI for Real Estate Agent, we see that the coefficient of determinations increases to respectively 96.4% and 59.3%, while the speed of adjustment stays the same. Substituting the two period lag of GSVI with a one period lag of the independent variable HPI creates major changes. The coefficient of determinations increases to respectively 98.8% and 65.1%, and we see that the one period lag of HPI now stands for most of the explanation in both the short and long run. Still, GSVI for Real Estate Agent is significant with a short run coefficient of 15.6% and long run coefficient of 16.2%. We find the greatest change in the speed of adjustment, which has increased to from -29.8% to -81.4%. These results show that even simple linear models, including only GSVI and a one period, lagged variable of HPI can explain the house prices.

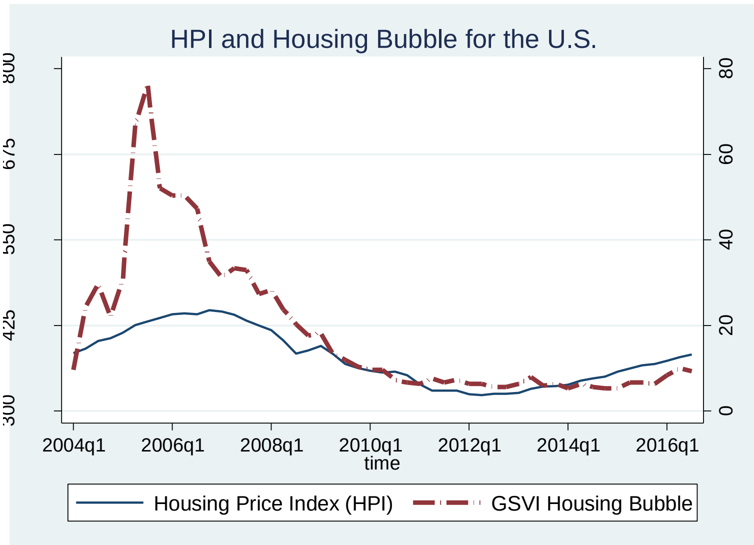

Figure 6-1: The figure display the Housing Price Index (HPI) on the left y-axis against Google Search Volume Index (GSVI) for Housing Bubble on the right y-axis for the United States.

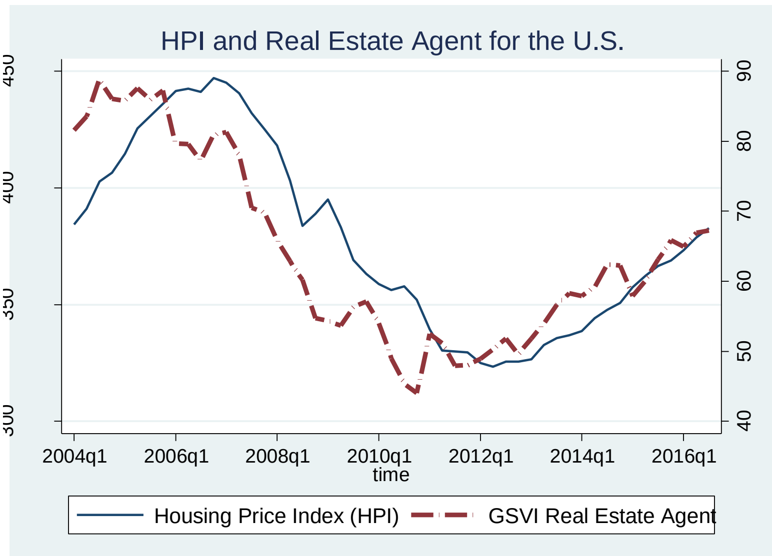

Figure 6-2: The figures display the Housing Price Index (HPI) on the left y-axis against Google Search Volume Index (GSVI) for Real Estate Agent on the right y-axis for the United States.

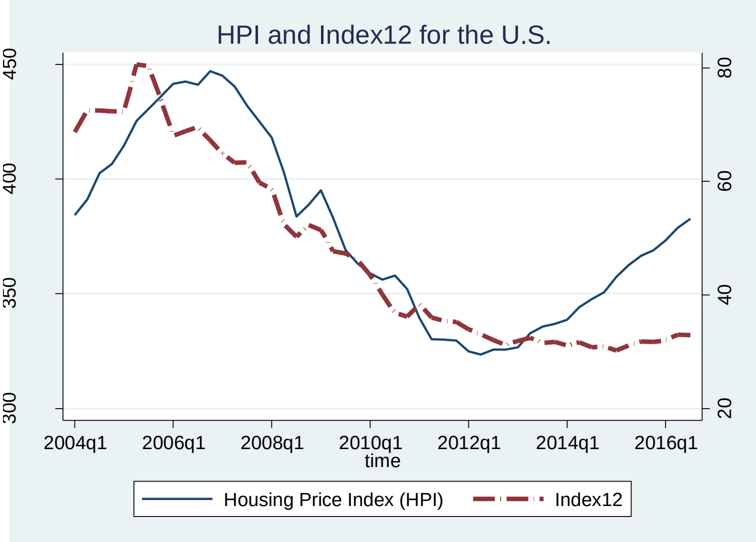

Figure 6-3: The figures display the Housing Price Index (HPI) on the left y-axis against Google Search Volume Index (GSVI) for a self-created index (index12) on the right y-axis for the United States. The Index consist of the average GSVI for the twelve single best search terms from an in-sample prediction test.

The graphs in Figure 6-1, Figure 6-2 and Figure 6-3 show that GSVI for the two search terms and Index12 behaved quite different. The search volume levels for Housing Bubble indicated a bubble in the United States housing market. Search term levels seem to be low, without any trend, before and after the housing bubble. The graph in the upper figure shows how GSVI for Housing Bubble have a rather extreme development in search volumes during the actual bubble, increasing several 100% in a short amount of time before falling back before the house prices start to decrease. Both graphs seem to hit bottom in 2012, but while house prices increase steadily each year, GSVI for Housing Bubble stays at a low level. Viewing the graph in Figure 6-1, it seems as search volume levels for Housing Bubble have high correlation during bubble periods and lower during normal economic times. Due to its explosive increase in search volume level during bubble periods and leading the house prices, GSVI for Housing Bubble could work as a strong bubble indicator on both country and state level.

GSVI for Real Estate Agent and Index12 did not indicate a Housing bubble in the United States as clearly as GSVI for Housing Bubble. Search volume levels for both Real Estate Agent and Index12 shows a falling trend from the top in 2005, indicating that housing would fall. The search volume levels did not display the same explosive increase in search volume levels during the bubble period as Housing Bubble. The search volume seems to be at a more normal level, increasing and decreasing before the Housing Price Index during the housing bubble. GSVI for Real Estate Agent troughs in 2011 while the graph of the HPI flattens out a year later in 2012. The graph displaying index12 in the lower figure, do not hit bottom before several years later in 2015 and while the other two graphs start increasing year by year from the trough, Index12 stays at a low level. Viewing the graphs in Figure 6-2 and Figure 6-3, Real Estate Agent seems to be highly correlated with to the house prices, both during bubble periods and normal times while Index12 seems to be correlated with the house prices only during the housing bubble. Also, both GSVI for Real Estate Agent and Index12 seems to be leading the house prices during the bubble period. Real Estate Agent also leads the house prices in the non-bubble period. From Figure 6-2, we see that GSVI for Real Estate Agent peaks before the HPI and starts falling first, hits bottom first and then start increasing before the house prices do. Viewing, Figure 6-2, Figure 6-3, Table 6-3, and Table 6-4, we find Search volume levels for Real Estate Agent to be leading the House price Index more than Housing Bubble and Index12 is leading the HPI.

Based on the above results we conclude that GSVI for Housing Bubble is well suited as a bubble indicator but not to explain the short and long run effects on the house prices in general. GSVI for Real Estate Agent have higher search volume levels throughout the whole period, are present in all 50 states, have the highest correlation with the house prices and performs the best in-sample prediction result. We, therefore, find search volume levels for Real Estate agent most fitting in our further research on short and long-run effects on the House Prices. In the next two sections, we will therefore only include GSVI for Real Estate Agent.

6.5 ECM Results for all 50 States Using Only Google Searches

In the table below, we present the results from the regression of the house prices at level for assessment of the long-run effects from GSVI for Real Estate Agent and the result from the error correction model to assess the short-run effects, and the speed of adjustment from Google searches for each the fifty states.

Linear Regression of HPI Using Only Google Searches. Long Run Effects							
Model Variables	**L1.HPI**	**P>Z**	**GSVI**	**P>Z**	**L2.GSVI**	**P>Z**	**R^2**
Average Results for the Real Bubble States							
Only GSVI			0.734	0.000			0.709
GSVI + L2.GSVI			0.622	0.005	0.198	0.325	0.822
L1.HPI + GSVI	0.836	0.00			0.162	0.001	0.985
Average Results for the Minor Bubble States							
Only GSVI			0.347	0.000			0.345
GSVI + L2.GSVI			0.54	0.089	-0.125	0.325	0.522
L1.HPI + GSVI	0.931	0.00			0.062	0.065	0.978
Average Results for the 30 states not defined							
Only GSVI			0.278	0.029			0.496
GSVI + L2.GSVI			0.652	0.143	0.136	0.243	0.611
L1.HPI + GSVI	0.916	0.00			0.037	0.118	0.971
Average Results for the Non-Bubble States							
Only GSVI			0.059	0.141			0.245
GSVI + L2.GSVI			-0.002	0.346	0.071	0.298	0.241
L1.HPI + GSVI	0.967	0.00			0.003	0.384	0.932

Table 6-5: The Table shows the long run result of an error correction model (ECM) of the Housing Price Index (HPI) using only Google Search Volume Index (GSVI) for Housing Bubble (HB) and Real Estate Agent (REA). L2 in front of a variable stands for a two period lag of the respective variable. LR R^2 is the long run coefficient of determinations. LR MAE is the Mean Absolute Error (MAE) between predicted value and real value of HPI at level.

ECM Using Only Google Searches to Explain the House Prices. Short Run Effects

Model Variables	**SA C**	**P>Z**	**L1 HPI**	**P>Z**	**GSVI**	**P>Z**	**L2 GSVI**	**P>Z**	**R^2**
Average Results for the Real Bubble States									
Only GSVI	-0.16	0.003			0.176	0.094			0.36
GSVI + L2.GSVI	-0.10	0.065			0.201	0.106	0.17	0.158	0.34
L1.HPI + GSVI	-0.58	0.009	1.074	0.000			0.12	0.050	0.71
Average Result for the Minor Bubble States									
Only GSVI	-0.08	0.068			0.003	0.515			0.17
GSVI + L2.GSVI	-0.07	0.185			0.062	0.402	0.03	0.344	0.17
L1.HPI + GSVI	-0.69	0.047	1.126	0.004			0.02	0.382	0.53
Average Results for the 30 states not defined									
Only GSVI	-0.09	0.123			0.045	0.319			0.18
GSVI + L2.GSVI	-0.09	0.135			0.043	0.356	0.04	0.268	0.19
L1.HPI + GSVI	-0.84	0.088	1.127	0.018			0.05	0.335	0.38
Average Results for the Non-Bubble States									
Only GSVI	-0.04	0.334			0.015	0.472			0.08
GSVI + L2.GSVI	-0.04	0.352			0.013	0.46	0.02	0.495	0.11
L1.HPI + GSVI	-0.96	0.159	0.928	0.055			0.01	0.538	0.22

Table 6-6: The Table shows the short run result of an error correction model (ECM) of the Housing Price Index (HPI) using only Google Search Volume Index (GSVI) for Real Estate Agent (REA). L2 in front of a variable stands for a two period lag of the respective variable. SR R^2 is the short-run coefficient of determination. SR MAE is the Mean Absolute Error between predicted change in HPI and real value. SA C is the coefficient for speed of adjustment and P>Z is the probability that the respective coefficient is significant.

From Table 6-5 and Table 6-6, we see the model using only GSVI for Real Estate to regress the house prices shows good in-sample predictive results. For the states experiencing a real bubble, we see the average long run coefficient is 73.4% and significant, and the average long run coefficient of determination is 70.9%. The average short-run coefficient is 17.6% and significant at 10% confidence interval, and the average short-run coefficient of determination is 36.3%. The speed of adjustment is -15.6%. Inspecting the full results more closely, see Appendix I, we find the in-sample prediction results to be significantly better for California and Florida than for Nevada and Arizona. The short-run coefficient of determination is respectively 57.3% and 50.8% for the former and respectively 15.2 and 21.9% for the latter.

Including a two period lag of GSVI for Real Estate Agent increases the long run coefficient of determination to 82.2%, while decreasing the short run coefficient of determination and speed of adjustment to respectively 34.3% and -10.1%. Substituting the two-period lag with a one period lag of the dependent variable HPI creates more major changes. Both the long and short run coefficient of determinations increases to respectively 98.5% and 71.4%, while the speed of adjustment increases to -58.1%. We find the same throughout the groups of real, minor, and non-bubble states.

Evaluating the other state groups in Table 6-5 and Table 6-6, we find the coefficient of determinants for both the long and short run to be largest for the real bubble states and least for the non-bubble states. For the minor bubble states and the thirty states not defined as either bubble nor non-bubble states, we find the opposite result. This might be explained by two factors; first is the general bubble that existed globally in the U.S. housing market. Secondly, we suspect the size of the population in each state to affects the quality of the respective Google Trend data in the state.

In our work with this paper, we also constructed a Vector error correction model (VECM) to investigate the relationship between Google search and the house prices at state level. Due to the rigidity of the model and problems interpreting the results from the baseline models, which had several long run relationships, we decided to use other models as derived in section 0. Never the less, we ran the model using GSVI for Real Estate Agent and Index12, separately, for all 50 states and will briefly discuss our findings even the result is not included in the Appendix. The result from the VECM was coinciding with those presented above. GSVI for Real Estate Agent is leading the house prices and have long run effects on the house prices in the real and minor bubble states. For the thirty states that did not experience any bubble, the effect from GSVI for Real Estate Agent was somewhat less and more equal to the effect running from the house prices and towards the Google searches. In the non-bubble states, we found the effect from the house prices towards the Google searches to be stronger and more significant than the other way around. The result for GSVI for Index12 showed similar tendency but weaker and less significant results. The VECM results reinforces our findings of a long run relationship between Google searches and the house prices at state level, where the former are leading in the states experiencing a bubble.

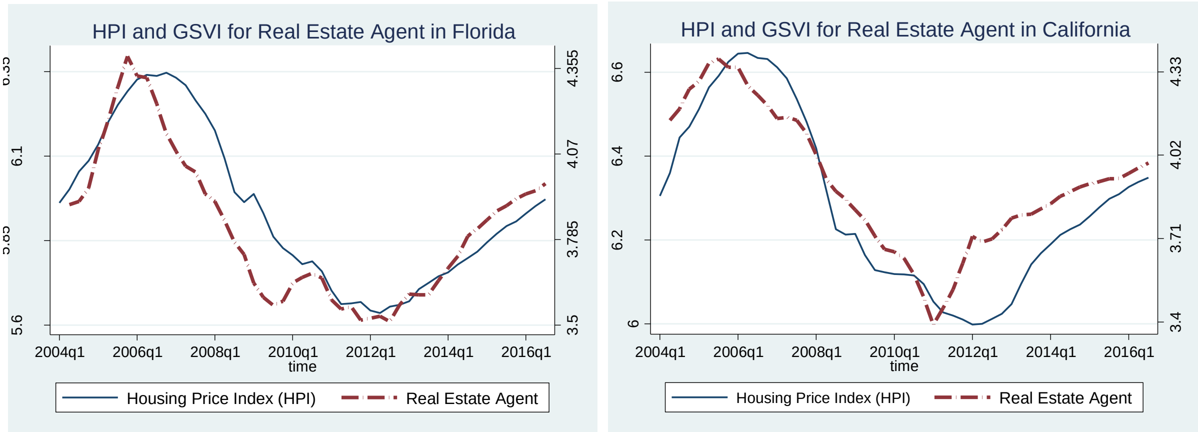

Figure 6-4: The figures labels the Housing Price Index (HPI) on the left y-axis and the Google Search Volume Index (GSVI) for Real Estate Agent on the right y-axis. The figures display the graphs for two of the states defined as real bubble states. Both time-series are transformed to logarithmic form and adjusted for inflation and seasonal effects.

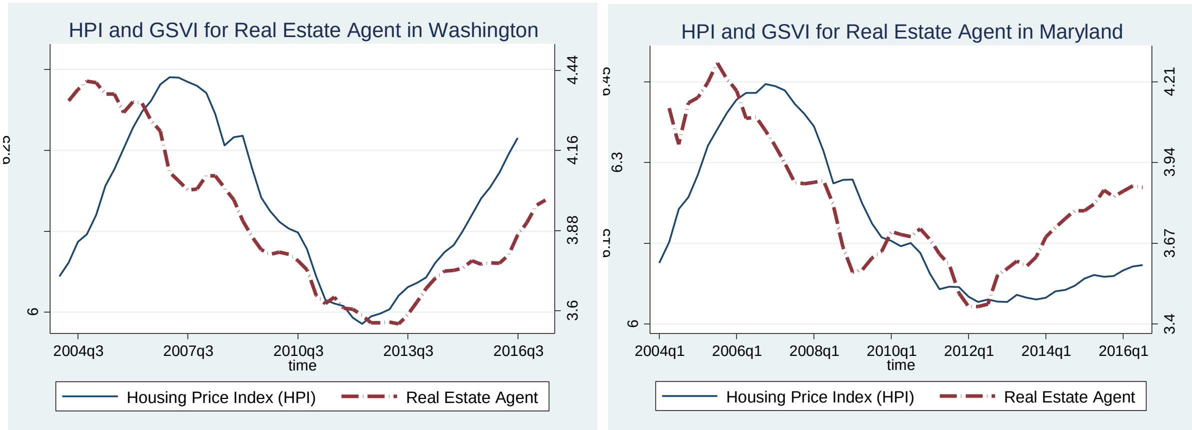

Figure 6-5: The figures labels the Housing Price Index (HPI) on the left y-axis and the Google Search Volume Index (GSVI) for Real Estate Agent on the right y-axis. The figures display the graphs for two of the states defined as minor bubble states. Both time-series are transformed to logarithmic form and adjusted for inflation and seasonal effects.

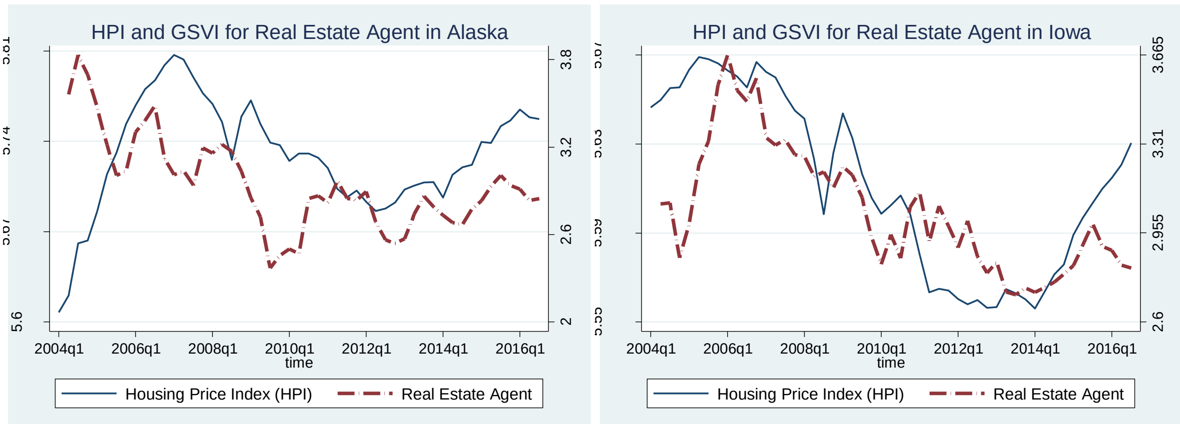

Figure 6-6: The figures labels the Housing Price Index (HPI) on the left y-axis and the Google Search Volume Index (GSVI) for Real Estate Agent on the right y-axis. The figures display the graphs for two of the states defined as non-bubble states. Both time-series are transformed to logarithmic form and adjusted for inflation and seasonal effects.

Figure 6-4, Figure 6-5 and Figure 6-6 display the GSVI for Real Estate Agent against the house prices for two of the real, minor and non-bubble states. Viewing the graphs, we see how the fit between the time-series changes in the different groups of states. Starting at the states experiencing a real bubble, we find GSVI for Real Estate Agent to fit the house prices extremely well, indicating a high correlation between the two time series for the whole period. Next, viewing the graphs in the two middle figures for the minor bubble states, we find the two time-series following closely but less than for the real bubble states. For the non-bubble states, we can still see that the two time-series moves together in the long run, but they do not fit as closely as for the real and minor bubble states. The tendency of higher correlation, between Google searches and the house price, the more of a bubble the respective state experienced is in accordance with the result we found in Table 6-5 and Table 6-6. In general, the GSVI for Real Estate Agent is leading the house prices in all the six states during the bubble period, but in the non-bubble period, the results are more coinciding.

6.6 ECM Results for all 50 States Using the Baseline Variables

In this section, we will go through and compare the results from the baseline model with and without the inclusion of Google searches. In addition to testing the effect of including Google searches in the model, we will compare the result to the effect of including the Consumer Confidence Index (CCI).

Model Description	**LR R^2**	**LR MAE**	**SR R^2**	**SR MAE**	**SA C**	**P>Z**
Average results for the Real Bubble States						
Google Model	0.985	2.355%	0.714	1.516%	-0.581	0.009
Baseline Model	0.992	1.494%	0.816	1.153%	-0.616	0.006
Baseline GSVI Model	0.993	1.440%	0.834	1.146%	-0.664	0.002
Baseline CCI Model	0.992	1.492%	0.824	1.182%	-0.594	0.008
Average results for the Minor Bubble States						
Google Model	0.978	1.391%	0.529	1.088%	-0.685	0.047
Baseline Model	0.987	1.017%	0.739	0.847%	-0.695	0.004
Baseline GSVI Model	0.988	0.972%	0.760	0.815%	-0.734	0.002
Baseline CCI Model	0.987	1.014%	0.749	0.833%	-0.697	0.002
Average results of the Thirty States not Defines as either Bubble nor non-bubble						
Google Model	0.971	1.128%	0.375	0.959%	-0.838	0.088
Baseline Model	0.979	0.879%	0.634	0.772%	-0.782	0.003
Baseline GSVI Model	0.980	0.852%	0.660	0.749%	-0.789	0.001
Baseline CCI Model	0.979	0.865%	0.648	0.753%	-0.754	0.007
Average results for the Non-Bubble States						
Google Model	0.932	0.850%	0.223	0.765%	-0.955	0.159
Baseline Model	0.944	0.715%	0.488	0.661%	-0.858	0.009
Baseline GSVI Model	0.943	0.707%	0.503	0.653%	-0.891	0.007
Baseline CCI Model	0.943	0.712%	0.499	0.652%	-0.856	0.010

Table 6-7: The table summarises three different versions of a baseline housing price model with Disposable Personal Income, Housing Permits Authorised, Unemployment Rate, Interest Rate and Population as explanatory variables. Also, a one period lag of the dependent variable is included. The "Baseline Model" includes the former variables, "Baseline Model Including GSVI" includes Google Search Volume Index (GSVI) for Real Estate Agent in addition to the other variables and "Baseline Model Including CCI" includes Consumer Confidence Index (CCI) instead of Google searches. In addition to this these three Baseline Models, we have "Model Only Using GSVI and L1.HPI" which is the best performing model using only GSVI for Real Estate and a one period lag of the dependent variable the Housing Price Index (HPI). The four models are assessed after the following criteria's; LR R^2 is the adjusted long run coefficient of determinations, LR MAE is the Mean Absolute Error (MAE) between predicted value and real value for HPI at level, SR R^2 is the adjusted short-run coefficient of determination, SR MAE is the MAE between predicted change in HPI and real value, SA C is the coefficient for speed of adjustment and P>Z is the probability that the coefficient is significant.

Viewing the result in Table 6-7, we see that all points of criteria are improved when including GSVI for Real Estate Agent in the baseline model. The adjusted coefficient of determination is increased for both the long and short run, and the speed of adjustment is both higher and more significant. These results apply for the real, minor and non-bubble states. In addition to the thirty states not defined as either bubble nor non-bubble states.

For the real bubble states, including GSVI for Real Estate Agent reduced the mean absolute error (MAE) on average with respectively 0.61% for the long run in-sample prediction and 3.78% for the short run in-sample prediction. For the minor bubble states, the MAE was reduced with respectively 4.42% for the long run in-sample prediction and 3.78% for the short run in-sample prediction. In the thirty states not defined as either bubble nor non-bubble states, there was the following improvement for the long and short run in-sample prediction MAE with respectively 3.1% and 2.97%. Last, for the non-bubble states, the average improvement in reduced MAE was respectively 1.11% and 1.21%.

Substituting Google searches with the Consumer Confidence Index (CCI) yields significantly worse results on all points of criteria except one, the short run MAE for the non-bubble states are on average reduced by 0.15%. Including CCI in the Baseline Model improves the MAE in both the long and short run but display a decreased coefficient of determination and lower speed of adjustment. Based on the results above, we conclude that GSVI for Real Estate Agent improves both the fitness of the Baseline Model and reduces the MAE of the in-sample prediction in both the long and short run. Also, the inclusion of GSVI for Real Estate Agent yields significantly better results than the inclusion of CCI.

Note: As described in the previously section, we also constructed a vector error correction model (VECM) using all the baseline variables. We included GSVI for Real Estate Agent and Index12, separately, for all the 50 states. Our findings was coinciding with those above.

Comparing the result from the model using only GSVI for Real Estate Agent and a one period lag of the dependent variable with the Baseline Model, we find the latter to perform better. The former model shows higher speed of adjustment for the thirty states not defined as either bubble nor non-bubble states and the non-bubble states. Assessing the long run coefficient of determination results, we find them to be coinciding with slightly better results for the Baseline Model. The major difference in performance is in the short run, where the Baseline Model display better fit. Still, we find the in-sample prediction results for such a simple model to be rather good.

7 Discussion

The purpose of this paper has been to operationalize points 2, 3, 4, 5 and 6 in Shiller's (2010) checklist for asset pricing bubbles, using Google Trends. Starting with 204 housing related search terms, we reduced them to 20 based on their correlation with the house prices in the identified bubble states. Next, we used a two-folded approach to operationalizing the points. First, we constructed a bubble identification test, based on differences in Google search volume during to the U.S. housing bubble compared to levels in normal economic times, for the 20 search terms. In addition, we created four indexes consisting of the average GSVI for the 20, 12, 6 and 3 most promising search terms, and tested them together with the single queries. Our belief was that housing related queries, and indexes, could capture the increased level of confidence and animal spirits which Akerlof and Shiller (2005) argue were visible during the boom period of the 06/07 U.S. housing bubble. We found Google Search Volume Index (GSVI) for Housing Bubble, Real Estate Agent and Index12 to lead the house prices during the bubble period and to indicate states experiencing a bubble. These properties are in accordance with Lind (2009) who argues that the most important aspect of a housing bubble indicator is its predictive abilities. He states that the indicator should be able to indicate that a period of dramatic price increases will be followed by a period of dramatic price decreases.

In addition to its predictive abilities, we need to assess the practical implications of computing GSVI for the two search terms and the best performing index, namely Index12. Therefore, to make a recommendation, we need to assess the simplicity and the robustness of the corresponding GSVI test performance results. With simplicity, it is referred to how easy it is to both compute the GSVI indicator and monitor the developments in the interest of the underlying search terms. In this respect, GSVI for the single search terms is easier to compute, as they do not involve downloading and calculating the averages of several queries. Increased complexity is only valuable if it yields improved performance. None of the indexes we created yielded better nor equally good results as the GSVI for Housing Bubble and Real Estate Agent. Also, when reviewing the full results from the bubble identification tests in Appendix G, we found GSVI for Housing Bubble to display significantly more robustness than Index12. GSVI for Housing Bubble shows especially good results for indicating states experiencing a housing bubble and detecting a global bubble in the United States.

When optimising to detect all the states experiencing a bubble, GSVI for Housing Bubble only erroneously included one non-bubble state and when optimising on not wrongly include any non-bubble states, it detected all four real bubble states and four out of six minor bubble states. It repeatedly produced the same results for a wide variety of tests. Search volume levels for Housing Bubble is low, without any trend, before and after the housing bubble period. During the actual bubble period, GSVI for Housing Bubble has a rather extreme development in search volumes, increasing several 100% in a short amount of time before falling back, leading the house prices. We regard these extreme developments in search volume levels to be changes in confidence and animal spirits. The level of confidence and animal spirit are fueled by rising house prices and stories, which is reinforced through the feedback loop (Shiller, 2005). We interpret the levels of GSVI for Housing Bubble to be a measure of points 2, 3, 4, and 5 from Shiller`s (2010) asset-pricing bubble checklist. The points “Great public excitement about said increases,” “An accompanying media frenzy,” “Stories of people earning a lot of money, causing envy among people who are not,” and “Growing interest in asset class among the general public” are all indicators of a bubble. Thus, we conclude that GSVI for Housing Bubble operationalizes several of the points in Shiller`s (2010) asset-pricing checklist, and are well suited as a bubble indicator. Our findings of GSVI for Housing Bubble as a strong bubble indicator is not in accordance with Fama (2014), and the efficient-market hypothesis, who rejects bubbles on empirical grounds by referring to the lack of reliable evidence that price declines are predictable and thus arguing that indicators cannot exist. Pentland (2010) found Google searches to precede purchase decisions, which is coinciding with our findings of Google searches leading the house prices. Several other economist have found online behavior to reveal consumers intention and predict purchase decisions (see, e.g. Kuruzovich et al. 2008, Horrigan 2008, Brynjolfson, Hu, and Rahman 2013).

Second, after ranking the 20 search terms, we further tested GSVI for Housing Bubble, Real Estate Agent and Index12, which was top three, to find their predictive power of the house prices. First, we predicted the house prices globally in the U.S. using only GSVI for each of the two queries and Index12. GSVI for Real Estate Agent displayed significant higher predictive power then the other two for the real, minor and non-bubble states. Next, we tested the correlation between the three GSVI and the HPI in the bubble period, the non-bubble period and the whole period and found Real Estate Agent to have the highest correlation with the house prices, especially during the non-bubble period.

We interpret these results as GSVI for Real Estate Agent captures more of the low levels of confidence and animal spirits that exist in normal economic periods. Point 5 in Shiller`s (2010) asset-pricing bubble checklist, “Growing interest in asset class among the general public”, is a point that could be sensitive to small changes in confidence and animal spirit and thereby being captured by changes in Google searches, also for periods without a housing bubble. We find GSVI for Real Estate Agent to best captures these changes and we therefore further test the predictive power at state level.

GSVI for Real Estate Agent and the HPI are cointegrated in 45 out of 50 states, and the former is leading the house prices in both the bubble and the non-bubble period. When testing the relationship between GSVI for Real Estate Agent and the HPI at state level, we found both short and long-term effects running from the former to the latter. These effects were significant in states experiencing a *real*, *minor* and no bubble. Constructing a simple linear model using only GSVI for Real Estate Agent and a one period lag of the dependent variable, HPI, produced good in-sample prediction results. The fit of the model and the *mean absolute error* results was best for the states experiencing a *real* bubble, followed by the states experiencing a *minor* bubble and least for the states experiencing no bubble. We observe that the predictive power of Google searches was even higher for house prices globally in the U.S than for the *real* bubble states. When moving from country to state level, the total amount of Google searches will be lower and we assume the quality of the data reduced. Thus, we expect GSVI, in general, to have higher explanatory power on the house prices in states with a large population compared to states with a low population.

Including GSVI for Real Estate Agent in our Baseline error correction model for the house prices, improved all points of criteria. The adjusted coefficient of determination increases for both the short and long run and the speed of adjustment is higher and more significant. Substituting Google searches with the well-established Consumer Confidence Index yielded worse result for all assessments. The results are valid for the *real*, *minor* and non-bubble states. In addition to the thirty states not defined as either bubble nor non-bubble states. Based on our findings of Google searches outperforming the CCI, we interpret our results as a operationalization of Shiller`s (2010) asset-pricing bubble checklist points.

Our results are similar to those of Wu and Brynjolfson (2009; 2015) who found evidence that queries submitted to Google`s search engine are correlated with a house price index – specifically the Case-Shiller Index – released by the Federal Housing Agency. Even though we used a different housing price index and different search terms and indexes, our results are coinciding that Google searches can improve the prediction of housing prices and trends. Choi and Varian (2009; 2011) found that queries can be useful leading indicators for subsequent customer purchases in situations where customers start planning purchases significantly in advance of their actual purchase decision. Their findings is especially applicable for our case study due to the time it takes to purchase a home. Furthermore, our findings of the short-term effect of Google Search volume levels coincides with the results found by Preis et al. (2010; 2013). They found Google search volume reflected the current state of the stock market. Bijl et al. (2015) investigated the predictive power of Google search volume on stock returns and found quarterly searches to be positively related to excess return without reversal. They further found that there is an economic value of including Google search statistics in forecasting models, which is in accordance with our own findings.

Others are more skeptical to the use of web searches in prediction. Goel et al. (2010) points out that even search data is easy to acquire and is often helpful in making forecasts, it may not provide dramatic increases in predictability. Our results strongly rejects this as we have found the inclusion of GSVI for Real Estate Agent to significantly improve the prediction in both states experiencing a bubble and the states that did not experiencing any bubble. In addition, inclusion of Google searches outperformed the well-established Consumer Confidence Index (CCI) for all state groups. Damien and Ahmed (2013) investigate previously results that Google search volume can predict future financial index returns but find that trading strategies based on financial related queries do not outperform strategies based on unrelated search terms. Their results differs from our findings whether Google have predictive power but there is a major difference between the liquidity of stocks and homes. Therefore, our findings of strong evidence for a long run effect running from GSVI for Real Estate Agent towards the House prices, is not contradicted by their results.

8 Conclusion

We have operationalized points 2, 3, 4, 5 and 6 in Shiller's list, using Google Trends data, and tested several Google Search Volume Indexes (GSVI) with good in-sample predictive abilities. Taking predictive abilities, simplicity and robustness into consideration, we conclude that the best candidate as a housing bubble indicator is GSVI for Housing Bubble. When optimising to detect all the states experiencing a bubble, GSVI for Housing Bubble erroneously included only one non-bubble state and when optimising on not wrongly including any non-bubble states, it detected all four *real* bubble states and four out of six *minor* bubble states. It repeatedly produced the same results for a wide variety of tests. For the states experiencing a housing bubble, GSVI for Housing Bubble displays relative low search volume levels, without any trends both before and after the bubble, but during the actual bubble period the search volume levels "explodes", increasing several 100%. Search volume levels for Housing Bubble globally in the U.S. displayed the same characteristics, leading the house prices and strongly indicating a real estate bubble. The extreme characteristics of GSVI for Housing Bubble during a bubble period, means there is no need to adjust the data for neither seasonally affects nor trends. Thus, simplifying the surveillance of the indicator.

GSVI for Real Estate Agent displays the highest correlation with the Housing Price Index (HPI) and yield the best in-sample predictive results of the house prices in both the short and long run. Also, GSVI for Real Estate Agent and the HPI are cointegrated in 45 out of 50 states, and the former is leading the house prices in both the bubble and the non-bubble period. When testing the relationship between GSVI for Real Estate Agent and the HPI, we found both short and long-term effects running from the former to the latter. These effects were significant in states experiencing a *real*, *minor* and no bubble. Constructing a simple linear model using only GSVI for Real Estate Agent and a one period lag of the dependent variable, HPI, produced good in-sample prediction results. The fit of the model and the *mean absolute error* results was best for the states experiencing a *real* bubble, followed by the states experiencing a *minor* bubble and least for the states experiencing no bubble. Predicting the house prices, using the same model, globally in the U.S. gave even better results than for the states experiencing a *real* housing bubble.

Including GSVI for Real Estate Agent in our Baseline error correction model for the house prices, improved all points of criteria. The adjusted coefficient of determination increases for both the short and long run and the speed of adjustment is higher and more significant. Substituting Google searches with the well-established Consumer Confidence Index yielded worse result for all assessments. The results are valid for the *real*, *minor* and non-bubble states. In addition to the thirty states not defined as either bubble nor non-bubble states.

Based on the results found in this paper, we conclude that GSVI for Housing Bubble can be a strong housing bubble indicator while GSVI for Real Estate Agent can predict the housing trend and be included in price models to improve their predictive abilities at state levels.

9 Further Research

To further assess the predictive abilities of the GSVI for Housing Bubble as a housing bubble indicator it should be tested on out-of-sample data for other countries that experienced a housing bubble at the same time as, or as a consequence of, the American sub-prime mortgage crises. Google Trends data are only available from 2004; therefore, further research is limited to housing bubbles between now and then. In the wake of the American housing market crash in 06/07 and the following financial crises, which spread to every part of the world, several countries experienced what can be characterized as a real housing bubble. Some of the potential countries coule be, e.g. Ireland, Spain, UK, Italy, Turkey, Denmark, Portugal, Brazil, South Korea, China, Mexico, India and Hong Kong. It would be interesting to find whether GSVI for a direct translation of Housing Bubble can capture the same level of confidence and animal spirit for some of these countries, as in the United States. Housing Bubble is a search term directly related to housing bubbles and not an everyday search query one would expect to yield high search volume levels during normal economic times. It will be relatively easy to both translate the term Housing Bubble and compute it for the respective countries. On the other hand, the different cultures in the specific countries will affect the way people think and therefore the way they use Google as a search engine.

For Real Estate Agent the next step is to start making an out-of-sample prediction and make a forecast of the future house prices at the state level in the United States. The relationship between in-sample prediction result of the house prices at the state level and the population in the respective state should be further investigated. We have found that the in-sample prediction result was significantly better in the states experiencing a real housing bubble, but not if this is due to the presence of animal spirits or if it is simply due to higher search volume levels or a mix. We repeatedly find that the states experiencing a real and minor bubble have significantly higher population than those states not experiencing a bubble. An exception is the non-bubble state Texas, which is the second most populated state in the United States. The correlation and in-sample prediction result for Texas were coinciding with the other non-bubble states and give reason to believe we have been able to capture the animal spirits described by Akerlof and Shiller (2005).

We started out with 204 search terms we found related to housing and animal spirits. After screening the search terms based on their correlation with the house prices in the identified bubble states, we ended up with twenty search queries. These were further tested in several in-sample bubble identification tests. Our focus, both when originally finding the 204 search terms and reducing them to twenty, was to find search terms that could measure the confidence and level of animal spirits during a housing bubble. There might be a major amount of search queries we never tested which have the same or higher correlation with the house prices and which can yield even better prediction results. Therefore, more search queries should be tested to forecast the house prices. Also, there should be done further research on whether the correlation between search terms and the house prices changes over time. Today there is far more users of the Internet and Google then in 2004 and new generations are growing up, using Google in a new way.

10 Reference List

Akerlof, G.A. and Shiller, R.J. (2009). Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism. New Jersey: Princeton University Press. ISBN 978-0-691-14233-3.

Benjamin, J., Chinloy, P. and Jud, D. (2004). "Real estate versus financial wealth in consumption". In: Journal of Real Estate Finance and Economics 29, pp. 341-354.

Bracke, P. (2013). "How long do housing cycles last? A duration analysis for 19 OECD countries". In: Journal of Housing Economics Vol. 22 No. 3, pp. 213–230.

Brunnermeier, M.K. and Oehmke, M. (2012). Bubbles, Financial Crises, and Systemic Risk. National Bureau of Economic Research, Working Paper No. 18398.

Brynjolfsson, E., Y. J. Hu, and M. S. Rahman. 2013. "Competing in the Age of Omnichannel Retailing." MIT Sloan Management Review 54 (4). http://sloanreview .mit.edu/article/competing-in-the-age-of-omnichannel-retailing/.

Campbell, J. and J. Cocco (2004), How Do Housing Price Affect Consumption? Evidence from Micro Data. Harvard Institute of Economic Research, Discussion Paper No. 2045.

Case, K.E., Quigley, J. and Shiller R. (2001). Comparing wealth effects: the stock market versus the housing market. National Bureau of Economic Research, Working Paper No. 8606.

Case, K.E. and Shiller, R.J. (2000). United States S&P Case-Shiller Home Price Index, available from: http://www.tradingeconomics.com/united-states/case-shiller-home-price-index

Case, K.E. and Shiller, R.J. (2004). Is there a Bubble in the Housing Market? Cowless foundation paper no. 10 89.

Chen, B., Schoeni, R. and Stafford, F. (2012). Mortgage Distress and Financial Liquidity: How U.S. Families are Handling Savings, Mortgages and Other Debts. Technical Series Paper 12-02, Survey Research Center, Institute for Social Research, University of Michigan.

Choi, H., and H. R. Varian. 2009. "Predicting the Present with Google Trends." Google Research Blog. http://static.googleusercontent.com/media/www.google .com/en/us/googleblogs/pdfs/google_predicting_the_present.pdf. comScore. 2012. "qSearch: A Comprehensive View of the Search Landscape." Technical Report, comScore. https://www.comscore.com/Products/Audience-Analytics /qSearch.

Damien, C. & Ahmed, B. H. A. 2013. Predicting financial markets with Google Trends and not so random keywords. arXiv preprint arXiv:1307.4643.

Engsted, T. (2014). Fama on bubbles. DEAS Working Paper Series from RePEc, 2014.

M. Ettredge, J. Gerdes, and G. Karuga. Using web-based search data to predict macroeconomic statistics. Communications of the ACM, 48(11):87–92, 2005. URL http://portal.acm.org/citation.cfm?id=1096010.

Fama, Eugene F. 2014. ”Two Pillars of Asset Pricing.” In: American Economic Review, 10 4(6), pp. 1467-1485.

Glaeser, E. L. 2008. “Housing Prices in the Three Americas.” New York Times, September 30.

Glaeser, E. L., and J. Gyourko. 2006. “Housing Dynamics.” NBER Working Paper no. 12787, Cambridge, MA.

Sharad Goel, Jake M. Hofman, S´ebastien Lahaie, David M. Pennock, and Duncan J. Watts. Predicting consumer behavior with web search. Proceedings of the National Academy of Sciences, 7(41):17486–17490, Sep 27 2010. doi: 10.1073/pnas.1005962107. URL http://www.pnas.org/content/early/2010/09/20/1005962107.

Hardaway, R. M. (2011). The Great American Housing Bubble. Santa Barbara, California: Praeger. e-Book ISBN: 9780313382291

Harding, D. and Pagan, A. (2002). ”Dissecting the cycle: a methodological investigation”. In: Journal of Monetary Economics Volume 49, Issue 2, pp. 365–381

Hawkins, Douglas M. "The problem of overfitting." Journal of chemical information and computer sciences 44.1 (2004): 1-12.

Horrigan, J. B. 2008. “The Internet and Consumer Choice: Online Americans Use Different Search and Purchase Strategies for Different Goods.” Technical Report, Pew Internet and American Life Project.

Jones, B. (2015), ”Asset Bubbles: Re-thinking Policy for the Age of Asset Management”, IMF Working Paper, Monetary and Capital Markets Department.

Kahneman, D. (2011), Thinking, Fast and Slow. New York: Farrar, Straus & Giroux. ISBN 0374275637.

Kindleberger, C, P and Aliber, R, Z. (2012) Manias, panics and crashes: a history of financial crises

Kuruzovich, J., S. Viswanathan, R. Agarwal, S. Gosain, and S. Weitzman. 2008. “Marketspace or Marketplace? Online Information Search and Channel Outcomes in Auto Retailing.” Information Systems Research 19 (2): 182–201.

Pentland, A. S. 2010. Honest Signals. Cambridge, MA: MIT Press.

Preis, T., Moat, H. S. & Stanley, H. E. 2013. Quantifying Trading Behavior in Financial Markets Using Google Trends. Sci. Rep., 3.

Preis, T., Reith, D. & Stanley, H. E. 2010. Complex dynamics of our economic life on different scales: insights from search engine query data

Lind, H. (2009). “Price bubbles in housing markets: concept, theory and indicators”. In: International Journal of Housing Markets and Analysis Vol. 2 No. 1, pp. 78-90.

New York Times (2010). Shiller’s List: How to Diagnose the Next Bubble. Published: 27th of January 2014. Available from http://dealbook.nytimes.com/2010/01/27/schillers-list-how-to-diagnose-the-next-bubble/ [3 December 2016].

Shiller, R.J. (2005). Irrational Exuberance. 2nd. New Jersey: Princeton University Press. ISBN 0-691- 12335-7.

Smith, Martha K. (2014-06-13). "Overfitting". University of Texas at Austin. Retrieved 2016-07-31

Stiglitz, J.E. (1990). “Symposium on bubbles”. In: Journal of Economic Perspectives Vol. 4 No. 2, pp. 13-18.

Taylor, A.M. (2015). ”Credit, Financial Stability, and the Macroeconomy”. Annual Review of Economics. Vol. 7: 309-339. http://pages.stern.nyu.edu/ bakos/wise/papers/wise2009-3b3paper.pdf

Elliott, Graham, Rothenberg, Thomas J., and James H. Stock. Efficient Tests for an Autoregressive Unit Root. Econometrica, 64:4, 1996, 813-836. http://cadmus.eui.eu/bitstream/handle/1814/3318/ECO2005-5.pdf

11 Appendix A

11.1 Seasonality Adjustments

Our seasonality adjustment method is described below. The method is based on the Centered Moving Average (CMA). In Equation (1) below T is the total number of observations in our data set. Each year has M periods. For quarterly data M = 4, and M = 12 for monthly data. X is simply the data point.

1. First we calculate the CMA using equation (1)

$$CMA_t = \frac{1}{2M}\sum_{i=1}^{M}\left(X_{t-\frac{M}{2}+(i-1)} + X_{t-\frac{M}{2}+i}\right)$$

2. After calculating the CMAs we first find the ratio αt = HPIt CMAt for all (remaining) observations.
3. We then find the unadjusted seasonality factor, γp = ¯ αt for each period p ∈ [1,..., M]. This is simply the average of αt in all Q1s, Q2s and so forth for quarterly data and Januarys, Februarys and so forth for monthly data.
4. The adjusted seasonality factor φp for each period is found by dividing γp by the average of all unadjusted seasonality factors, ¯ γ , i.e. φp = γp ¯ γ
5. The seasonally adjusted X is found by dividing by the adjusted seasonality factor, i.e. Xseas.adj = Xt φp

From (1) we see that M 2 +1 ≤ t ≤ T − M 2 , which implies that we lose M 2 data points at the start of the time series and M 2 at the end, in total M data points.

12 Appendix B

12.1 Search Terms

List of Alphabetically Sorted Search Terms	
A	Acres, Acres of Land, Affordable Housing, Analyst
B	Backyard, Beach Front, Broker, Bubble, Building a House, Building Cost, Buying Out
C	CBS Constructed Homes, Consumer Loans, Consumer Credit, Consumer Lending, Condos, Credit
D	Debt, Disposable Income, Down Payment, Duplex Home, Dwelling, Dwellings
E	Equity, Equity Requirement
F	Financial, Financial Analysis, First Time Homebuyer, Future Interest
G	Gated Communities, GDP
H	Home Equity, Home Equity Loan, Homes in up and Coming Communities, House Analysis,
I	Income, Income Change, Income Increase, Income Raise, Increasing Property Prices Increasing Real Estate Prices, Inflation, Installments, Interest Forecast, Interest, Interest Rate
L	Land Price, Land Prices, Leasing, Lending, Lending Standard, Low Down Payment, Low
M	Middle Class Homes, Mortgage, Mortgage Payment, Mortgage Requirements
N	Net Immigration, New Buildings, Newly Renovated, Number of Completed Homes
O	One Story Home, Overpriced, Overvaluation
P	Part Payment, Patio, Peak, Pet Approval, Pool, Pricing, Property Bubble, Property, Property Investment, Property Tax, Property Under Construction, Population
R	Raising Property, Real Estate, Real Estate Advisor, Real Estate Agent, Real Estate Bubble, Real Estate Broker, Realtor, Real Estate Listings
S	Salary Increase, Salary Change, Salary Raise, School District, Second Mortgage
T	Turmoil, Two Storey Home, Two Storey House
U	Unemployment, Unemployment Rate
V	Vacation House, Valuation
W	Wage, Wages, Wage Increase, Wage raise, Waterfront Property
Z	Zero Interest Rate

Table 12-1: The table presents the 204 search term, originally tested, sorted alphabetically.

13 Appendix C

13.1 The 50 United States Sorted After their Total Price Fall from Top to Bottom

Rank	State	3 years	5 years	Top HPI	Peak	Bottom	Trough	Price Fall
1	Nevada	65.1%	79.5%	491.2	Q1 2006	191.4	Q2 2012	-61.0%
2	Arizona	55.2%	68.8%	506.2	Q4 2006	247.4	Q3 2011	-51.1%
3	Florida	50.2%	78.4%	570.9	Q4 2006	280.4	Q2 2012	-50.9%
4	California	56.2%	84.9%	770.1	Q2 2006	402.7	Q1 2012	-47.7%
5	Michigan	3.0%	9.3%	394.5	Q2 2005	240.1	Q2 2012	-39.1%
6	Rhode Island	36.5%	72.5%	726.0	Q1 2006	448.2	Q4 2013	-38.3%
7	Maryland	42.5%	72.1%	630.2	Q4 2006	420.4	Q1 2013	-33.3%
8	Idaho	35.3%	40.0%	398.5	Q1 2007	266.7	Q2 2011	-33.1%
9	Oregon	34.2%	45.1%	533.6	Q2 2007	357.7	Q2 2012	-33.0%
10	Washington	36.1%	43.7%	580.0	Q1 2007	396.1	Q2 2012	-31.7%
11	Georgia	6.1%	9.8%	382.7	Q4 2006	262.0	Q2 2012	-31.5%
12	New Jersey	26.3%	53.2%	682.8	Q4 2006	469.2	Q4 2013	-31.3%
13	New Hampshire	21.0%	44.0%	561.8	Q1 2006	388.8	Q1 2013	-30.8%
14	Minnesota	14.8%	30.2%	442.7	Q1 2006	306.4	Q2 2012	-30.8%
15	Connecticut	25.6%	43.1%	560.8	Q1 2006	389.8	Q1 2014	-30.5%
16	Illinois	11.9%	21.5%	440.0	Q4 2006	306.1	Q1 2013	-30.4%
17	Delaware	28.7%	47.6%	591.9	Q4 2006	420.2	Q1 2014	-29.0%
18	Massachusetts	24.4%	50.2%	880.5	Q2 2005	628.6	Q4 2012	-28.6%
19	Ohio	2.8%	7.1%	328.2	Q2 2005	241.4	Q1 2014	-26.4%
20	Hawaii	46.5%	78.9%	631.3	Q1 2007	466.4	Q2 2012	-26.1%
21	Virginia	34.9%	54.9%	552.1	Q4 2006	408.0	Q2 2012	-26.1%
22	New Mexico	26.6%	34.1%	382.6	Q1 2007	288.6	Q1 2014	-24.6%
23	Utah	30.4%	30.2%	439.9	Q3 2007	333.4	Q4 2003	-24.2%
24	New York	19.8%	42.0%	760.4	Q4 2006	577.9	Q1 2014	-24.0%
25	Maine	15.6%	34.9%	600.6	Q4 2006	458.3	Q1 2014	-23.7%
26	Wisconsin	12.7%	18.6%	387.0	Q1 2006	297.3	Q1 2014	-23.2%
27	Missouri	6.7%	13.5%	351.3	Q4 2006	275.8	Q1 2014	-21.5%
28	South Carolina	12.7%	15.4%	395.0	Q4 2006	310.5	Q1 2014	-21.4%
29	North Carolina	10.2%	12.7%	387.6	Q2 2007	310.0	Q4 2013	-20.0%
30	Alabama	11.5%	14.5%	349.1	Q2 2007	280.7	Q4 2013	-19.6%
31	Mississippi	11.9%	13.4%	301.6	Q1 2007	243.7	Q4 2013	-19.2%

32	Pennsylvania	19.3%	32.0%	463.4	Q4 2006	375.7	Q1 2014	-18.9%
33	Indiana	1.5%	4.8%	306.5	Q2 2005	249.8	Q1 2014	-18.5%
34	Colorado	14.2%	21.7%	427.9	Q4 2006	349.5	Q1 2012	-18.3%
35	Vermont	23.5%	40.5%	533.9	Q4 2006	440.3	Q1 2014	-17.5%
36	Tennessee	9.7%	12.8%	350.6	Q2 2007	292.1	Q1 2013	-16.7%
37	Montana	20.8%	35.8%	431.5	Q3 2007	363.1	Q2 2012	-15.9%
38	Arkansas	9.3%	13.6%	299.2	Q1 2007	252.1	Q2 2012	-15.7%
39	West Virginia	13.0%	16.8%	259.5	Q4 2006	219.0	Q1 2013	-15.6%
40	Kentucky	3.3%	6.6%	340.4	Q4 2006	292.2	Q1 2014	-14.1%
41	Kansas	2.6%	6.3%	280.6	Q4 2006	241.9	Q1 2014	-13.8%
42	Nebraska	4.7%	7.4%	302.5	Q2 2005	262.2	Q4 2012	-13.3%
43	Wyoming	24.2%	38.6%	323.8	Q3 2007	281.9	Q1 2012	-13.0%
44	Louisiana	15.2%	21.1%	284.2	Q1 2007	251.4	Q1 2013	-11.5%
45	Alaska	22.1%	31.5%	332.6	Q1 2007	294.7	Q2 2012	-11.4%
46	Texas	6.0%	10.0%	257.5	Q2 2007	232.7	Q1 2012	-9.6%
47	Iowa	5.4%	9.6%	289.8	Q2 2005	270.0	Q3 2008	-6.8%
48	South Dakota	3.9%	6.8%	331.1	Q1 2007	309.1	Q3 2012	-6.6%
49	Oklahoma	2.5%	5.2%	231.8	Q1 2007	222.6	Q3 2008	-4.0%
50	North Dakota	12.4%	19.4%	280.0	Q1 2007	271.6	Q3 2008	-3.0%

Table 13-1: The table display the fifty states sorted according to their price fall from the peak to the trough."3 years" and "5 years" is the percentage price increase the last three and five years before the price top in each respective state. "Top HPI" and "Bottom" is the highest and lowest value for the Housing Price index (HPI) in each respective state. "Peak" and "Trough" is the quarter and year for the highest and lowest value of HPI. "Price Fall" is the percentage price fall from peak to trough in each respective state.

14 Appendix D

14.1 Stationarity Test of the Variables at Level for all 50 States

Country General Variables	**ln CCI**	**ln IR**	**ln DPI**
t-statistics	-1.493	-0.843	-1.234

Table 14-1: The table shows the results from the Dickey-Fuller Generalized Least Square unit root test of the following time series at Level. The natural logarithm to Consumer Confidence Index (CCI), 1 + Interest Rate in percentage (IR) and Disposable Personal Income (DPI). The three time-series are all general for the United States.

State Specific Variables	**ln HPI**	**ln UR**	**ln DPO**	**ln HPA**	**ln GSVI**
State Name	t-statistics	t-statistics	t-statistics	t-statistics	t-statistics
Nevada	-1.4	-3.1 *	-2.3	-1.4	-1.2
Arizona	-1.5	-3.9 ***	-2.8	-1.1	-0.9
Florida	-1.3	-2.9	-2	-1	-0.9
California	-1.7	-3.2 **	-2.2	-0.9	-1.4
Maryland	-1.3	-2.3	-2.1	-1.7	-1.4
Idaho	-1.4	-1.3	-1.7	-1.9	-1.7
Oregon	-1.4	-1.3	-1.7	-1.9	-1.4
Washington	-1.4	-3.3 **	-2.1	-1.8	-1.0
Hawaii	-1.3	-2.7	-2.2	-1.5	-5.2 ***
Virginia	-1.3	-2.1	-2.2	-1.2	-1.6
Rhode Island	-0.9	-2.6	-2.5	-1.5	-1.9
Michigan	-0.8	-2.2	-1.6	-1.7	-2.4 **
Georgia	-1.0	-1.8	-1.2	-1.0	-0.6
New Jersey	-1.1	-2.6	-2.1	-1.5	-1.0
New Hampshire	-0.8	-2.5	-2.1	-2.2 *	-4.5 ***
Minnesota	-0.9	-2.2	-3.7 **	-2.4 **	-1.3
Connecticut	-0.5	-2.1	-2.7	-1.8	-1.5
Illinois	-0.7	-1.9	-3.3 **	-1.2	-1.5
Delaware	-1.0	-2.7	-2.7	-1.6	-0.5
Massachusetts	-1.0	-2.8	-2.2	-1.6	-1.6
Ohio	-0.6	-2.1	-2.5	-1.8	-1.1
New Mexico	-1.1	-2.8	-2.5	-1.0	-1.0
Utah	-1.5	-2.1	-3.8 ***	-1.9	-1.9
New York	-1.1	-2.6	-2.8	-2.3 **	-1.6
Maine	-1.0	-2.3	-2.8	-2.5 **	-2.1 *

Wisconsin	-0.8	-2.5		-2.4		-2.3	**	-0.8	
Missouri	-0.9	-2.8		-1.8		-1.4		-0.9	
South Carolina	-1.3	-2.1		-3.2	**	-1.4		-1.2	
Alabama	-1.0	-2.5		-3.2	**	-0.9		-1.0	
Mississippi	-0.9	-1.6		-3.4	**	-1.3		-2.5	**
Pennsylvania	-1.2	-2.1		-2.4		-1.7		-2.2	*
Indiana	-0.7	-1.9		-3.0	*	-1.9		-2.0	*
Colorado	-0.6	-2.5		-3.6	**	-1.4		-1.7	
Vermont	-1.0	-1.9		-4.2	***	-3.2	***	-2.3	**
Tennessee	-1.2	-2.1		-1.7		-1.1		-1.1	
Montana	-0.8	-2.1		-3.0	*	-2.9	***	-4.4	***
Arkansas	-1.1	-2.2		-3.1	*	-1.9		-1.6	
West Virginia	-1.1	-2.3		-3.0	*	-1.8		-3.4	***
Kentucky	-1.0	-2.1		-3.4	**	-1.6		-2.4	**
Kansas	-1.0	-2.5		-2.8		-1.8		-1.6	
Nebraska	-1.1	-2.4		-2.2		-3.1	***	-2.0	*
Wyoming	-0.8	-2.7		-2.6		-3.8	***	-1.8	
Louisiana	-1.2	-2.7		-2.4		-1.8		-1.3	
Alaska	-0.9	-2.5		-2.7		-3.9	***	-2.5	**
Texas	0.0	-2.6		-2.2		-1.6		-1.0	
Iowa	-1.2	-2.8		-2.2		-3.4	***	-3.2	***
South Dakota	-0.3	-2.5		-2.0		-4.9	***	-2.1	*
Oklahoma	-1.3	-3.5	***	-2.6		-1.7		-0.7	
North Dakota	-1.4	-2.6		-2.6		-3.0	***	-5.1	***

Table 14-2: The table show the result from the Dickey-Fuller Generalized Least Square (DF-GLS) unit root test of the following time series in Level. The natural logarithm to Housing Price Index (HPI), 1+Unemployment Rate in percentage (UR), Housing Permits Authorized (HPA), Population (PO) and Google Search Volume Index (GSVI). The five time-series are state specific for each of the 50 states. The DF-GLS tests are performed with the No Trend option for all variables except Unemployment Rate.

14.2 Stationarity Test of the First Differenced Variables for all 50 States

Country General Variables	**∆ ln CCI**		**∆ ln IR**		**∆ ln DPI**	
t-statistics	-5.333	***	-3.557	***	-4.968	***

Table 14-3: The table shows the results from the Dickey-Fuller Generalized Least Square unit root test of the following first differenced time series. The natural logarithm to Consumer Confidence Index (CCI), 1 + Interest Rate in percentage (IR) and Disposable Personal Income (DPI). The three time-series are general for the United States. The tests are performed with the no trend option for all variables.

State Specific Variables	**∆ ln HPI**		**∆ ln UR**		**∆ ln DPO**		**∆ ln HPA**		**∆ ln GSVI**	
States	t-statistics		t-statistics		t-statistics		t-statistics		t-statistics	
Nevada	-1.8		-1.6		-5.1	***	-6.5	***	-8.4	***
Arizona	-2	*	-2	*	-5.6	***	-4.2	***	-5.5	***
Florida	-2.1	*	-1.9	*	-5.0	***	-3.2	***	-5.1	***
California	-1.9	*	-1.6		-2.6	***	-3.4	***	-4.9	***
Maryland	-2.1	*	-2.7	***	-2.4	***	-5.8	***	-3.8	***
Idaho	-2.9	***	-2.4	**	-5.0	***	-5.1	***	-3.9	***
Oregon	-2.9	***	-2.4	**	-5.0	***	-5.1	***	-8.0	***
Washington	-2.3	**	-2.8	***	-2.5	***	-5.0	***	-4.9	***
Hawaii	-1.9	*	-2.6	***	-2.4	***	-6.6	***	-9.3	***
Virginia	-2.6	***	-2.7	***	-2.4	***	-5.5	***	-4.6	***
Rhode Island	-2.5	**	-2.7	***	-2.72		-5.0	***	-6.0	***
Michigan	-4.4	***	-3.6	***	-5.0	***	-5.9	***	-1.2	
Georgia	-4.0	***	-2.1	*	-5.2	***	-6.7	***	-3.5	***
New Jersey	-2.6	***	-2.8	***	-3.1	***	-4.6	***	-1.6	
New Hampshire	-3.1	***	-3.6	***	-4.6	***	-5.5	***	-2.8	***
Minnesota	-4.6	***	-2.9	***	-3.5	***	-3.9	***	-6.6	***
Connecticut	-2.9	***	-2.5	***	-3.5	***	-4.7	***	-4.1	***
Illinois	-3.2	***	-3.1	***	-4.9	***	-4.4	***	-3.7	***
Delaware	-2.7	***	-3.0	*	-4.9	***	-8.3	***	-0.8	
Massachusetts	-3.1	***	-2.2	*	-2.9	*	-3.0	***	-3.4	***
Ohio	-5.4	***	-3.0	***	-2.9	*	-4.7	***	-5.9	***
New Mexico	-2.9	***	-3.9	***	-4.7	***	-6.4	***	-7.9	***
Utah	-3.1	***	-2.9	***	-6.6	***	-5.1	***	-4.5	***
New York	-3.1	***	-2.6	**	-3.2	**	-3.9	***	-1.4	
Maine	-3.0	***	-3.6	***	-5.0	***	-3.8	***	-2	*
Wisconsin	-3.7	***	-3.6	***	-4.7	***	-5.1	***	-2.3	**

Missouri	-4.3	***	-2.4	**	-5.0	***	-5.6	***	-4.5	***
South Carolina	-4.1	***	-3.2	***	-2.9	*	-5.2	***	-2.5	**
Alabama	-4.1	***	-3.3	***	-2.9	*	-8.0	***	-4.0	***
Mississippi	-4.1	***	-3.3	***	-4.9	***	-9.1	***	-7.3	***
Pennsylvania	-3.2	***	-3.0	***	-3.0	*	-5.2	***	-4.5	***
Indiana	-5.9	***	-3.3	***	-2.9	***	-4.9	***	-2.2	***
Colorado	-3.6	***	-1.8		-3.0	*	-4.8	***	-7.5	***
Vermont	-2.9	***	-4.0	***	-6.5	***	-4.4	***	-7.3	***
Tennessee	-4.0	***	-3.8	***	-5.0	***	-5.7	***	-6.1	***
Montana	-3.0	***	-2.9	***	-3.5	***	-4.6	***	-5.9	***
Arkansas	-3.6	***	-3.8	***	-7.2	***	-4.5	***	-4.6	***
West Virginia	-4.4	***	-4.5	***	-3.3	**	-6.2	***	-1.9	
Kentucky	-5.1	***	-3.3	***	-4.9	***	-7.3	***	-6.6	***
Kansas	-4.9	***	-2.6	***	-3.5	**	-5.6	***	-9.7	***
Nebraska	-4.6	***	-3.0	***	-2.4		-3.9	***	-8.4	***
Wyoming	-2.6	***	-3.6	***	-2.9	*	-5.9	***	-8.2	***
Louisiana	-4.0	***	-5.7	***	-3.5	***	-4.1	***	-4.6	***
Alaska	-3.4	***	-3.2	***	-3.7	***	-4.2	***	-3.9	***
Texas	-3.5	***	-2.6	***	-2.6		-4.8	***	-4.8	***
Iowa	-5.0	***	-3.2	***	-2.4		-4.4	***	-10.8	***
South Dakota	-4.2	***	-3.6	***	-2.4		-9.3	***	-4.3	***
Oklahoma	-5.0	***	-3.5	***	-3.2	**	-7.5	***	-2.0	*
North Dakota	-4.3	***	-4.9	***	-2.7		-3.7	***	-10.0	***

Table 14-4: The table shows the results from the Dickey-Fuller Generalized Least Square (DF-GLS) unit root test of the first difference of the following time series. The tests are performed with the No Trend option for all variables. The natural logarithm to Housing Price Index (HPI), 1+Unemployment Rate in percentage (UR). Housing Permits Authorized (HPA), Population (PO) and Google Search Volume Index (GSVI). The five time-series are state specific for each of the fifty states. The DF-GLS tests are performed with the No Trend option for all variables except Unemployment Rate.

15 Appendix E

Tests of Cointegration

15.1 Test of Cointegration among all Variables for all 50 States

Johansen Cointegration Test						
Maximum Rank	**Standard Model**		**CCI Model**		**GSVI Model**	
	5% Critical Value		**5% Critical Value**		**5% Critical Value**	
0	94.15		124.24		124.24	
1	68.52		94.15		94.15	
2	47.21		68.52		68.52	
3	29.68		47.21		47.21	
4	15.41		29.68		29.68	
5	3.76		15.41		15.41	
6			3.76		3.76	
State Name	**No of CR**	**Trace Statistics**	**No of CR**	**Trace Statistics**	**No of CR**	**Trace Statistics**
Nevada	2	40.3627	3	43.7198	3	41.3831
Arizona	3	23.9541	3	45.1838	4	23.6582
Florida	3	28.8519	3	44.5998	4	28.8787
California	3	25.6202	5	10.3180	5	10.3595
Maryland	2	45.2226	4	24.8237	3	43.2934
Idaho	3	16.3943	3	41.7381	4	16.1368
Oregon	3	16.3943	3	16.3943	4	16.1368
Washington	3	17.6006	4	15.6217	3	40.8641
Hawaii	3	22.5517	3	45.0867	4	22.5375
Virginia	3	23.9379	4	24.3824	3	34.4096
Rhode Island *	3	24.4648	4	19.0211	4	22.7024
Michigan	2	31.6738	3	39.7405	3	31.0343
Georgia	2	43.4937	3	44.6179	3	40.4820
New Jersey	3	23.9767	4	18.7800	4	21.8624
New Hampshire *	2	35.5841	3	41.7453	3	35.5357
Minnesota	2	42.2711	3	33.3372	3	45.0350
Connecticut	2	46.2505	3	46.1145	3	44.0273
Illinois	2	32.5926	3	31.6180	3	28.6395
Delaware	2	37.3162	2	64.4077	3	40.5506
Massachusetts	2	41.3113	3	39.4561	3	29.6629
Ohio	1	66.0071	2	60.5332	2	53.4324
New Mexico	2	40.6647	3	41.9667	3	36.4998

Utah	2	45.7809	3	43.7825	4	22.8082
New York	2	44.4772	3	31.2727	3	43.4184
Maine *	2	39.8359	3	30.2518	3	39.7884
Wisconsin	2	35.5670	3	36.8299	2	61.2751
Missouri	2	42.3427	2	67.6795	3	37.6629
South Carolina	3	22.2126	3	44.4371	4	22.9054
Alabama	2	42.3577	2	67.0684	2	67.5585
Mississippi	2	41.8859	2	62.5591	3	39.9526
Pennsylvania	3	25.7204	4	25.0029	4	25.4845
Indiana	2	41.9070	3	45.1118	3	43.1139
Colorado	2	31.3916	3	38.1430	3	31.0629
Vermont	3	22.4250	3	40.7536	4	22.2068
Tennessee	1	68.1641	3	37.7474	2	61.8224
Montana	4	8.8941	4	27.6188	5	8.7579
Arkansas	3	23.3949	3	45.6645	2	63.6456
West Virginia	2	45.6584	3	47.1587	3	42.4379
Kentucky	1	61.5666	1	93.5387	1	93.6420
Kansas	2	45.8627	3	42.8021	3	43.0793
Nebraska	2	26.7690	2	63.7105	2	63.6124
Wyoming	3	20.0871	4	22.6715	4	19.8079
Louisiana	1	56.0549	1	91.7284	2	53.4080
Alaska	2	39.9938	3	42.5329	2	67.8711
Texas	2	33.3084	3	32.9837	2	66.0308
Iowa	2	30.7638	3	32.0965	2	65.6164
South Dakota	1	67.6594	3	42.0215	2	65.7291
Oklahoma *	1	43.2096	2	38.1231	2	38.9640
North Dakota	1	60.7251	1	88.9476	2	62.0380

Table 15-1: The table shows the result from the Cointegration test implemented by vecrank in Stata, which is based on Johansen`s method. The test check if there is one or more cointegrating relationships among variables in the three models Standard, CCI and GSVI. The Standard Model consist of the variables Housing Price Index (HPI), Unemployment Rate (UR), Interest Rate (IR), Housing Permits Authorized, Population (PO) and Disposable Personal Income (DPI). The CCI Model includes the same variables in addition to the Consumer Confidence Index (CCI). The GSVI Model includes the same variables as the Standard Model in addition to the Google Search Volume Index (GSVI). All three models are tested with only one lag. The null hypothesis are that there are Maximum Rank (0, 1, 2, …., n-1, where n is number of variables in the model) cointegrating relationships among variables. * Indicates collinearity in the model in the specific state. The Stata function **noreduce** ***have been used on these models.*** **Noreduce** ***do not perform checks and corrections for collinearity among lags of dependent variables.***

15.2 Test of Cointegration among Housing Price Index and Google Search Volume Index for Real Estate Agent in all 50 States

Johansen Cointegration Test of Housing Price Index and Google Search Volume Index					
State Name	No of CE	Trace Statistics	5% Critical Value	Max Statistics	5% Critical Value
Nevada	1	0.0661 **	3.76	0.0661	3.76
Arizona	1	0.8579 **	3.76	0.8579	3.76
Florida	1	1.3191 **	3.76	1.3191	3.76
California	1	1.2571 **	3.76	1.2571	3.76
Maryland	1	1.2694 **	3.76	1.2694	3.76
Idaho	1	0.6684 **	3.76	0.6684	3.76
Oregon	0	13.1988	15.41	13.1988	15.41
Washington	1	1.1240 **	3.76	1.1240	3.76
Hawaii	1	3.5763 **	3.76	3.5763	3.76
Virginia	1	2.0193 **	3.76	2.0193	3.76
Rhode Island	1	0.5224 **	3.76	0.5224	3.76
Michigan	1	2.5222 **	3.76	2.5222	3.76
Georgia	1	1.2616 **	3.76	1.2616	3.76
New Jersey	1	0.3889 **	3.76	0.3889	3.76
New Hampshire	1	0.4260 **	3.76	0.4260	3.76
Minnesota	1	1.0042 **	3.76	1.0042	3.76
Connecticut	1	0.0656 **	3.76	0.0656	3.76
Illinois	1	0.8502 **	3.76	0.8502	3.76
Delaware	1	0.1084 **	3.76	0.1084	3.76
Massachusetts	1	1.0299 **	3.76	1.0299	3.76
Ohio	1	3.0872 **	3.76	3.0872	3.76
New Mexico	1	0.3967 **	3.76	0.3967	3.76
Utah	1	1.1838 **	3.76	1.1838	3.76
New York	1	1.0557 **	3.76	1.0557	3.76
Maine	1	0.4965 **	3.76	0.4965	3.76
Wisconsin	1	0.8712 **	3.76	0.8712	3.76
Missouri	1	1.2948 **	3.76	1.2948	3.76
South Carolina	1	0.7458 **	3.76	0.7458	3.76
Alabama	1	1.2938 **	3.76	1.2938	3.76
Mississippi	1	0.3842 **	3.76	0.3842	3.76
Pennsylvania	1	1.0007 **	3.76	1.0007	3.76
Indiana	1	2.1225 **	3.76	2.1225	3.76
Colorado	1	1.3479 **	3.76	1.3479	3.76
Vermont	1	1.7154 **	3.76	1.7154	3.76

Tennessee	1	0.5677 **	3.76	0.5677	3.76
Montana	1	3.9230	3.76	3.9230	3.76
Arkansas	0	14.8727	15.41	13.1080	14.07
West Virginia	1	0.7659 **	3.76	0.7659	3.76
Kentucky	1	0.9502 **	3.76	0.9502	3.76
Kansas	1	1.1206 **	3.76	1.1206	3.76
Nebraska	1	0.8089 **	3.76	0.8089	3.76
Wyoming	0	8.1668	3.76	8.1668	3.76
Louisiana	1	1.7536 **	3.76	1.7536	3.76
Alaska	0	7.5477	3.76	7.5477	3.76
Texas	0	8.7813	15.41	5.1980	14.07
Iowa	1	1.0014 **	3.76	1.0014	3.76
South Dakota	1	0.1783 **	3.76	0.1783	3.76
Oklahoma	1	0.7905 **	3.76	0.7905	3.76
North Dakota	1	0.8032 **	3.76	0.8032	3.76

***Table 15-2: The table shows the result from the Cointegration test implemented by vecrank in Stata, which is based on Johansen`s method. The test check if there is Cointegration between the Housing Price Index time-series and the Google Search Volume Index time*-series, individually, in each of the 50 states. The null hypothesis are that there are Maximum Rank (0 or 1) cointegrating relationships among variables. ** = 5% significance level for one cointegrating relationship among variables**

16 Appendix F

Identified Bubble States and Ranking

16.1 Identified Bubble States and Ranked after their Total Price Fall

Rank	State	3 Years Increase	5 Years Increase	Top HPI	Date Top	Bottom HPI	Date Bottom	Price Fall
1	Nevada	65.1%	79.5%	491.2	Q1 2006	191.4	Q2 2012	-61.0%
2	Arizona	55.2%	68.8%	506.2	Q4 2006	247.4	Q3 2011	-51.1%
3	Florida	50.2%	78.4%	570.9	Q4 2006	280.4	Q2 2012	-50.9%
4	California	56.2%	84.9%	770.1	Q2 2006	402.7	Q1 2012	-47.7%
5	Maryland	42.5%	72.1%	630.2	Q4 2006	420.4	Q1 2013	-33.3%
6	Oregon	34.2%	45.1%	533.6	Q2 2007	357.7	Q2 2012	-33.0%
7	Washington	36.1%	43.7%	580.0	Q1 2007	396.1	Q2 2012	-31.7%
8	New Jersey	26.3%	53.2%	682.8	Q4 2006	469.2	Q4 2013	-31.3%
9	Connecticut	25.6%	43.1%	560.8	Q1 2006	389.8	Q1 2014	-30.5%
10	Virginia	34.9%	54.9%	552.1	Q4 2006	408.0	Q2 2012	-26.1%

Table 16-1: The table lists the identified bubble states. The states are ranked based on the total price fall after the Housing Price Index (HPI) peak. The top four states are defined as real bubble states while the following six are defined as minor bubble states. "3 Years Increase" shows the three-year HPI increase before the top. "5 Years Increase" shows the five-year HPI increase before the top. "Top HPI" shows the Housing Price Index (HPI) peak, and "Date top" shows when the HPI peaked. " HPI Bottom" shows the lowest HPI value after the peak, and "Date bottom" shows when the HPI was lowest. "Total fall" shows the total HPI fall from peak to bottom.

16.2 Identified Non-bubble States and Ranked after their Total Price Fall

Rank	State	3 Years Increase	5 Years Increase	Top HPI	Date Top	Bottom HPI	Date Bottom	Price Fall
1	North Dakota	12.4%	19.4%	280.0	Q1 2007	271.6	Q3 2008	-3.0%
2	Oklahoma	2.5%	5.2%	231.8	Q1 2007	222.6	Q3 2008	-4.0%
3	South Dakota	3.9%	6.8%	331.1	Q1 2007	309.1	Q3 2012	-6.6%
4	Iowa	5.4%	9.6%	289.8	Q2 2005	270.0	Q3 2008	-6.8%
5	Texas	6.0%	10.0%	257.5	Q2 2007	232.7	Q1 2012	-9.6%
6	Alaska	22.1%	31.5%	332.6	Q1 2007	294.7	Q2 2012	-11.4%
7	Louisiana	15.2%	21.1%	284.2	Q1 2007	251.4	Q1 2013	-11.5%
8	Wyoming	24.2%	38.6%	323.8	Q3 2007	281.9	Q1 2012	-13.0%
9	Nebraska	4.7%	7.4%	302.5	Q2 2005	262.2	Q4 2012	-13.3%
10	Kansas	2.6%	6.3%	280.6	Q4 2006	241.9	Q1 2014	-13.8%

Table 16-2: The table lists the ten defined non-bubble states. The states are ranked based on the total price fall after the Housing Price Index (HPI) peak. These are the states that experienced the least, if any, decrease in house prices during the U.S. real estate bubble in 06/07. "3 Years Increase" shows the three-year HPI increase before the top. "5 Years Increase" shows the five-year HPI increase before the top. "Top HPI" shows the Housing Price Index (HPI) peak, and "Date top" shows when the HPI peaked. " HPI Bottom" shows the lowest HPI value after the peak, and "Date bottom" shows when the HPI was lowest. "Total fall" shows the total HPI fall from peak to bottom.

17 Appendix G

Tests of Google Search Volume Indexes

17.1 Test of Average Google Search Volume Indexes

17.1.1 Average Google Search Volume for all 20 Search Terms

One in a Row Red Flag Test													
M Value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.25	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	0	0	0	0	0	0	0	0	0	/ 4	3
# 6 MBS	6	5	3	2	0	0	0	0	0	0	0	0	/ 6	1
# 10 NBS	6	3	0	0	0	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	32	19	2	0	0	0	0	0	0	0	0	0	/ 40	0
Score	0	8	15	2	0	0	0	0	0	0	0	0	Max	15

Two in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	1	0	0	0	0	0	0	0	0	/ 4	3
# 6 MBS	5	3	2	0	0	0	0	0	0	0	0	/ 6	1
# 10 NBS	6	1	0	0	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	28	8	0	0	0	0	0	0	0	0	0	/ 40	0
Score	-1	12	5	0	0	0	0	0	0	0	0	Best score	12

Three in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	1	0	0	0	0	0	0	0	0	/ 4	3
# 6 MBS	0	3	2	0	0	0	0	0	0	0	0	/ 6	1
# 10 NBS	6	1	0	0	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	28	8	0	0	0	0	0	0	0	0	0	/ 40	0
Score	-1	12	5	0	0	0	0	0	0	0	0	Best score	12

Eight in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	2	0	0	0	0	0	0	0	0	0	/ 4	3
# 6 MBS	4	2	0	0	0	0	0	0	0	0	0	/ 6	1
# 10 NBS	1	0	0	0	0	0	0	0	0	0	0	/ 10	-3
# 30 NBS	8	0	0	0	0	0	0	0	0	0	0	/ 40	0
Score	13	8	0	0	0	0	0	0	0	0	0	Best score	13

Table 17-1: The table shows the result from four in-sample bubble identification tests. An index (Index20) consisting of the Average Google Search Volume Index (GSVI) for all 20 search terms are used to flag a state as a bubble state if the respective state has GSVI levels above a multiplier (M) times a normal level one time, two in a row, three in a row and eight in a row. The test has been performed for numerous values of M as seen in "M Value". "# 4 RBS" stands for how many, of the four; Real Bubble States (RBS) have been detected. "# 6 MBS" stands for how many, of the six; Minor Bubble States (MBS) has been detected. "# 10 NBS" stands for how many, of the ten; Non-bubble States (NBS) has wrongly been detected. "# 30 NDS" stands for how many, of the thirty; states not defined as neither bubble nor non-bubble states are detected. The index receives 3 points for identifying a real bubble state, 1 point for identifying a minor bubble state and are deducted 3 points for wrongly identifying a non-bubble state as a bubble state. This test is performed for each value of the multiplier M and the highest value are called "max" and placed in the bottom right cell.

17.1.2 Average Google Search Volume for the 12 Single Best Search Terms

One in a Row Red Flag Test												
M Value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of Total	Points
# 4 RBS	4	4	4	4	4	4	4	1	0	0	0	/ 4	3
# 6 MBS	6	6	6	5	3	3	2	1	0	0	0	/ 6	1
# 10 NBS	9	8	7	3	0	0	0	0	0	0	0	/ 10	-3
# 30 NBS	39	37	33	24	8	1	0	0	0	0	0	/ 40	0
Score	-9	-6	-3	8	15	15	14	4	0	0	0	Max	15
Two in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	2	0	0	0	0	0	/ 4	3
# 6 MBS	6	6	5	3	3	1	0	0	0	0	0	/ 6	1
# 10 NBS	8	8	3	1	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	38	34	24	10	0	0	0	0	0	0	0	/ 40	0
Score	-6	-6	8	12	15	7	0	0	0	0	0	Best score	15
Three in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	2	0	0	0	0	0	/ 4	3
# 6 MBS	6	6	5	3	3	1	0	0	0	0	0	/ 6	1
# 10 NBS	8	8	3	1	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	38	34	23	10	0	0	0	0	0	0	0	/ 40	0
Score	-6	-6	8	12	15	7	0	0	0	0	0	Best score	15
Eight in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	3	0	0	0	0	0	0	0	/ 4	3
# 6 MBS	5	5	3	2	1	0	0	0	0	0	0	/ 6	1
# 10 NBS	5	3	1	0	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	28	19	5	1	0	0	0	0	0	0	0	/ 40	0
Score	2	8	12	11	1	0	0	0	0	0	0	Best score	12

Table 17-2: The table shows the result from four in-sample bubble identification tests. An index (Index12) consisting of the Average Google Search Volume Index (GSVI) for the 12 single best search terms are used to flag a state as a bubble state if the respective state has GSVI levels above a multiplier (M) times a normal level one time, two in a row, three in a row and eight in a row. The test has been performed for numerous values of M as seen in "M Value". "# 4 RBS" stands for how many, of the four; Real Bubble States (RBS) have been detected. "# 6 MBS" stands for how many, of the six; Minor Bubble States (MBS) has been detected. "# 10 NBS" stands for how many, of the ten; Non-bubble States (NBS) has wrongly been detected. "# 30 NDS" stands for how many, of the thirty; states not defined as neither bubble nor non-bubble states are detected. The index receives 3 points for identifying a real bubble state, 1 point for identifying a minor bubble state and are deducted 3 points for wrongly identifying a non-bubble state as a bubble state. This test is performed for each value of the multiplier M and the highest value are called "max" and placed in the bottom right cell

17.1.3 Average Google Search Volume Index for the 6 Single Best Search Terms

One in a Row Red Flag Test												
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	4	4	3	0	0	0	/ 4	3
# 6 MBS	6	6	6	6	4	3	3	2	1	0	0	/ 6	1
# 10 NBS	8	8	5	4	3	0	0	0	0	0	0	/ 10	-3
# 30 NDS	38	36	32	27	21	10	2	0	0	0	0	/ 40	0
Score	-6	-6	3	6	7	15	15	11	1	0	0	Best score	15
Two in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	4	4	3	1	0	0	/ 4	3
# 6 MBS	6	6	6	6	5	6	3	3	1	0	0	/ 6	1
# 10 NBS	9	8	8	6	4	3	1	1	0	0	0	/ 10	-3
# 30 NDS	39	37	37	33	31	20	11	4	0	0	0	/ 40	0
Score	-9	-6	-6	0	5	9	12	9	4	0	0	Best score	12
Three in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	4	2	1	0	0	0	/ 4	3
# 6 MBS	6	6	6	4	3	1	1	0	0	0	0	/ 6	1
# 10 NBS	7	7	4	3	1	0	0	0	0	0	0	/ 10	-3
# 30 NDS	36	34	26	17	6	1	0	0	0	0	0	/ 40	0
Score	-3	-3	6	7	12	13	7	3	0	0	0	Best score	13
Eight in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	4	2	1	0	0	0	/ 4	3
# 6 MBS	6	6	6	4	3	1	1	0	0	0	0	/ 6	1
# 10 NBS	7	6	4	3	1	0	0	0	0	0	0	/ 10	-3
# 40 non-BS	36	33	26	16	6	1	0	0	0	0	0	/ 40	0
Score	-3	0	6	7	12	13	7	3	0	0	0	Best score	13

Table 17-3: The table shows the result from four in-sample bubble identification tests. An index (Index12) consisting of the Average Google Search Volume Index (GSVI) for the 12 single best search terms are used to flag a state as a bubble state if the respective state has GSVI levels above a multiplier (M) times a normal level one time, two in a row, three in a row and eight in a row. The test has been performed for numerous values of M as seen in "M Value". "# 4 RBS" stands for how many, of the four; Real Bubble States (RBS) have been detected. "# 6 MBS" stands for how many, of the six; Minor Bubble States (MBS) has been detected. "# 10 NBS" stands for how many, of the ten; Non-bubble States (NBS) has wrongly been detected. "# 30 NDS" stands for how many, of the thirty; states not defined as neither bubble nor non-bubble states are detected. The index receives 3 points for identifying a real bubble state, 1 point for identifying a minor bubble state and are deducted 3 points for wrongly identifying a non-bubble state as a bubble state. This test is performed for each value of the multiplier M and the highest value are called "max" and placed in the bottom right cell.

17.1.4 Google Search Volume Index for the 3 Single Best Search Terms

One in a Row Red Flag Test												
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	4	2	1	0	0	0	/ 4	3
# 6 MBS	6	6	6	6	5	3	1	1	0	0	0	/ 6	1
# 10 NBS	7	8	7	5	3	1	0	0	0	0	0	/ 10	-3
# 30 NDS	37	37	35	29	19	9	2	0	0	0	0	/ 40	0
Score	-3	-6	-3	3	8	12	7	4	0	0	0	Best score	12
Two in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	4	4	2	1	0	0	0	/ 4	3
# 6 MBS	6	6	6	6	5	3	1	1	0	0	0	/ 6	1
# 10 NBS	7	8	7	5	3	1	0	0	0	0	0	/ 10	-3
# 30 NDS	37	37	35	29	19	9	2	0	0	0	0	/ 40	0
Score	-3	-6	-3	3	8	12	7	4	0	0	0	Best score	12
Three in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	2	0	0	0	0	0	0	/ 4	3
# 6 MBS	6	5	3	2	1	0	0	0	0	0	0	/ 6	1
# 10 NBS	4	3	2	0	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	27	18	8	1	0	0	0	0	0	0	0	/ 40	0
Score	6	8	9	14	7	0	0	0	0	0	0	Best score	14
Eight in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	Of total	Points
# 4 RBS	4	4	4	4	2	1	1	0	0	0	0	/ 4	3
# 6 MBS	6	6	5	4	2	2	0	0	0	0	0	/ 6	1
# 10 NBS	6	4	2	2	0	0	0	0	0	0	0	/ 10	-3
# 30 NDS	31	27	17	7	0	0	0	0	0	0	0	/ 40	0
Score	0	6	11	10	8	5	3	0	0	0	0	Best score	11

Table 17-4: The table shows the result from four in-sample bubble identification tests. An index (Index12) consisting of the Average Google Search Volume Index (GSVI) for the 12 single best search terms are used to flag a state as a bubble state if the respective state has GSVI levels above a multiplier (M) times a normal level one time, two in a row, three in a row and eight in a row. The test has been performed for numerous values of M as seen in "M Value". "# 4 RBS" stands for how many, of the four; Real Bubble States (RBS) have been detected. "# 6 MBS" stands for how many, of the six; Minor Bubble States (MBS) has been detected. "# 10 NBS" stands for how many, of the ten; Non-bubble States (NBS) has wrongly been detected. "# 30 NDS" stands for how many, of the thirty; states not defined as neither bubble nor non-bubble states are detected. The index receives 3 points for identifying a real bubble state, 1 point for identifying a minor bubble state and are deducted 3 points for wrongly identifying a non-bubble state as a bubble state. This test is performed for each value of the multiplier M and the highest value are called "max" and placed in the bottom right cell.

17.2 Test of Google Search Volume Index for Housing Bubble

One in a Row Red Flag Test													
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	10	Of total	Points
# 4RBS	4	4	4	4	4	4	4	4	4	4	4	4	/ 4	3
# 6MBS	6	6	6	6	6	6	6	6	6	6	6	4	/ 6	1
# NBS	1	1	1	1	1	1	1	1	1	1	1	0	/ 10	-3
# 30 NDS	19	19	19	19	19	19	19	19	19	19	15	7	/ 40	0
Score	15	15	15	15	15	15	15	15	15	15	15	16	Max	16
Two in a Row Red Flag Test														
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	10	Of total	Points
# 4RBS	4	4	4	4	4	4	4	4	4	4	4	4	/ 4	3
# 6MBS	6	6	6	6	6	6	6	6	6	6	6	4	/ 6	1
# NBS	1	1	1	1	1	1	1	1	1	1	1	0	/ 10	-3
# 30 NDS	19	19	19	19	19	19	19	19	19	19	15	7	/ 40	0
Score	15	15	15	15	15	15	15	15	15	15	15	16	Max	16
Three in a Row Red Flag Test														
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	10	Of total	Points
# 4RBS	4	4	4	4	4	4	4	4	4	4	4	0	/ 4	3
# 6MBS	6	6	6	6	6	6	6	6	6	6	2	1	/ 6	1
# NBS	1	1	1	1	1	1	1	1	1	1	0	0	/ 10	-3
# 30 NDS	19	19	19	19	19	19	19	19	18	10	2	0	/ 40	0
Score	15	15	15	15	15	15	15	15	15	15	14	1	Max	15
Eight in a Row Red Flag Test														
M value	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.5	5	7	10	Of total	Points
# 4RBS	4	4	4	4	3	3	3	3	2	0	0	0	/ 4	3
# 6MBS	6	6	6	6	6	6	6	6	6	1	0	0	/ 6	1
# NBS	1	1	1	1	1	1	1	1	1	0	0	0	/ 10	-3
# 30 NDS	19	19	19	17	17	17	16	12	10	0	0	0	/ 40	0
Score	15	15	15	15	12	12	12	12	9	1	0	0	Max	15

Table 17-5: The table shows the result from four in-sample bubble identification tests. An index (Index12) consisting of the Average Google Search Volume Index (GSVI) for the 12 single best search terms are used to flag a state as a bubble state if the respective state has GSVI levels above a multiplier (M) times a normal level one time, two in a row, three in a row and eight in a row. The test has been performed for numerous values of M as seen in "M Value". "# 4 RBS" stands for how many, of the four; Real Bubble States (RBS) have been detected. "# 6 MBS" stands for how many, of the six; Minor Bubble States (MBS) has been detected. "# 10 NBS" stands for how many, of the ten; Non-bubble States (NBS) has wrongly been detected. "# 30 NDS" stands for how many, of the thirty; states not defined as neither bubble nor non-bubble states are detected. The index receives 3 points for identifying a real bubble state, 1 point for identifying a minor bubble state and are deducted 3 points for wrongly identifying a non-bubble state as a bubble state. This test is performed for each value of the multiplier M and the highest value are called "max" and placed in the bottom right cell.

18 Appendix H

Tests of Correlation

18.1 Test of Correlation between Google Search Volume Index for Real Estate Agent and the Housing Price Index

Correlation Between Housing Price and Google Search Volume Index for Real Estate Agent With Different Lags

State Name	Zero Lags	Two Lags	Four Lags	Six Lags	Eight Lags	Best Result
Nevada	0.8744	0.8573	0.8067	0.7304	0.6462	0.8744
Arizona	0.724	0.8104	0.8455	0.8318	0.7748	0.8455
Florida	0.8986	0.9567	0.9348	0.8571	0.7335	0.9567
California	0.9315	0.9627	0.9285	0.8436	0.698	0.9627
Average RBS	**0.8571**	**0.8968**	**0.8789**	**0.8157**	**0.7131**	**0.9098**
Maryland	0.7064	0.8572	0.9291	0.9399	0.9246	0.9399
Oregon	0.4753	0.5117	0.5392	0.5693	0.6199	0.6199
Washington	0.6372	0.7512	0.8168	0.8063	0.7421	0.8168
New Jersey	0.4235	0.5987	0.7354	0.8293	0.8836	0.8836
Virginia	0.5522	0.7196	0.8066	0.8521	0.8601	0.8601
Connecticut	0.6746	0.7665	0.8405	0.8804	0.8798	0.8804
Average MBS	**0.5782**	**0.7009**	**0.7779**	**0.8129**	**0.8184**	**0.8335**
Kansas	0.7049	0.7524	0.7299	0.7094	0.6956	0.7524
Nebraska	0.7049	0.6528	0.5585	0.4219	0.3134	0.7049
Wyoming	0.544	0.4739	0.3714	0.2321	-0.0529	0.544
Louisiana	0.2339	0.3976	0.5809	0.6749	0.6725	0.6749
Alaska	-0.034	0.2326	0.4925	0.5879	0.6279	0.6279
Texas	0.2705	0.2183	0.1118	-0.0353	-0.2115	0.2705
Iowa	0.7532	0.6693	0.6048	0.4926	0.3262	0.7532
South Dakota	-0.1095	0.0481	0.2301	0.2685	0.36	0.36
Oklahoma	0.4528	0.4909	0.5349	0.5526	0.5631	0.5631
North Dakota	0.3074	0.2778	0.2645	0.2324	0.2192	0.3074
Average NBS	**0.3828**	**0.4213**	**0.4479**	**0.4137**	**0.3514**	**0.5558**

Table 18-1: The table display the correlation between Google Search Volume Index (GSVI) for Real Estate Agent the Housing Price Index (HPI) for real, minor and non-bubble states. The average for each state group is also included: RBS stands for Real Bubble States, MBS stands for Minor Bubble States and NBS stands for non-bubble states. GSVI for Real Estate Agent are tested without lags and with 2, 4, 6 and 8 lags. "Best Result" shows the results the lagged value of GSVI for Real Estate Agent yielding the highest correlation with the HPI.

18.2 Test of Correlation between Google Search Volume Index for Housing Bubble and the Housing Price Index

Correlation Between Housing Price and Google Search Volume Index for Housing Bubble With Different Lags						
State Name	**The Whole Period**			**Best**	**Bubble**	**Non-Bubble**
	Zero lags	**Two lags**	**Four lags**	**Result**	**Period**	**Period**
Nevada	0.4855	0.3435	0.207	0.4855	0.3014	-0.1095
Arizona	0.8393	0.8554	0.8226	0.8554	0.7007	0.3966
Florida	0.8869	0.8712	0.8073	0.8869	0.7764	0.4784
California	0.925	0.8728	0.7742	0.925	0.9382	0.7925
Average RBS	**0.7841**	**0.7357**	**0.6528**	**0.7882**	**0.6792**	**0.3895**
Maryland	0.8873	0.9186	0.8943	0.9186	0.6376	0.6334
Oregon	0.6734	0.6852	0.6972	0.6972	0.3079	-0.2231
Washington	0.7362	0.7662	0.7427	0.7662	0.3853	0.6166
New Jersey	0.9051	0.9385	0.9383	0.9385	0.6864	0.407
Virginia	0.6931	0.8082	0.8542	0.8542	0.4787	-0.4125
Connecticut	0.7232	0.6947	0.6466	0.7232	0.5771	0.3662
Average MBS	**0.7697**	**0.8019**	**0.7956**	**0.8163**	**0.5122**	**0.2313**

Table 18-2: The table display the correlation between Google Search Volume Index (GSVI) for Housing Bubble and the Housing Price Index (HPI) for real and minor bubble states. The average for each state group is also included: RBS stands for Real Bubble States and MBS stands for Minor Bubble States. GSVI for Housing Bubble are tested without lags and with 2 and 4 lags. "Best Result" shows the results the lagged value of GSVI for Housing Bubble yielding the highest correlation with the HPI. In addition to being tested with different lags for the whole period, the correlation in the bubble period, Q1 2004 – Q2 2010, and in the non-bubble period, Q3 2010 – Q3 2016, is also tested.

19 Appendix I

Linear Regression of the Housing Price Index

19.1 Linear Regression of the Housing Price Index (HPI) Using Only Lagged Values of HPI and Google Search Volume Index

19.1.1 Results from the Linear Regression of Housing Price Index by Only Using Google Search Volume Index for Real Estate Agent

Short and Long Run Effects from GSVI on the House Prices									
		Long Run Effects			**Speed of Adjustment**		**Short Run Effects**		
Rank	**State Name**	**LR C**	**P>Z**	**LR^2**	**SA C**	**P>Z**	**SR C**	**P>Z**	**R^2**
1	Nevada	0.651	0.000	0.637	-0.101	0.006	0.037	0.097	0.152
2	Arizona	0.679	0.000	0.524	-0.101	0.007	0.152	0.260	0.219
3	Florida	0.848	0.000	0.808	-0.196	0.000	0.243	0.012	0.573
4	California	0.757	0.000	0.868	-0.225	0.000	0.274	0.008	0.508
Average RBS		**0.734**	**0.000**	**0.709**	**-0.156**	**0.003**	**0.176**	**0.094**	**0.363**
5	Maryland	0.452	0.000	0.499	-0.147	0.000	-0.042	0.487	0.359
6	Oregon	0.154	0.000	0.226	-0.028	0.343	0.014	0.669	0.018
7	Washington	0.276	0.000	0.406	-0.080	0.016	0.152	0.130	0.189
8	New Jersey	0.473	0.001	0.179	-0.049	0.024	-0.125	0.270	0.118
9	Connecticut	0.596	0.000	0.455	-0.059	0.024	0.011	0.835	0.113
10	Virginia	0.135	0.000	0.305	-0.100	0.001	0.007	0.697	0.191
Average MBS		**0.347**	**0.000**	**0.345**	**-0.077**	**0.068**	**0.003**	**0.515**	**0.165**
11	Michigan	0.755	0.000	0.727	-0.170	0.000	0.169	0.004	0.340
12	Rhode Island	0.367	0.000	0.360	-0.066	0.002	0.022	0.466	0.185
13	Idaho	0.277	0.000	0.522	-0.071	0.068	0.054	0.136	0.093
14	Oregon	0.369	0.000	0.943	-0.187	0.032	0.216	0.000	0.320
15	New Hampshire	0.358	0.000	0.263	-0.033	0.120	0.033	0.132	0.097
16	Minnesota	0.480	0.000	0.852	-0.137	0.030	0.073	0.076	0.122
17	Illinois	0.581	0.000	0.826	-0.084	0.091	0.115	0.055	0.174
18	Delaware	0.050	0.137	0.060	-0.008	0.691	0.031	0.000	0.255
19	Massachusetts	0.655	0.000	0.777	-0.137	0.000	0.191	0.002	0.474
20	Ohio	0.489	0.000	0.812	-0.165	0.000	0.061	0.100	0.303

21	Hawaii	0.072	0.342	0.037	-0.036	0.381	-0.014	0.449	0.048
22	New Mexico	0.115	0.000	0.303	-0.063	0.049	-0.006	0.788	0.087
23	Utah	0.253	0.000	0.361	-0.104	0.012	0.018	0.496	0.138
24	New York	0.458	0.000	0.420	-0.070	0.007	0.131	0.034	0.282
25	Maine	0.199	0.000	0.403	-0.058	0.049	0.050	0.029	0.208
26	Wisconsin	0.298	0.000	0.588	-0.114	0.000	-0.033	0.195	0.269
27	Missouri	0.286	0.000	0.764	-0.167	0.000	0.036	0.321	0.292
28	South Carolina	0.184	0.000	0.624	-0.136	0.002	0.006	0.760	0.179
29	North Carolina	0.246	0.000	0.719	-0.164	0.000	0.042	0.217	0.277
30	Alabama	0.259	0.000	0.711	-0.195	0.000	0.009	0.865	0.337
31	Mississippi	0.097	0.000	0.353	-0.020	0.567	0.002	0.809	0.006
32	Pennsylvania	0.384	0.000	0.432	-0.072	0.020	0.089	0.135	0.152
33	Indiana	0.284	0.000	0.777	-0.068	0.214	0.017	0.379	0.032
34	Colorado	0.294	0.000	0.542	0.125	0.003	-0.015	0.771	0.123
35	Vermont	0.024	0.377	0.017	-0.058	0.141	-0.010	0.453	0.069
36	Tennessee	0.157	0.000	0.587	-0.115	0.033	0.039	0.212	0.141
37	Montana	0.077	0.003	0.140	-0.068	0.083	0.008	0.513	0.053
38	Arkansas	0.073	0.000	0.270	-0.078	0.007	0.000	0.998	0.113
39	West Virginia	0.049	0.023	0.164	0.005	0.900	-0.006	0.156	0.034
40	Kentucky	0.136	0.000	0.514	-0.063	0.187	0.031	0.014	0.108
Average 30 NDS		**0.278**	**0.029**	**0.496**	**-0.086**	**0.123**	**0.045**	**0.319**	**0.177**
41	Kansas	0.112	0.000	0.497	-0.040	0.363	-0.003	0.838	0.016
42	Nebraska	0.136	0.000	0.497	-0.024	0.618	0.018	0.357	0.020
43	Wyoming	0.044	0.000	0.296	-0.089	0.040	0.006	0.206	0.111
44	Louisiana	0.032	0.157	0.055	-0.088	0.024	-0.005	0.850	0.074
45	Alaska	-0.004	0.865	0.001	-0.138	0.010	-0.002	0.865	0.154
46	Texas	0.064	0.005	0.073	0.070	0.000	0.097	0.046	0.238
47	Iowa	0.119	0.000	0.567	0.004	0.926	0.018	0.101	0.042
48	South Dakota	-0.013	0.241	0.012	-0.008	0.847	-0.007	0.535	0.010
49	Oklahoma	0.043	0.001	0.205	-0.072	0.177	0.018	0.448	0.041
50	North Dakota	0.065	0.005	0.095	0.013	0.429	-0.003	0.507	0.017
Average NBS		**0.059**	**0.141**	**0.245**	**-0.043**	**0.334**	**0.015**	**0.472**	**0.079**

Table 19-1: The table display the results from linear regressing the Housing Price Index (HPI) using only Google Search Volume Index (GSVI) levels for Real Estate agent. LR C is the long run coefficient for GSVI, P>Z is the probability that the variable is significant and LR r^2 is the long run coefficient of determinations. SA C is the coefficient for Speed of adjustment and P>Z is the probability that the coefficient is significant. SR C is the Short run coefficient for GSVI, P>Z is the probability that the variable is significant and SR r^2 is the Song run coefficient of determinations. Average RBS is the average of values for the real bubble states, MBS stands for minor bubble states, 30 NDS stands for the 30 states, not defines as either bubbler nor non-bubble states and NBS stands for non-bubble states.

19.1.2 Results from the Linear Regression of Housing Price Index by Only Using Google Search Volume Index for Real Estate Agent and a two Period Lag

Short and Long Run Effects from GSVI on the House Prices												
Rank	**Long Run Effects**					**Speed of Adjustment**		**Short Run Effects**				
State Name	**GSVI**	**P>Z**	**L2 GSVI**	**P>Z**	**R^2**	**SA C**	**P>Z**	**GSVI**	**P>Z**	**L2 GSVI**	**P>Z**	**R^2**
Nevada	0.368	0.000	0.438	0.000	0.782	0.055	0.227	0.028	0.234	0.030	0.377	0.050
Arizona	0.708	0.019	0.060	0.863	0.657	-0.109	0.007	0.197	0.180	0.140	0.251	0.232
Florida	0.808	0.000	0.105	0.358	0.917	-0.189	0.000	0.290	0.007	0.253	0.001	0.592
California	0.603	0.000	0.188	0.077	0.933	-0.159	0.027	0.287	0.004	0.273	0.003	0.496
Ave RBS	**0.622**	**0.005**	**0.198**	**0.325**	**0.822**	**-0.101**	**0.065**	**0.201**	**0.106**	**0.174**	**0.158**	**0.343**
Maryland	0.766	0.000	-0.243	0.012	0.755	-0.153	0.000	0.029	0.638	0.141	0.042	0.361
Oregon	0.068	0.428	0.115	0.171	0.283	-0.023	0.476	0.014	0.693	-0.021	0.561	0.019
Washington	0.340	0.100	-0.020	0.934	0.564	-0.083	0.068	0.145	0.191	0.137	0.150	0.216
New Jersey	1.245	0.000	-0.644	0.046	0.410	-0.042	0.047	0.095	0.422	-0.064	0.470	0.100
Connecticut	0.592	0.004	0.101	0.634	0.590	-0.030	0.518	0.067	0.145	-0.030	0.518	0.160
Virginia	0.226	0.000	-0.058	0.154	0.528	-0.097	0.001	0.021	0.325	0.021	0.322	0.181
Ave MBS	**0.540**	**0.089**	**-0.125**	**0.325**	**0.522**	**-0.071**	**0.185**	**0.062**	**0.402**	**0.031**	**0.344**	**0.173**
Michigan	0.189	0.274	0.636	0.001	0.887	-0.223	0.000	0.120	0.082	0.100	0.000	0.390
Rhode Island	0.378	0.000	0.086	0.548	0.519	-0.086	0.001	0.035	0.214	-0.007	0.768	0.261
Idaho	0.084	0.266	0.229	0.014	0.611	-0.078	0.106	0.059	0.136	0.004	0.004	0.082
Oregon	0.065	0.400	0.308	0.000	0.945	-0.140	0.108	0.200	0.003	0.086	0.200	0.351
New Hampshire	0.254	0.012	0.278	0.008	0.418	-0.046	0.065	0.016	0.456	0.030	0.053	0.133
Minnesota	0.225	0.000	0.286	0.000	0.897	-0.085	0.380	0.066	0.230	0.085	0.100	0.133
Illinois	0.355	0.000	0.251	0.013	0.871	-0.119	0.017	0.150	0.062	-0.003	0.965	0.171
Delaware	-0.141	0.000	0.392	0.000	0.421	-0.039	0.156	0.029	0.072	0.014	0.003	0.197
Massach-usetts	0.567	0.000	0.112	0.338	0.867	-0.123	0.002	0.147	0.002	0.151	0.001	0.438
Ohio	0.418	0.000	0.081	0.327	0.886	-0.179	0.000	0.080	0.038	0.070	0.112	0.276
Hawaii	0.012	0.833	0.101	0.210	0.081	0.011	0.742	-0.018	0.335	-0.018	0.154	0.064
New Mexico	0.001	0.993	0.158	0.041	0.454	-0.080	0.044	0.004	0.875	-0.027	0.365	0.122
Utah	0.269	0.000	0.049	0.495	0.535	-0.082	0.087	0.013	0.715	0.089	0.001	0.214
New York	0.524	0.000	0.037	0.794	0.598	-0.072	0.006	0.023	0.658	0.167	0.000	0.386
Maine	0.133	0.038	0.103	0.105	0.510	-0.073	0.018	0.015	0.508	0.038	0.028	0.233
Wisconsin	0.340	0.000	-0.024	0.806	0.707	-0.106	0.004	-0.003	0.937	0.042	0.227	0.190
Missouri	0.274	0.000	0.024	0.749	0.850	-0.193	0.000	0.066	0.061	0.012	0.675	0.271
South Carolina	0.206	0.000	-0.002	0-97	0.753	-0.163	0.002	0.015	0.671	0.019	0.284	0.174

North Carolina	0.286	0.000	-0.023	0.713	0.832	-0.154	0.006	0.042	0.235	0.110	0.009	0.282
Alabama	0.340	0.000	-0.061	0.352	0.855	-0.188	0.001	0.035	0.595	0.110	0.006	0.302
Mississippi	0.036	0.119	0.108	0.000	0.506	-0.006	0.895	0.001	0.897	0.002	0.804	0.002
Pennsyl -vania	0.342	0.003	0.108	0.363	0.544	-0.048	0.106	0.101	0.128	0.067	0.162	0.192
Indiana	0.111	0.011	0.184	0.000	0.832	-0.140	0.098	0.047	0.071	-0.012	0.474	0.098
Colorado	-0.08	0.341	0.381	0.000	0.548	0.132	0.001	-0.055	0.283	0.138	0.001	0.257
Vermont	0.064	0.074	0.007	0.891	0.144	-0.064	0.053	-0.002	0.831	-0.010	0.298	0.101
Tennessee	0.104	0.024	0.072	0.152	0.683	-0.114	0.069	0.040	0.257	0.023	0.506	0.136
Montana	0.026	0.375	0.082	0.001	0.234	-0.041	0.399	0.009	0.478	-0.002	0.856	0.020
Arkansas	0.136	0.000	-0.051	0.079	0.460	-0.091	0.003	0.008	0.552	0.009	0.577	0.132
West Virginia	0.01	0.525	0.058	0.051	0.235	0.021	0.661	-0.009	0.299	0.000	0.000	0.032
Kentucky	0.037	0.000	0.124	0.000	0.657	-0.130	0.010	0.045	0.004	-0.011	0.403	0.177
Ave NDS	**0.652**	**0.143**	**0.136**	**0.243**	**0.611**	**-0.090**	**0.135**	**0.043**	**0.356**	**0.043**	**0.268**	**0.194**
Kansas	0.087	0.021	0.050	0.299	0.585	-0.042	0.354	-0.003	0.872	-0.002	0.898	0.015
Nebraska	0.048	0.221	0.101	0.031	0.491	-0.001	0.987	0.015	0.472	0.022	0.321	0.045
Wyoming	0.032	0.004	0.006	0.813	0.226	-0.085	0.041	0.009	0.195	0.002	0.822	0.100
Louisiana	0.012	0.630	0.022	0.483	0.069	-0.150	0.003	-0.005	0.683	0.000	0.985	0.164
Alaska	0.047	0.331	0.005	0.929	0.158	-0.112	0.009	0.000	0.991	-0.016	0.531	0.115
Texas	-0.19	0.009	0.264	0.001	0.157	0.042	0.045	0.069	0.208	0.130	0.000	0.370
Iowa	0.020	0.475	0.103	0.001	0.592	-0.017	0.758	0.021	0.079	-0.011	0.259	0.065
South Dakota	0.020	0.349	-0.021	0.290	0.021	-0.004	0.916	-0.009	0.460	0.006	0.500	0.019
Oklahoma	-0.02	0.631	0.080	0.135	0.284	-0.077	0.174	0.022	0.475	-0.005	0.827	0.045
North Dakota	0.017	0.460	0.076	0.003	0.173	0.022	0.237	-0.003	0.579	0.006	0.211	0.066
Ave NBS	**- 0.002**	**0.346**	**0.071**	**0.298**	**0.241**	**-0.042**	**0.352**	**0.013**	**0.460**	**0.015**	**0.495**	**0.110**

Table 19-2: The table show the results from the linear regression of the Housing Price Index (HPI) using only Google Search Volume Index (GSVIt) and GSVIt-2 levels for Real Estate agent. LR C is the long run coefficient for GSVI, P>Z is the probability that the coefficient is significant, L2.GSVI is the long run coefficient for GSVIt-2, and LR r^2 is the long run coefficient of determinations. SA C is the coefficient for Speed of adjustment, SR C is the Short run coefficient for GSVI, L2.GSVI is the short run coefficient for GSVIt-2 and SR r^2 is the short run coefficient of determinations. Average RBS is the average of values for the real bubble states, MBS stands for minor bubble states, 30 NDS stands for the 30 states, not defines as either bubbler nor non-bubble states and NBS stands for non-bubble states.

19.1.3 Results from the Linear Regression of Housing Price Index (HPI) by Only Using One Period Lag of HPI and Google Search Volume Index for Real Estate Agent

Short and Long Run Effects from GSVI on the House Prices

Rank	**Long Run Effects**					**Speed of Adjustment**		**Short Run Effects**				
State Name	**HPI L1**	**P>Z**	**GSVI**	**P>Z**	**R^2**	**SA C**	**P>Z**	**HPI L1**	**P>Z**	**GSVI**	**P>Z**	**R^2**
Nevada	0.913	0.000	0.081	0.003	0.983	-0.802	0.000	1.464	0.000	0.056	0.000	0.703
Arizona	0.899	0.000	0.122	0.000	0.979	-0.457	0.032	1.068	0.000	0.100	0.154	0.675
Florida	0.789	0.000	0.223	0.000	0.991	-0.598	0.001	0.847	0.000	0.174	0.023	0.718
California	0.744	0.000	0.222	0.000	0.987	-0.468	0.004	0.915	0.000	0.146	0.021	0.759
Ave RBS	**0.836**	**0.000**	**0.162**	**0.001**	**0.985**	**-0.581**	**0.009**	**1.074**	**0.000**	**0.119**	**0.050**	**0.714**
Maryland	0.885	0.000	0.104	0.000	0.987	-0.755	0.000	1.029	0.000	-0.021	0.594	0.656
Oregon	0.967	0.000	0.007	0.387	0.961	-1.053	0.154	1.716	0.024	0.002	0.909	0.500
Washington	0.912	0.000	0.052	0.000	0.971	-0.467	0.044	1.053	0.000	0.107	0.035	0.627
New Jersey	0.966	0.000	0.090	0.000	0.985	-0.511	0.031	1.018	0.000	-0.071	0.276	0.508
Connecticut	0.943	0.000	0.083	0.000	0.988	-0.711	0.044	0.960	0.001	0.049	0.301	0.340
Virginia	0.910	0.000	0.036	0.000	0.978	-0.612	0.008	0.979	0.000	0.024	0.179	0.545
Ave MBS	**0.931**	**0.000**	**0.062**	**0.065**	**0.978**	**-0.685**	**0.047**	**1.126**	**0.004**	**0.015**	**0.382**	**0.529**
Michigan	0.911	0.000	0.061	0.229	0.985	-1.440	0.005	1.685	0.000	0.117	0.027	0.434
Rhode Island	0.936	0.000	0.064	0.000	0.991	-0.697	0.015	0.953	0.000	0.054	0.013	0.472
Idaho	0.921	0.000	0.031	0.007	0.964	-0.685	0.059	1.290	0.001	0.039	0.118	0.529
Oregon	0.729	0.000	0.100	0.000	0.981	-0.785	0.001	1.011	0.000	0.160	0.002	0.489
New Hampshire	0.972	0.000	0.022	0.281	0.983	-0.794	0.283	1.202	0.074	0.025	0.219	0.289
Minnesota	0.860	0.000	0.068	0.006	0.978	-0.975	0.014	1.281	0.001	0.061	1.220	0.318
Illinois	0.892	0.000	0.071	0.008	0.987	-0.601	0.145	0.944	0.004	0.068	0.139	0.326
Delaware	1.012	0.000	-0.013	0.022	0.979	-0.476	0.286	0.908	0.048	0.012	0.292	0.447
Massachusetts	0.836	0.000	0.130	0.000	0.988	-0.670	0.005	0.758	0.000	0.173	0.004	0.527
Ohio	0.846	0.000	0.078	0.000	0.986	-0.965	0.001	1.012	0.000	0.060	0.072	0.340
Hawaii	0.946	0.000	-0.014	0.438	0.945	-0.526	0.009	1.214	0.000	-0.021	0.001	0.630
New Mexico	0.940	0.000	0.018	0.000	0.973	-0.613	0.084	1.066	0.001	-0.007	0.735	0.388
Utah	0.908	0.000	0.053	0.000	0.959	-0.692	0.002	1.187	0.000	0.026	0.169	0.523
New York	0.915	0.000	0.087	0.000	0.980	-0.649	0.008	0.896	0.000	0.127	0.009	0.502
Maine	0.940	0.000	0.024	0.014	0.975	-0.823	0.440	1.038	0.002	0.045	0.046	0.334
Wisconsin	0.906	0.000	0.044	0.000	0.984	-0.946	0.001	1.087	0.000	-0.009	0.710	0.370
Missouri	0.848	0.000	0.053	0.000	0.983	-0.936	0.001	1.036	0.000	0.045	0.242	0.385

South Carolina	0.886	0.000	0.030	0.002	0.969	-0.989	0.001	1.252	0.000	0.007	0.742	0.360
North Carolina	0.848	0.000	0.049	0.000	0.977	-0.848	0.002	1.084	0.000	0.223	0.468	0.439
Alabama	0.843	0.000	0.056	0.000	0.979	-1.034	0.000	1.057	0.000	0.023	0.628	0.445
Mississippi	0.976	0.000	0.002	0.696	0.960	-1.384	0.320	1.693	0.250	0.000	0.956	0.128
Pennsylvania	0.910	0.000	0.064	0.000	0.967	-0.567	0.028	0.951	0.000	0.079	0.089	0.382
Indiana	0.937	0.000	0.011	0.499	0.972	-1.056	0.055	1.314	0.016	0.002	0.899	0.146
Colorado	1.127	0.000	-0.045	0.004	0.962	-0.143	0.619	0.742	0.019	-0.041	0.512	0.337
Vermont	0.955	0.000	0.026	0.000	0.961	-1.076	0.001	1.116	0.000	0.014	0.147	0.393
Tennessee	0.884	0.000	0.027	0.011	0.954	-0.865	0.010	1.177	0.001	0.043	0.115	0.344
Montana	0.930	0.000	0.000	0.956	0.935	-0.889	0.025	1.348	0.007	0.000	0.974	0.341
Arkansas	0.932	0.000	0.016	0.000	0.971	-0.841	0.006	1.004	0.000	0.009	0.466	0.293
West Virginia	1.008	0.000	-0.010	0.132	0.944	-1.190	0.035	1.369	0.014	-0.013	0.010	0.183
Kentucky	0.935	0.000	0.010	0.228	0.956	-0.974	0.169	1.141	0.112	0.026	0.035	0.155
Ave NDS	**0.916**	**0.000**	**0.037**	**0.118**	**0.971**	**-0.838**	**0.088**	**1.127**	**0.018**	**0.045**	**0.335**	**0.375**
Kansas	0.965	0.000	0.001	0.910	0.953	-1.266	0.091	1.540	0.044	-0.001	0.946	0.126
Nebraska	0.970	0.000	0.001	0.931	0.946	-1.417	0.097	1.772	0.033	-0.001	0.935	0.164
Wyoming	0.909	0.000	-0.001	0.804	0.947	-0.891	0.003	1.250	0.000	-0.003	0.599	0.362
Louisiana	0.883	0.000	0.014	0.013	0.908	-1.045	0.005	1.097	0.002	0.007	0.430	0.309
Alaska	0.921	0.000	0.012	0.017	0.915	-0.887	0.018	1.172	0.002	-0.012	0.675	0.246
Texas	1.083	0.000	-0.008	0.146	0.960	-0.172	0.537	0.688	0.004	0.058	0.264	0.404
Iowa	1.000	0.000	-0.007	0.397	0.944	-1.136	0.052	-1.415	0.031	0.011	0.235	0.171
South Dakota	0.990	0.000	0.007	0.100	0.897	-0.880	0.207	1.189	0.071	-0.004	0.683	0.154
Oklahoma	0.934	0.000	0.005	0.179	0.885	-1.445	0.057	1.628	0.028	0.023	0.345	0.123
North Dakota	1.012	0.000	0.001	0.872	0.990	-0.725	0.456	0.973	0.326	-0.001	0.678	0.077
Ave NBS	**0.967**	**0.000**	**0.003**	**0.384**	**0.932**	**-0.955**	**0.159**	**0.928**	**0.055**	**0.009**	**0.538**	**0.223**

Table 19-3: The table display the results from the linear regression of the Housing Price Index (HPI) using only a one period lagged value of HPI (L.HPI) and Google Search Volume Index (GSVIt) levels for Real Estate agent. L.HPI is the long run coefficient for HPIt-1, P>Z is the probability that the coefficient is significant, GSVI is the long run coefficient for GSVIt, and LR r^2 is the long run coefficient of determinations. SA C is the coefficient for Speed of adjustment, L.HPI is the Short run coefficient for HPIt-1, GSVI is the short run coefficient for GSVIt and SR r^2 is the short run coefficient of determinations. Average RBS is the average of values for the real bubble states, MBS stands for minor bubble states, 30 NDS stands for the 30 states, not defines as either bubbler nor non-bubble states and NBS stands for non-bubble states.

19.2 Linear Regression of the Housing Price Index (HPI) Using the Baseline Model Variables

19.2.1 Results from the Linear Regression of Housing Price Index (HPI) Using the Baseline Variables

Long and Short Run Effects							
Rank	**State Name**	**LR R^2**	**LR MAE**	**SR R^2**	**SR MAE**	**SA C**	**P>Z**
1	Nevada	0.993	1.856%	0.798	1.148%	-0.550	0.006
2	Arizona	0.992	1.398%	0.844	1.129%	-0.720	0.001
3	Florida	0.994	1.168%	0.830	1.136%	-0.652	0.000
4	California	0.990	1.553%	0.790	1.201%	-0.542	0.017
Average RBS		**0.992**	**1.494%**	**0.816**	**1.153%**	**-0.616**	**0.006**
5	Maryland	0.987	1.114%	0.749	0.882%	-0.712	0.000
6	Oregon	0.985	1.136%	0.768	0.946%	-0.656	0.001
7	Washington	0.984	1.071%	0.748	0.939%	-0.556	0.024
8	New Jersey	0.991	0.968%	0.725	0.809%	-0.678	0.001
9	Connecticut	0.991	0.897%	0.673	0.762%	-0.776	0.000
10	Virginia	0.985	0.919%	0.773	0.743%	-0.790	0.000
Average MBS		**0.987**	**1.017%**	**0.739**	**0.847%**	**-0.695**	**0.004**
11	Michigan	0.991	1.100%	0.661	1.021%	-0.653	0.001
12	Rhode Island	0.993	0.986%	0.745	0.866%	-0.812	0.000
13	Idaho	0.985	1.136%	0.768	0.946%	-0.656	0.001
14	Georgia	0.987	1.074%	0.664	0.979%	-0.734	0.001
15	New Hampshire	0.989	0.939%	0.619	0.826%	-0.760	0.003
16	Minnesota	0.985	1.101%	0.630	1.040%	-0.756	0.001
17	Illinois	0.992	0.821%	0.669	0.736%	-0.829	0.001
18	Delaware	0.988	1.102%	0.651	0.904%	-0.878	0.000
19	Massachusetts	0.988	0.888%	0.598	0.819%	-0.778	0.002
20	Ohio	0.986	0.891%	0.558	0.758%	-0.643	0.006
21	Hawaii	0.982	0.973%	0.787	0.835%	-0.798	0.000
22	New Mexico	0.980	0.963%	0.635	0.811%	-0.735	0.001
23	Utah	0.975	1.091%	0.715	0.943%	-0.697	0.005
24	New York	0.988	0.695%	0.761	0.620%	-0.889	0.000
25	Maine	0.984	0.842%	0.666	0.687%	-0.836	0.000
26	Wisconsin	0.985	0.854%	0.587	0.704%	-0.632	0.006

27	Missouri	0.982	0.825%	0.614	0.656%	-0.665	0.004
28	South Carolina	0.973	0.908%	0.612	0.748%	-0.745	0.003
29	North Carolina	0.980	0.756%	0.671	0.634%	-0.575	0.017
30	Alabama	0.978	0.766%	0.627	0.672%	-0.794	0.010
31	Mississippi	0.970	0.809%	0.503	0.699%	-0.918	0.001
32	Pennsylvania	0.976	0.745%	0.655	0.641%	-0.747	0.002
33	Indiana	0.977	0.725%	0.561	0.645%	-0.592	0.008
34	Colorado	0.981	0.823%	0.672	0.828%	-0.762	0.005
35	Vermont	0.963	0.861%	0.566	0.763%	-1.080	0.000
36	Tennessee	0.963	0.770%	0.579	0.729%	-0.768	0.001
37	Montana	0.966	0.736%	0.668	0.723%	-0.969	0.000
38	Arkansas	0.972	0.680%	0.553	0.615%	-1.042	0.000
39	West Virginia	0.950	0.865%	0.492	0.768%	-1.098	0.000
40	Kentucky	0.964	0.631%	0.534	0.542%	-0.616	0.012
Average 30 NDS		**0.979**	**0.879%**	**0.634**	**0.772%**	**-0.782**	**0.003**
41	Kansas	0.961	0.714%	0.443	0.683%	-0.824	0.010
42	Nebraska	0.961	0.659%	0.485	0.622%	-0.724	0.023
43	Wyoming	0.963	0.800%	0.596	0.742%	-0.923	0.000
44	Louisiana	0.911	0.877%	0.460	0.786%	-1.076	0.001
45	Alaska	0.922	0.838%	0.485	0.753%	-0.948	0.000
46	Texas	0.977	0.590%	0.668	0.538%	-0.615	0.019
47	Iowa	0.951	0.586%	0.460	0.539%	-0.685	0.024
48	South Dakota	0.912	0.681%	0.458	0.639%	-0.832	0.009
49	Oklahoma	0.891	0.731%	0.455	0.669%	-0.847	0.004
50	North Dakota	0.992	0.675%	0.367	0.638%	-1.101	0.001
Average NBS		**0.944**	**0.715%**	**0.488**	**0.661%**	**-0.858**	**0.009**

Table 19-4: The table shows the results for the Baseline Model for the Housing Price Index (HPI) in all 50 states. The model includes the following independent variables: HPI with one lagged period, Unemployment Rate, Interest Rate, Population, Personal Disposable Income and Housing Permits Authorized. All variables are transformed to the natural logarithm. LR R^2 is the long run coefficient of determinations, LR error is the Mean Absolute Error between predicted value and real value for HPI at level, SR R^2 is the short run coefficient of determination, SR error is the MAE between predicted change in HPI and real value, SA C is the coefficient for speed of adjustment and P>Z is the probability that the coefficient is significant.

19.2.2 Results from the Linear Regression of Housing Price Index (HPI) Using the Baseline Variables and Including Google Search Volume Index for Real Estate Agent

Baseline Model Including GSVI		**Long and Short Run Effects**					
Rank	**State Name**	**LR R^2**	**LR MAE**	**SR R^2**	**SR MAE**	**SA C**	**P>Z**
1	Nevada	0.994	1.710%	0.840	1.286%	-0.639	0.000
2	Arizona	0.992	1.398%	0.845	1.077%	-0.683	0.004
3	Florida	0.994	1.158%	0.846	1.053%	-0.697	0.000
4	California	0.990	1.495%	0.804	1.168%	-0.636	0.004
Average RBS		**0.993**	**1.440%**	**0.834**	**1.146%**	**-0.664**	**0.002**
5	Maryland	0.990	0.938%	0.811	0.739%	-0.924	0.000
6	Oregon	0.985	1.128%	0.769	0.943%	-0.663	0.001
7	Washington	0.985	1.019%	0.767	0.886%	-0.573	0.012
8	New Jersey	0.992	0.904%	0.760	0.770%	-0.731	0.000
9	Connecticut	0.991	0.895%	0.704	0.705%	-0.809	0.000
10	Virginia	0.985	0.867%	0.785	0.725%	-0.809	0.000
Average MBS		**0.988**	**0.959%**	**0.766**	**0.795%**	**-0.752**	**0.002**
11	Michigan	0.990	1.087%	0.666	1.012%	-0.676	0.001
12	Rhode Island	0.993	0.988%	0.761	0.868%	-0.755	0.000
13	Idaho	0.985	1.128%	0.769	0.939%	0.659	0.001
14	Georgia	0.987	1.078%	0.680	0.982%	-0.779	0.000
15	New Hampshire	0.989	0.939%	0.619	0.826%	-0.760	0.003
16	Minnesota	0.985	1.085%	0.653	1.011%	-0.807	0.000
17	Illinois	0.994	0.674%	0.742	0.650%	-1.014	0.000
18	Delaware	0.987	1.010%	0.658	0.890%	-0.889	0.000
19	Massachusetts	0.988	0.838%	0.634	0.763%	-0.909	0.001
20	Ohio	0.987	0.837%	0.672	0.658%	-0.959	0.000
21	Hawaii	0.982	0.961%	0.785	0.842%	-0.813	0.000
22	New Mexico	0.985	0.858%	0.699	0.779%	-0.933	0.000
23	Utah	0.975	1.098%	0.718	0.934%	-0.709	0.005
24	New York	0.988	0.694%	0.765	0.615%	-0.851	0.000
25	Maine	0.984	0.828%	0.676	0.667%	-0.787	0.000
26	Wisconsin	0.987	0.718%	0.660	0.618%	-0.868	0.000
27	Missouri	0.983	0.749%	0.638	0.664%	-0.657	0.004
28	South Carolina	0.975	0.941%	0.640	0.777%	-0.708	0.003

29	North Carolina	0.983	0.744%	0.709	0.608%	-0.642	0.002
30	Alabama	0.978	0.732%	0.652	0.618%	-0.846	0.000
31	Mississippi	0.971	0.798%	0.523	0.670%	-0.992	0.001
32	Pennsylvania	0.977	0.733%	0.675	0.610%	-0.830	0.000
33	Indiana	0.979	0.711%	0.603	0.625%	-0.741	0.001
34	Colorado	0.981	0.824%	0.671	0.827%	-0.761	0.005
35	Vermont	0.962	0.863%	0.585	0.746%	-1.090	0.000
36	Tennessee	0.962	0.771%	0.591	0.713%	-0.739	0.002
37	Montana	0.967	0.734%	0.714	0.686%	-0.934	0.000
38	Arkansas	0.971	0.677%	0.550	0.617%	-1.031	0.000
39	West Virginia	0.952	0.848%	0.505	0.728%	-1.072	0.000
40	Kentucky	0.966	0.625%	0.588	0.523%	-0.785	0.000
Average 30 NDS		**0.980**	**0.852%**	**0.660**	**0.749%**	**-0.789**	**0.001**
41	Kansas	0.960	0.715%	0.458	0.682%	-0.826	0.009
42	Nebraska	0.966	0.621%	0.545	0.580%	-0.897	0.004
43	Wyoming	0.962	0.800%	0.599	0.732%	-0.906	0.000
44	Louisiana	0.912	0.858%	0.479	0.792%	-1.109	0.001
45	Alaska	0.929	0.822%	0.510	0.743%	-0.997	0.001
46	Texas	0.977	0.603%	0.670	0.538%	-0.619	0.014
47	Iowa	0.950	0.586%	0.463	0.537%	-0.690	0.023
48	South Dakota	0.909	0.678%	0.463	0.632%	-0.834	0.011
49	Oklahoma	0.893	0.715%	0.478	0.649%	-0.923	0.002
50	North Dakota	0.991	0.676%	0.369	0.644%	-1.105	0.001
Average NBS		**0.943**	**0.707%**	**0.503**	**0.653%**	**-0.891**	**0.007**

Table 19-5: The able shows the results for the Baseline Model, including Google Search Volume Index (GSVI) for Real Estate Agent, for the Housing Price Index (HPI) in all 50 states. The model includes the following independent variables: HPI with one lagged period, GSVI, Unemployment Rate, Interest Rate, Population, Personal Disposable Income and Housing Permits Authorized. All variables are transformed to the natural logarithm. LR R^2 is the long run coefficient of determinations, LR error is the Mean Absolute Error between predicted value and real value for HPI at level, SR R^2 is the short run coefficient of determination, SR error is the MAE between predicted change in HPI and real value, SA C is the coefficient for speed of adjustment and P>Z is the probability that the coefficient is significant.

19.2.3 Results from the Linear Regression of Housing Price Index (HPI) Using the Baseline Variables and Including the Consumer Confidence Index (CCI)

Baseline Model Including CCI		**Long and Short Run Effects**					
Rank	**State Name**	**LR R^2**	**LR MAE**	**SR R^2**	**SR MAE**	**SA C**	**P>Z**
1	Nevada	0.994	1.710%	0.840	1.286%	-0.639	0.000
2	Arizona	0.992	1.398%	0.845	1.077%	-0.683	0.004
3	Florida	0.994	1.158%	0.846	1.053%	-0.697	0.000
4	California	0.990	1.495%	0.804	1.168%	-0.636	0.004
Average RBS		**0.993**	**1.440%**	**0.834**	**1.146%**	**-0.664**	**0.002**
5	Maryland	0.990	0.938%	0.811	0.739%	-0.924	0.000
6	Oregon	0.985	1.128%	0.769	0.943%	-0.663	0.001
7	Washington	0.985	1.019%	0.767	0.886%	-0.573	0.012
8	New Jersey	0.992	0.904%	0.760	0.770%	-0.731	0.000
9	Connecticut	0.991	0.895%	0.704	0.705%	-0.809	0.000
10	Virginia	0.985	0.867%	0.785	0.725%	-0.809	0.000
Average MBS		**0.988**	**0.959%**	**0.766**	**0.795%**	**-0.752**	**0.002**
11	Michigan	0.990	1.087%	0.666	1.012%	-0.676	0.001
12	Rhode Island	0.993	0.988%	0.761	0.868%	-0.755	0.000
13	Idaho	0.985	1.128%	0.769	0.939%	0.659	0.001
14	Georgia	0.987	1.078%	0.680	0.982%	-0.779	0.000
15	New Hampshire	0.989	0.939%	0.619	0.826%	-0.760	0.003
16	Minnesota	0.985	1.085%	0.653	1.011%	-0.807	0.000
17	Illinois	0.994	0.674%	0.742	0.650%	-1.014	0.000
18	Delaware	0.987	1.010%	0.658	0.890%	-0.889	0.000
19	Massachusetts	0.988	0.838%	0.634	0.763%	-0.909	0.001
20	Ohio	0.987	0.837%	0.672	0.658%	-0.959	0.000
21	Hawaii	0.982	0.961%	0.785	0.842%	-0.813	0.000
22	New Mexico	0.985	0.858%	0.699	0.779%	-0.933	0.000
23	Utah	0.975	1.098%	0.718	0.934%	-0.709	0.005
24	New York	0.988	0.694%	0.765	0.615%	-0.851	0.000

25	Maine	0.984	0.828%	0.676	0.667%	-0.787	0.000
26	Wisconsin	0.987	0.718%	0.660	0.618%	-0.868	0.000
27	Missouri	0.983	0.749%	0.638	0.664%	-0.657	0.004
28	South Carolina	0.975	0.941%	0.640	0.777%	-0.708	0.003
29	North Carolina	0.983	0.744%	0.709	0.608%	-0.642	0.002
30	Alabama	0.978	0.732%	0.652	0.618%	-0.846	0.000
31	Mississippi	0.971	0.798%	0.523	0.670%	-0.992	0.001
32	Pennsylvania	0.977	0.733%	0.675	0.610%	-0.830	0.000
33	Indiana	0.979	0.711%	0.603	0.625%	-0.741	0.001
34	Colorado	0.981	0.824%	0.671	0.827%	-0.761	0.005
35	Vermont	0.962	0.863%	0.585	0.746%	-1.090	0.000
36	Tennessee	0.962	0.771%	0.591	0.713%	-0.739	0.002
37	Montana	0.967	0.734%	0.714	0.686%	-0.934	0.000
38	Arkansas	0.971	0.677%	0.550	0.617%	-1.031	0.000
39	West Virginia	0.952	0.848%	0.505	0.728%	-1.072	0.000
40	Kentucky	0.966	0.625%	0.588	0.523%	-0.785	0.000
Average 30 NDS		**0.980**	**0.852%**	**0.660**	**0.749%**	**-0.789**	**0.001**
41	Kansas	0.960	0.715%	0.458	0.682%	-0.826	0.009
42	Nebraska	0.966	0.621%	0.545	0.580%	-0.897	0.004
43	Wyoming	0.962	0.800%	0.599	0.732%	-0.906	0.000
44	Louisiana	0.912	0.858%	0.479	0.792%	-1.109	0.001
45	Alaska	0.929	0.822%	0.510	0.743%	-0.997	0.001
46	Texas	0.977	0.603%	0.670	0.538%	-0.619	0.014
47	Iowa	0.950	0.586%	0.463	0.537%	-0.690	0.023
48	South Dakota	0.909	0.678%	0.463	0.632%	-0.834	0.011
49	Oklahoma	0.893	0.715%	0.478	0.649%	-0.923	0.002
50	North Dakota	0.991	0.676%	0.369	0.644%	-1.105	0.001
	Average NBS	**0.943**	**0.707%**	**0.503**	**0.653%**	**-0.891**	**0.007**

Table 19-6: The table shows the results for the Baseline Model, including the Consumer Confidence Index (CPI) for the Housing Price Index (HPI) in all 50 states. The model includes the following independent variables: HPI with one lagged period, Unemployment Rate, Interest Rate, Population, Personal Disposable Income and Housing Permits Authorized. All variables are transformed to the natural logarithm. LR R^2 is the long run coefficient of determinations, LR error is the Mean Absolute Error between predicted value and real value for HPI at level, SR R^2 is the short run coefficient of determination, SR error is the MAE between predicted change in HPI and real value, SA C is the coefficient for speed of adjustment and P>Z is the probability that the coefficient is significant.

20 Appendix J

Linear Regression Models of the Housing Price Index

20.1 Linear regression models of HPI globally in the U.S.

For all the models shown in this sub-chapter, the following abbreviations are applicable:

HPI_t = The House Price Index at time t
$GSVI_{w,t}$ = Google Search Volume Index for search term w, at time t
$\epsilon_{HPI,t-1}$ = The error correction term

The long run effect from GSVI for Housing Bubble on HPI

$$HPI_t = \alpha + \beta GSVI_{HB,t} \tag{20}$$

The Short run effect from GSVI for Housing Bubble and the speed of adjustment

$$\Delta HPI_t = \alpha + \beta \Delta GSVI_{HB,t} + \gamma * \epsilon_{HPI,t-1} \tag{21}$$

The long run effect from GSVI for Real Estate Agent on HPI

$$HPI_t = \alpha + \beta GSVI_{REA,t} \tag{22}$$

The Short run effect from GSVI for Real Estate Agent and the speed of adjustment

$$\Delta HPI_t = \alpha + \beta \Delta GSVI_{REA,t} + \gamma \epsilon_{HPI,t-1} \tag{23}$$

The long run effect from GSVI for Housing Bubble and a two period lag of Housing Bubble on HPI

$$HPI_t = \alpha + \beta_1 GSVI_{HB,t} + \beta_2 GSVI_{HB,t-2} \tag{24}$$

The Short run effect from GSVI for Housing Bubble and a two period lag of Housing Bubble and the speed of adjustment

$$\Delta HPI_t = \alpha + \beta_1 * \Delta GSVI_{HB,t} + \beta_2 * \Delta GSVI_{HB,t-2} + \gamma * \epsilon_{HPI,t-1} \tag{25}$$

The long run effect from GSVI for Real Estate Agent and a two period lag on HPI

$$HPI_t = \alpha + \beta_1 * GSVI_{REA,t} + \beta_2 * GSVI_{REA,t-2} \tag{26}$$

The Short run effect from GSVI for Real Estate Agent and a two period lag of Real Estate Agent and the speed of adjustment

$$\Delta HPI_t = \alpha + \beta_1 * \Delta GSVI_{REA,t} + \beta_2 * \Delta GSVI_{REA,t-2} + \gamma * \epsilon_{HPI,t-1} \tag{27}$$

For all the models shown in this sub-chapter, the following abbreviations are applicable:

HPI_t = The House Price Index at time t
$GSVI_{w,t}$ = Google Search Volume Index for search term w, at time t
$\epsilon_{HPI,t-1}$ = The error correction term

The long run effect from GSVI for Real Estate Agent and Housing Bubble on HPI

$$\widehat{HPI}_t = \alpha + \beta_1 * GSVI_{REA,t} + \beta_2 * GSVI_{HB,t} \tag{28}$$

The Short run effect from GSVI for Real Estate Agent and Housing Bubble and the speed of adjustment

$$\Delta\widehat{HPI}_t = \alpha + \beta_1 * \Delta GSVI_{REA,t} + \beta_2 * \Delta GSVI_{HB,t} + \epsilon_{HPI,t-1} \tag{29}$$

The long run effect from one period lagged HPI and GSVI for Housing Bubble on HPI

$$\widehat{HPI}_t = \alpha + \beta_1 * HPI_{t-1} + \beta_2 * GSVI_{HB,t} \tag{30}$$

The Short run effect from one period lagged HPI and GSVI for Housing Bubble and the speed of adjustment

$$\Delta\widehat{HPI}_t = \alpha + \beta_1 * \Delta HPI_{t-1} + \beta_2 * \Delta GSVI_{HB,t} + \gamma * \epsilon_{HPI,t-1} \tag{31}$$

The long run effect from one period lagged HPI and GSVI for Real Estate Agent on HPI

$$\widehat{HPI}_t = \alpha + \beta_1 * HPI_{t-1} + \beta_2 * GSVI_{REA,t} \tag{32}$$

The Short run effect from one period lagged HPI and GSVI for Real Estate Agent and the speed of adjustment

$$\Delta\widehat{HPI}_t = \alpha + \beta_1 * \Delta HPI_{t-1} + \beta_2 * \Delta GSVI_{REA,t} + \gamma * \epsilon_{HPI,t-1} \tag{33}$$

The long run effect from one period lagged HPI and GSVI for Real Estate Agent and Housing Bubble on HPI

$$\widehat{HPI}_t = \alpha + \beta_1 * HPI_{t-1} + \beta_2 * GSVI_{REA,t} + \beta_3 * GSVI_{HB,t} \tag{34}$$

The Short run effect from one period lagged HPI and GSVI for Real Estate Agent and Housing Bubble and the speed of adjustment

$$\Delta\widehat{HPI}_t = \alpha + \beta_1 * \Delta HPI_{t-1} + \beta_2 * \Delta GSVI_{REA,t} + \beta_3 * \Delta GSVI_{HB,t} + \gamma * \epsilon_{HPI,t-1} \tag{35}$$

20.2 Linear Regression Models of HPI for all 50 United States Using Only Google Searches

For all the models shown in this sub-chapter, the following abbreviations are applicable:

$HPI_{s,t}$ = The House Price Index for state s, at time t

$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t

$\epsilon_{HPI,t-1}$ = The error correction term from the previous period $t-1$

20.2.1 Linear regression of HPI with only GSVI as independent variable

The long run effect

$$HPT_{s,t} = \alpha + \beta * GSVI_{REA,s,t} \quad (36)$$

The Short run effect and the speed of adjustment

$$\Delta HPT_{s,t} = \alpha + \beta * \Delta GSVI_{REA,s,t} + \gamma * \epsilon_{HPI,s,t-1} \quad (37)$$

20.2.2 Linear Regression of HPI Using Only GSVI and Two Period lag of GSVI

The long run effect

$$HPT_{s,t} = \alpha + \beta_1 * GSVI_{REA,s,t} + \beta_2 * GSVI_{REA,s,t-2} \quad (38)$$

The Short run effect and the speed of adjustment

$$\Delta HPT_{s,t} = \alpha + \beta_1 * \Delta GSVI_{REA,s,t} + \beta_2 * \Delta GSVI_{REA,s,t-2} + \gamma * \epsilon_{HPI,s,t-1} \quad (39)$$

20.2.3 J.2.3 Linear Regression of HPI Using Only One Period Lag of HPI and GSVI

The long run effect

$$HPT_{s,t} = \alpha + \beta_1 * HPI_{REA,s,t-1} + \beta_2 * GSVI_{REA,s,t} \quad (40)$$

The Short run effect and the speed of adjustment

$$\Delta HPT_{s,t} = \alpha + \beta_1 * \Delta HPI_{REA,s,t-1} + \beta_2 * \Delta GSVI_{REA,s,t} + \gamma * \epsilon_{HPI,s,t-1} \quad (41)$$

20.3 The Baseline Error Correction Model for all 50 states

For all the models shown in this sub-chapter, the following abbreviations are applicable:

$HPI_{s,t}$ = The House Price Index for state s, at time t
DPI_t = Disposable Personal Income at time t
$HPA_{s,t}$ = Housing Permits Authorized for state s, at time t
$UR_{s,t}$ = Unemployment Rate for state s, at time t
IR_t = Interest Rate at time t
$PO_{s,t}$ = Population in state s, at time t
$GSVI_{w,s,t}$ = Google Search Volume Index for search term w, in state s, at time t
CCI_t = The Consumer Confidence Index
$\epsilon_{HPI,t-1}$ = The error correction term

20.3.1 The Baseline Model

The long run effect

$$HPT_{s,t} = \alpha + \beta_1 HPI_{s,t-1} + \beta_2 UR_{s,t} + \beta_3 PO_{s,t} + \beta_4 DPI_t + \beta_5 IR_t + \beta_6 HPA_{s,t} \qquad (42)$$

The Short run effect and the speed of adjustment

$$\Delta HPT_{s,t} = \alpha + \beta_1 \Delta HPI_{s,t-1} + \beta_2 \Delta UR_{s,t} + \beta_3 \Delta PO_{s,t} + \beta_4 \Delta DPI_t + \beta_5 \Delta IR_t + \beta_6 \Delta HPA_{s,t} + \gamma \epsilon_{HPI,s,t-1} \qquad (43)$$

20.3.2 The Baseline Model Including GSVI for Real Estate Agent

The long run effect

$$HPI_{s,t} = \alpha + \beta_1 HPI_{s,t-1} + \beta_2 UR_{s,t} + \beta_3 PO_{s,t} + \beta_4 DPI_t + \beta_5 IR_t + \beta_6 HPA_{s,t} + \beta_7 GSVI_{REA,s,t} \qquad (44)$$

The Short run effect and the speed of adjustment

$$\Delta HPT_{s,t} = \alpha + \beta_1 \Delta HPI_{s,t-1} + \beta_2 \Delta UR_{s,t} + \beta_3 \Delta PO_{s,t} + \beta_4 \Delta DPI_t + \beta_5 \Delta IR_t + \beta_6 \Delta HPA_{s,t} + \beta_7 \Delta GSVI_{REA,s,t} + \gamma \epsilon_{HPI,s,t-1} \qquad (45)$$

20.3.3 The Baseline Model Including the Consumer Confidence Index (CCI)

The long run effect

$$HPI_{s,t} = \alpha + \beta_1 HPI_{s,t-1} + \beta_2 UR_{s,t} + \beta_3 PO_{s,t} + \beta_4 DPI_t + \beta_5 IR_t + \beta_6 HPA_{s,t} + \beta_7 CCI_t \qquad (46)$$

The short run effect and the speed of adjustment

$$\Delta HPT_{s,t} = \alpha + \beta_1 \Delta HPI_{s,t-1} + \beta_2 \Delta UR_{s,t} + \beta_3 \Delta PO_{s,t} + \beta_4 \Delta DPI_t + \beta_5 \Delta IR_t + \beta_6 \Delta HPA_{s,t} + \beta_7 \Delta CCI_{REA,s,t} + \gamma \epsilon_{HPI,s,t-1} \qquad (47)$$