IZA DP No. 9427

Trend-Spotting in the Housing Market

Nikos Askitas

October 2015

Trend-Spotting in the Housing Market

Nikos Askitas
IZA

ABSTRACT

Trend-Spotting in the Housing Market[*]

I create a time series of weekly ratios of Google searches, in the US, on buying and selling in the Real Estate Category of Google Trends. I call this ratio the Google US Housing Market BUSE Index or simply the BUSE index. It expresses the number of "buy"-searches for each "sell"-search which, by means of certain regularity assumptions on the distribution of Internet users, I think is a good proxy of the number of prospective home buyers for each prospective home seller in the pool of prospective housing market participants. I show this ratio to have several unique, desirable properties which make it useful for understanding and nowcasting the US housing market. Firstly it has a significant correlation with the US national S&P/Case-Shiller Home Price Index. Since the latter is monthly and published as a three-month moving average with a two month lag and the Google Trends data is weekly we can have a short term nowcasting of housing prices in the US. In the seasonal variations of this ratio the BUSE index recaptures traces of prospect theory whose applicability in the housing market has been well documented. I show how these Google data can be used to create a consistent narrative of the post bubble burst dynamics in the US housing market and propose the BUSE index as an instrument for monitoring housing market conditions.

Corresponding author:

Nikos Askitas
Institute for the Study of Labor (IZA)
Schaumburg-Lippe-Strasse 5-9
53113 Bonn
Germany
E-mail: askitas@iza.org

---

1. INTRODUCTION

The US housing market is the subject of much research for many good reasons. A house is an asset and it is at the same time a home. As an asset it is related to a homeowner's long term expectations while as a home it defines a homeowner's lifestyle and forms her life attitudes. As a commodity it is related to a large supply chain of constructions materials and home equipment and it generates a significant number of jobs for construction, maintenance and sales. For these reasons it is also often an instrument of government intervention to the economy as a whole, a fact which contributes to the inherent endogeneity in the formation of home prices. These facts also explain why conventional economic wisdom is not entirely unfounded to maintain that as the housing market goes so goes the rest of the economy.

There is surely no shortage of house price indices in the US housing market, a fact which underlines the importance of this market, so why would we need another one? In order to answer this question but also in order to explain my choices in this paper I need to briefly discuss the available indices.[1] There are five main housing price indices two of which are so called "median house price indices", the other three being "repeat sales indices". The former kind consists of the index by the National Association of Realtors (NAR) and the one by the Census Bureau (CB) while the latter kind consists of an index by the Federal Housing Finance Agency (FHFA) and two proprietary, one from CoreLogic (CL) and the well known S&P/Case-Shiller national index (S&P). These indices have various pros and cons and exhibit differences which may be explained by their methodologies. In short, median house price indices are blind to intrinsic, hedonic value while the repeat sales indices use a prior sale as a proxy for the hedonic value. The FHFA index is a repeat sales index which however is based only on sales of houses securitised by Fannie Mae and Freddie Mac while S&P and CL are based on "arm's length transactions" with the CL being slightly broader. I choose the S&P index because its data is readily available on the S&P website.

So why do we need yet another index? If we could in some way capture buying thoughts of prospective home buyers or selling thoughts of prospective home sellers in real time we should be able to monitor housing market conditions regardless of the fundamentals that may be driving the housing market. A simple ratio of the number of buying thoughts to the number of selling thoughts ought to tell us something about the formation of upcoming home prices. Underwriting standards, interest rates, mortgage rates, lending trends and practices, inflow of foreign capital, prevalence of securitisation of mortgages, government programs for affordable housing, tax incentives for home buyers, labor market conditions, whatever the fundamentals are each time they ought to shape that ratio and in fact they ought to be captured by that ratio. So the answer to the question of "why yet another index" is that I seek to construct an index which is buzz based and contains the sperm of what later becomes price. It is for that reason that "arm's length transactions" are better suited as a target variable.

But where can we find data to build a time series of the number of buying thoughts to the number of selling thoughts in the population? The answer to this question may be different

---

[1]The information is largely based on information from the "St. Louis Fed On the Economy blog" and was pointed out to me by Christian Zimmermann.

in each era but in ours the place to search for this type of data is the internet. Newspapers played a decisive role in the early history of "speculative bubbles" which are nothing else but market expectations of one kind or another. Technological improvement (vacuum tube amplification) brought the price of the telephone down in the early 1900 and helped its adoption by households, making long distance calls affordable. The subsequent proliferation of the telephone played a role in the "volatile stock market of the 1920's"[2]. The stock market boom of the 1990s was similarly accompanied by another technological revolution: the advent of the internet. Email, Google search, chat rooms and their more evolved variants of Social Media as we know them today brought on the era of a more intensive "interpersonal contagion of ideas". It is therefore not far fetched to search for market clues in internet data, especially at a time when virtually every market has an online component (Askitas and Zimmermann (2015)).

In Askitas and Zimmermann (2011) we could show how by looking at Google Search intensities for "hardship letter" we may nicely nowcast mortgage delinquency rates in the US housing market during the surge of mortgage delinquencies. In this paper I follow and adapt an idea in Askitas (2015) and look at ratios of searches containing the word "buy" to searches containing the word "sell", in the Google category Real Estate. These are the buying and selling thoughts of our prospective market participants. I thus obtain an index, which I call the Google US Housing Market BUSE Index or simple the BUSE index. It captures the relative proliferation of prospective buyers to prospective sellers in the pool of prospective buyers and sellers i.e. the pool of all prospective market participants[3].

Since sooner or later the Efficient Market Hypothesis weighs in on asset price movements, even if it some times gets obstructed by various types of irrationalities, predicting the (far) future is a futile exercise. Consistent with Choi and Varian (2012) I don't hence claim to predict future prices but simply to nowcast their formation in the present. In the literature which uses Google Trends to forecast economic variables the standard approach is to enhance a standard seasonal autoregressive model with Google Trends categorical data and record improvements of the mean square error as in Choi and Varian (2012). The novelty of this paper is that we use the Google Trends category "Real Estate" but take the ratio of buy to sell searches therein. This technique has only been used in Askitas (2015) in another context with very good results. When someone searches information on selling in the Real Estate category then they are very likely a prospective seller of real estate and when they are searching for information on buying they are prospective buyers. Diluting or concentrating prospective buyers then in this pool ought to have a significance in the housing market as it ought to be related to supply and demand as it indeed does. In forming the ratio of searches I get rid of the denominator in the Google Trends data and hence observe ratios of the real number of such searches which I can then more justifiably view as the concentration of prospective buyers in the pool of prospective housing market participants.

In the post 2006 Bubble burst US housing market this ratio correlates strongly and negatively with housing prices as expressed by the S&P/Case-Shiller U.S. National Home Price Index

---

[2]Rober J. Shiller, Irrational Exuberance, Third Edition pp.101, 181, 182.

[3]Adding "build" searches appears to sharpen the results. This I believe is due to the fact that builders may be former buyers and data on building permits shows that building new homes is currently on the rise.

and this is the main result of this paper. In the boom phase (which we can unfortunately only observe since January 2004) as prices increase prospective buyers are being increasingly diluted in a pool where sellers proliferate, setting the stage for a downturn. When the probability that a house on sale will be sold reaches a trigger threshold (which we estimate around 15%) the bust phase is initiated with falling prices and an increase in the concentration of prospective buyers among decreasing prospective sellers. The movement of the BUSE index counter to prices is consistent with a phase difference between buyers and sellers: in a boom sellers are accelerating their entry in the pool only after buyers start slowing down due to the high prices and that in a bust sellers are living at accelerating rates after buyers start slowing their exit down. A more technical way to say this is that the percentage of change of buyers and sellers is related so that when one reaches its local extremum the other one changes concavity.

Observing the seasonal properties of this ratio we see that while the relative intensities of both "buy" and "sell" searches have a Christmas time trough the ratio of buyers to sellers exhibits a peak, a phenomenon consistent we think with prospect theory which postulates that all other things kept equal a loss hurts more than a comparable gain pleases. The trough in both buy and sell searches means that Housing market participation is viewed or felt as incompatible to the Christmas time, family oriented, hedonic bliss. The fact now that the buy to sell ratio spikes tells us that selling is more incompatible than buying. The intensity of the peak at the lowest point of the bust is much higher than at the peak of the housing bubble strengthening my point that what we observe is indeed an aggregate form of prospect theory in action (Kahneman and Tversky (1979)).

I also take a look at the dynamics of sales and inventories of existing homes, the housing prices and the BUSE index and find a narrative which sheds light into the post bubble burst dynamics. In order to better allow the dynamics to emerge I apply certain a smoothing technique to the time series by breaking each one of them into twelve month-based annual series, imputing missing values linearly in between and taking point-wise averages of all twelve series. This method allows us to plot what a trained eye sees in time series with periodicities of varying frequency and it returns quieter results than doing a month fixed effects smoothing. A certain pork-cycle-like pattern emerges (Hanau (1928)) among sales of existing homes (which we think of as a proxy for demand), their inventory (supply), S$P house prices and the BUSE index: rising sales (indicating increasing demand) pull prices up and draw sellers into the market (supply) while draining the market of prospective buyers faster than of prospective sellers. The sales peak first, the prices peak subsequently in tandem with the bottoming out of the BUSE index while the inventory peaks last and we are in a bust. In the second phase sales bottom out first followed by the prices hitting the lowest point in tandem with the peak of the BUSE index and finally a bottoming out of the inventory. The market is on the rise again. At the end of this stage it would appear as though we are getting ready for a bust.

The rest of this paper is structured as follows. In Section 2 I discuss the data sources, the data and its transformations. In Section 3 I describe the dynamics of the post bubble burst US housing market showing how the Google BUSE Index may be used to create a narrative of the dynamics of the housing market. In that section I pose the question of whether the market is about to rinse and start over entering another bust. In Section 4 I do some forecasting exercises

and in Section 5 I close with conclusions.

2. DATA

Google Trends data[4] is relative data. Within an aggregation time unit i (which can be a hour, a day or a week) we take the number x_i of searches which include our keyword of interest $\mathtt{x}$ and divide that by the total number of searches T_i in the same aggregation time unit i, so that we form x_i/T_i. Moreover if we are observing a certain time period (which can be seven days for hourly data, three months for daily data and everything since 2004 for weekly data) then $i = 1\ldots n$ for some n ($n = 7 \times 24$ in case of hourly data or about 3×30 in case of daily or the number of weeks since 2004 in case of weekly data). If then $M_n = \max_{i=1\ldots n}\{x_i/T_i\}$ then the time series we get from Google is:

$$G_i(\mathtt{x}) = \frac{100 \cdot x_i}{T_i \cdot M_n}$$

or setting $c_n = 100/M_n$

$$G_i(\mathtt{x}) = \frac{x_i}{T_i}c_n. \tag{2.1}$$

Google uses undisclosed, proprietary algorithms to classify and group searches into categories such as Travel, Real Estate, Business, Health etc. The final piece of Google trends nomenclature we need to explain in order to proceed with the description of the data is the exclusion mechanism. One can ask for all searches containing a certain keyword without searches which contain certain others; up to 30 keywords can be excluded. For examples drawing the time series for "$\mathtt{x} \quad - \mathtt{y_1} \quad \cdots \quad - \mathtt{y''_{30}}$" will produce the relative volumes of all searches which contain the word $\mathtt{x}$ without those that contain any of $\mathtt{y_1}, \cdots, \mathtt{y_{30}}$.

For obvious reasons I restrict my attention to the Google category "Real Estate". In analogy to Askitas (2015), where I looked for searches $\mathtt{yes}$ $\mathtt{-no}$ and $\mathtt{no}$ $\mathtt{-yes}$ to successfully and precisely nowcast the Greek Referendum of July 5 2015, I exploit the dichotomy between $\mathtt{buy}$ and $\mathtt{sell}$ in the Real Estate category. In other words I look for two time series:

$$\text{"}\mathtt{buy} \quad - \mathtt{sell}\text{"} \quad \text{and} \quad \text{"}\mathtt{sell} \quad - \mathtt{buy}\text{"}.$$

I thus get two time series which may be thought of as the buy and sell "buzz" (i.e. search intensities for "buy" and "sell") in the Google category "Real Estate". These time series look as depicted at the top of Figure 1.

Search intensities are vulnerable to ambient search noise and shocks from irrelevant keywords in other words from random variation of the denominator in equation (2.1) hence I will be looking at the "buy to sell ratio" just like I did with the "no to yes" ratio in Askitas (2015). In other words the series that I will form is the point-wise ratios of the BUY and the SELL series. This series has the advantage that it equals the ratio of the absolute number of buy

---

[4]The general description of the data in this section draws heavily from the data section of Askitas (2015)
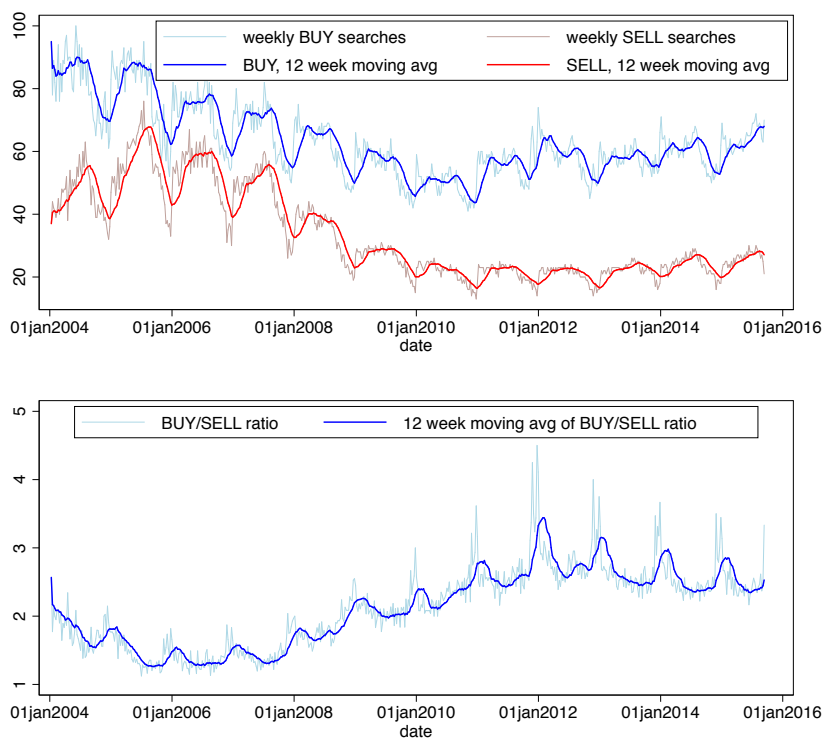
FIGURE 1.— BUY searches are "buy -sell" and SELL searches are "sell -buy". Twelve week moving averages are also displayed to better depict the trends.

The time series is aggregated and published on a weekly basis. Data Source: Google Trends (www.google.com/trends).

searches to the absolute number of sell searches in other words it is no longer vulnerable to the denominator of equation (2.1). The series and its twelve week moving average are depicted at the bottom of Figure 1.

Notice that while , in Figure 1, in both the `sell` and the `buy` searches we observe seasonal Christmas lows the ratio peaks. In other words during low relative volumes for buy and sell we have more prospective buyers than sellers.

Another version of what we discussed so far can be drawn with `buy`, `sell` and `build` and by forming the ratio of `buy` searches to the sum of `sell` and `build` searches. The ratio drawn in this way has a better correlation with the Shiller House Price Index.

I now use the thirty keywords exclusion option in Google Trends to provide support for the plausibility of my identification strategy. In the next session I will compare the buy to sell ratio with housing prices and that will of course be the ultimate test. By successively excluding terms we can have a good picture of the sort of searches which contain buy or sell in the Real Estate category. The results are in Figure (2). By the additional keywords it can be seen that we can reasonably hope that "buy searches" broadly identify (house) buyers and that "sell searches" broadly identify (house) sellers. The order by which we subtract keywords is significant. Earlier terms have a larger share in the respective searches.
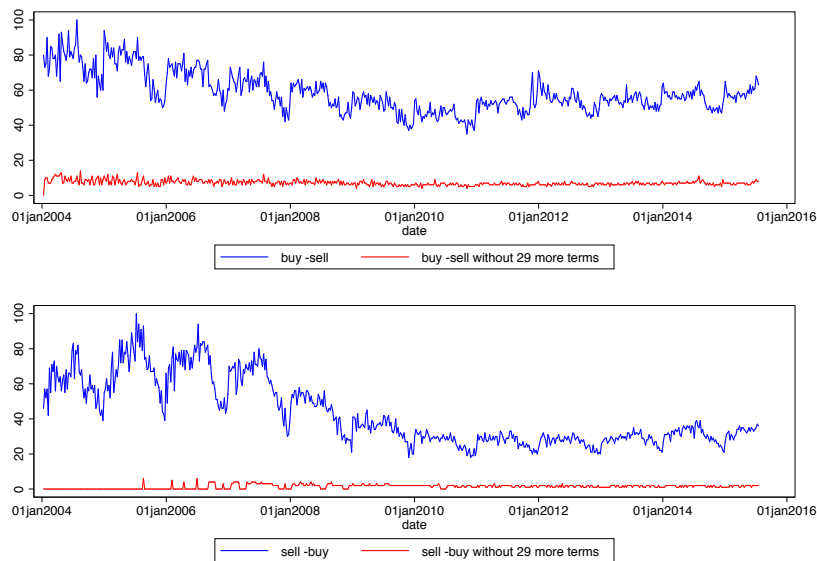

FIGURE 2.— The breakdown of the buy (top) and the sell (bottom) searches by additional keyword.

BUY, terms excluded: -sell -house -houses -home -homes -property -estate -apartments -owner -condo -land -rent -things -when -timeshare -best -apartment -condos -short -foreclosures -looking -flats -townhouse -timeshares -foreclosure -orlando -cabins -before -case.
SELL, terms excluded: -buy -to -house -how -home -homes -estate -help -timeshare -houses -assist -my -land -sale -property -short -condo -condos -share -timeshares -apartments -you -rent -online -u -shares.
Data Source: Google Trends (www.google.com/trends) and own calculations.

Finally notice that using the buy to sell ratio q we can now find the shares of buyers and sellers, as in Askitas (2015) in the space of the buy and sell searches as follows. The percentage of buyers is $100q/(1+q)$ and the percentage of sellers is give by $100/(1+q)$.

In conclusion this paper's identification strategy for the choice of keywords is to first choose the Google Trends category "Real Estate" to establish relevance to the housing market, to then look at buy and sell searches therein and by excluding terms establish that most if not all of the searches are made from buyers and sellers respectively. The ratio is now a ratio of prospective buyers to sellers.

3. BOOM BUST DYNAMICS IN THE POST BUBBLE BURST TIME

Admittedly one would have had to wait a very long time in order to see variable dynamics in the US housing market as the S&P/Case-Shiller U.S. National Home Price Index rose more than 7 fold from 25.18 in February of 1975 to its peak of 184.62 in August of 2006. But we are in the post bubble burst era and the dynamics are there and depict remarkable regularity which I organised in Figure 4 and I now want to describe. In order to take seasonal variation and random noise out I apply a certain smoothing to all series which captures the intuition that a trained eye applies to such series by ignoring seasonal variations in order to see the trend. For each series $S = (S_i : i = 1, \ldots, n)$ I create 12 sub-series $S^j = (S_i^j : i = 1, \ldots, n)$ one for each month $j = 1, \ldots, 12$. Each S^j is formed from S as follows. First I restrict S to the j-th month with missing values everywhere else. Then I fill in the missing values by linear imputation between border values. Finally I take point-wise averages to form the smoothened series. In Figure 3 I demonstrate this for the ratio of buy to sell searches and the months j=6, 12 i.e. I smoothen using only 2 instead of 12 months in order not to clatter the graph.
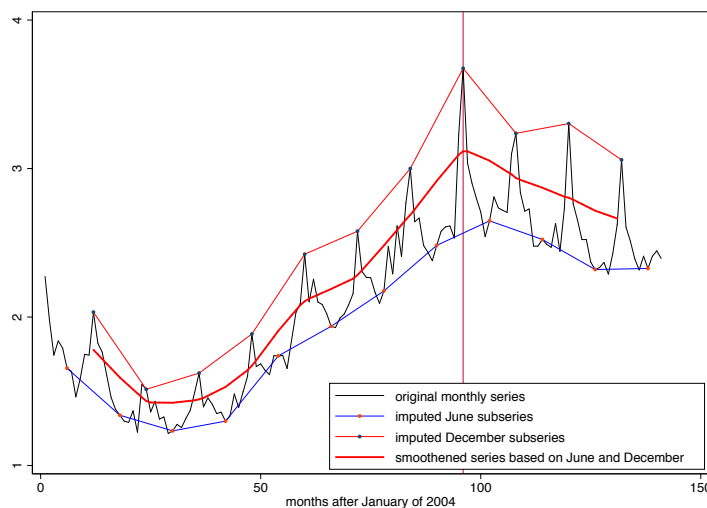

FIGURE 3.— Smoothing the monthly BUSE index using June and December values

Data Source: Google Trends (www.google.com/trends) and own calculations.

I take smoothings of the S&P/Case-Shiller U.S. National Home Price Index and the Google BUSE index (top), the inventory of existing homes on sale and the sales of existing home (middle) and the probability of a house on sale to be sold (bottom). The latter probability is simply the ratio of sales to inventory.
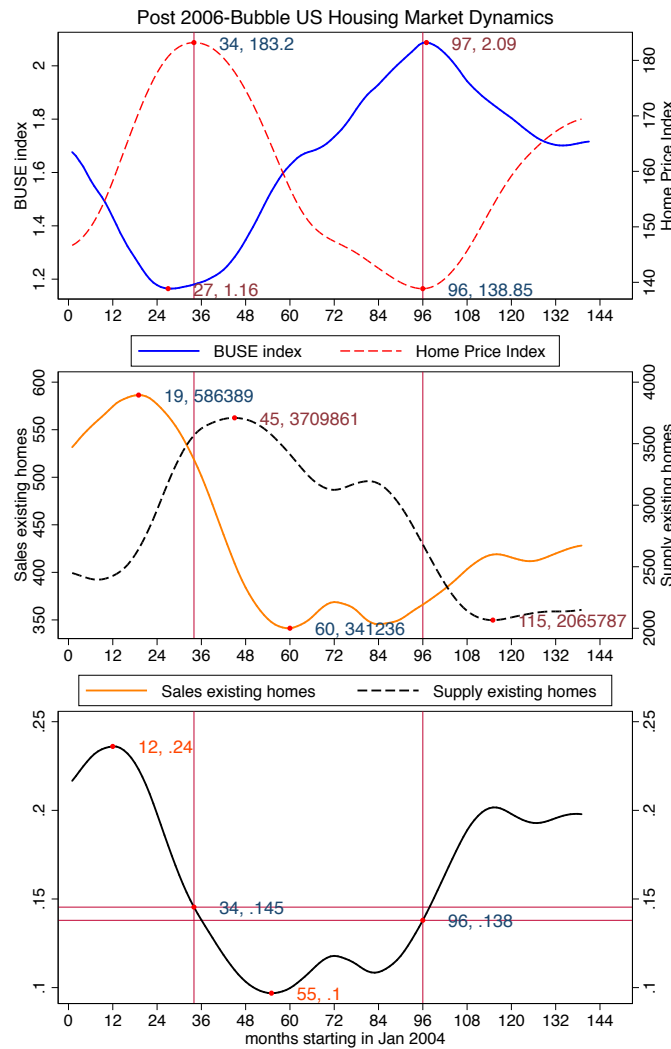

FIGURE 4.— The dynamics of prices and market participation

The series have been smoothened in order to better recognise the cyclical pattern.
Data Source: Google Trends (www.google.com/trends), FRED (research.stlouisfed.org), S&P Dow Jones Indices (us.spindices.com) and own calculations.

Notice how the Google BUSE index moves counter to price (i.e. prospective buyers are being diluted in the pool of prospective market participants during price hikes and their concentration increases on falling prices) and how the two indices reach their (opposite) local extrema simultaneously and of course turn around in tandem. We notice that when the sale probability

is below .138/.145 we have decreasing prices and increasing BUSE index while above .138/.145 we see the opposite. All in all a remarkably consistent and regular picture.

I distinguish three phases. In the first phase we have increasing prices with the Shiller National House Price Index peaking at 183.2 points in Oct of 2006. In the second phase we observe decreasing prices with the Shiller House Price Index bottoming out at 138.85 points in December 2012. In the third phase house prices return to an increasing trend which continues to date although it seems to be slowing down.

The turnaround of the S&P Index is preceded by the turnaround of sales and succeeded by the turnaround of supply while the threshold for a house price turnaround appears to be around probability of sale of .138/.145 (bottom of Figure 4). Finally increasing prices are accompanied by a decreasing share of prospective buyers in the pool of prospective market participants (top of Figure 4) while decreasing prices happen at the same time as it becomes increasingly likely to find a prospective buyer in the pool of prospective market participants. Notice that house prices and the BUSE index reach their opposite local extrema almost simultaneously with the BUSE index bottoming out 5 months in advance of the prices peak. Notice that the BUSE index is measured before the Shiller price index is made known (2 month lag) and that therefore it is the BUSE index which shapes the prices and not the other way around. In fact the BUSE index is the aggregate expression of real time market dynamics whose expression in the S%P price is made known with a delay of two months. Of course market participants (at least those who just bought a home) are known to actually do know current price trends (Case et al. (2012)) as they only exaggerate their 10 year expectations.

In the first phase a high probability of sales indicates increasing demand which drives prices up and leads to increasing supply of houses on sale. At the same time though prospective buyers become more and more rare in the pool of prospective market participants which sets the stage for a price bust. The bust then comes when there is 1.23 prospective buyers for every seller or builder. When prospective buyers become sufficiently rare i.e the BUSE index reaches a minimum the market can no longer sustain its price level. When the prices are dropping the supply of houses on sale is decreasing and the sales are also decreasing. Prospective buyers proliferate setting the stage for a price stabilisation and turnaround. First the sales turn around and when there are 2 prospective buyers per seller or builder the prices are climbing again. It is also the point when the probability of a sale breaks through its critical threshold. This means real market conditions are such that one "feels" an improvement in the chances of selling a house on sale and also senses the number of prospective buyers is increasing hence sellers are starting to become more demanding. In the third phase we have climbing prices again sales and in particular supply of houses on sale is recovering extremely slowly. This may be due to the fact that many of the owners who would like to sell are still under water.

4. NOWCASTING

There are several approaches to dealing with mixed frequency data, in this case weekly Google Trends data and monthly home sales, home supply and Shiller index. We choose the simplest one by reducing the higher frequency data to the lowest one. We do this by taking the weekly Google Trends series and averaging out by month. This method is viable for a

TABLE I

Nowcasting the Shiller US Home Price Index using Google Trends

	P	$\tilde{P}$	ΔP	$\Delta \tilde{P}$
	coef./p-value	coef./p-value	coef./p-value	coef./p-value
B	−37.269***			
	(.000)			
$\tilde{B}$		−44.064***		
		(.000)		
p			21.321***	
			(.000)	
$\tilde{p}$				18.769***
				(.000)
const.	219.719***	230.921***	−3.149***	−2.821***
	(.000)	(.000)	(.000)	(.000)
Adj. R^2	.658***	.798***	.592***	.872***
No. of cases	139	139	138	138

BUSE Index

forecasting practitioner as well since we can have a month measurement as soon as we have at least one weekly measurement in it. Figure 4 suggest one should estimate at least two models. If P_i is the monthly National home Price Index (and $\tilde{P}_i$ is its smoothing from Section 3), B_i is the monthly BUSE index[5] (and $\tilde{B}_i$ is its smoothing) and p_i is the monthly probability that a house will be sold conditional on it being on sale (and $\tilde{p}_i$ is its smoothing) then one should write down and estimate two equations:

$$P_i = \alpha B_i + \beta \quad \text{and} \quad \Delta P_i = \gamma p_i + \delta \tag{4.1}$$

The first one of these equations is based on the observation that the S&P Index is strongly and inversely correlated with the BUSE index and the second equation expresses what one observes in Figure 4 namely that, analytically expressed $\frac{dP}{dt} \cdot (p - \mu) > 0$ for some μ close to .14 or so.

---

[5]To be more precise I first reduce the `buy` and `sell` Google series to monthly ones and then take 3 month moving averages since the S&P Index is also a three month moving average

I estimate the equations once for the three month moving averages and once for the smoothened series and the results of these regressions are listed in the table below. Notice that in the third model I estimate an equation $\Delta P = 21.374p - 3.155$ which can be rewritten as $\Delta P = 21.374(p - .14760924)$ allowing me to recover the turnaround threshold seen in Figure 4. Interestingly about 60% of the variance of ΔP is explained by $p - .14760924$. In the smooth version in model 4, 87.3% of the variance of $\Delta \tilde{P}$ is explained by $\tilde{p} - .15$. It is this last equation which made this observation possible which I do not know to have been made before. The first two models are the ones which convince us that the BUSE index will be interesting to monitor at least in the years ahead.

5. CONCLUSIONS

I used the ratio of buy to sell Google searches in the Google category Real Estate and showed that one can thus nowcast the S&P Home Price Index by means of the BUSE index. The index can also be used to better understand the dynamics of supply and demand in the US housing market. Prices are formed based on beliefs, expectations and a host of intangibles. In a highly connected world these often spread in an epidemiological manner. They are shaped by the aggregate buzz of an always on ambient backdrop of pessimism or optimism. Fundamental factors like mortgage interest rates, underwriting standards, short terms interest rates etc also influence the market of course but ultimately for any values of these one can observe how prices create ambient sentiment and how the latter feeds into the market and its price formation processes. I am of course just painting a macro picture because I only have access to aggregate data but one can only imagine the deep and profound insights of market behaviour which could be gained with access to micro data where the techniques in Varian (2014) could come to use.

The Google BUSE index explains about 70% of the housing price variation and I am aware that if it becomes part of the toolkit of market participants it will become just another factor shaping strategic behaviour in that market. That may well change its effectiveness but it will certainly make for a more informed understanding of market dynamics and applicable strategies and may help home buyers and sellers better understand the often seemingly puzzling market dynamics.

REFERENCES

ASKITAS, N. (2015): "Calling the Greek Referendum on the Nose with Google Trends," *SSRN*.

ASKITAS, N. AND K. F. ZIMMERMANN (2011): "Detecting Mortgage Delinquencies," *IZA Discussio Paper Series, No 5895*.

——— (2015): "The internet as a data source for advancement in social sciences," *International Journal of Manpower*, 36, 2–12.

CASE, K. E., R. J. SHILLER, AND A. K. THOMPSON (2012): "What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets," *Cowles Foundation Discussion Paper*.

CHOI, H. AND H. VARIAN (2012): "Predicting the Present with Google Trends," *Economic Record*, 88, 2–9.

HANAU, A. (1928): "Die Prognose der Schweinepreise," *Vierteljahrshefte zur Konjunkturforschung*, 47.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

VARIAN, H. R. (2014): "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28, 3–28.