

## **Data Shop**

*Data Shop*, a department of Cityscape, presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to [david.a.vandenbroucke@hud.gov](mailto:david.a.vandenbroucke@hud.gov) for consideration.

---

# **Trend-Spotting in the Housing Market**

Nikos Askitas

Institute for the Study of Labor

---

## **Abstract**

*I create a time series of weekly ratios of Google searches in the United States on buying and selling in the real estate category of Google Trends, whereby I call this ratio the Google U.S. Housing Market BUSE index, or simply the BUSE index.<sup>1</sup> It expresses the number of “buy” searches for each “sell” search, which I consider to be a good proxy of the number of prospective homebuyers for each prospective homeseller in the pool of prospective housing market participants by means of certain regularity assumptions on the distribution of Internet users. The BUSE index—which can be perceived as a behavioral macroeconomic indicator—has several unique, desirable properties, which make it useful for understanding and nowcasting the U.S. housing market. It has a significant correlation with the Standard & Poor’s/Case-Shiller® U.S. National Home Price Index. Because the latter is monthly and is published as a 3-month moving average with a 2-month lag and the Google Trends data are weekly, the result is a short-term nowcast of housing prices in the United States. I show how these Google data can be used to create a consistent narrative of the post-bubble-burst dynamics in the U.S. housing market and propose the BUSE index as an instrument for monitoring housing market conditions in real time.*

---

<sup>1</sup> BUSE index = buyers sellers index.

## Introduction

The U.S. housing market is the subject of much research for many good reasons. A house is simultaneously an asset and a home. As an asset, it is related to a homeowner's long-term expectations and, as a home, it defines a homeowner's lifestyle and forms his or her life attitudes. As a commodity, a home is related to a large supply chain of construction materials and home equipment, and it generates a significant number of jobs in construction, maintenance, and sales, to name a few. For these reasons, a home is also often an instrument of government intervention to the economy as a whole, a fact that contributes to the inherent endogeneity in the formation of home prices. For these reasons, conventional economic wisdom is not entirely unfounded in maintaining that *as the housing market goes, so goes the rest of the economy.*

The U.S. housing market certainly has no shortage of house price indices, an additional fact that underlines the importance of this market—which then begs the question, “Why would we need another one?” To answer this question and to explain my choices in this article, I first briefly discuss the available indices.<sup>2</sup> This housing market has five main housing price indices, two of which are so-called *median house price indices*, with the other three being *repeat sales indices*. The former type comprises the indices of the National Association of Realtors® and the U.S. Census Bureau, and the latter type comprises an index by the Federal Housing Finance Agency (FHFA) and two proprietary indices—one is from CoreLogic, Inc. (CoreLogic), and the other is the well-known Standard & Poor's (S&P)/Case-Shiller® U.S. National Home Price Index (Case-Shiller index).

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, but the repeat sales indices use a prior sale as a proxy for the hedonic value. The FHFA index is a repeat sales index, albeit based only on sales of houses securitized by Fannie Mae and Freddie Mac, but the Case-Shiller and CoreLogic indices are based on “arm’s length transactions,” with the CoreLogic index being slightly broader. I chose the Case-Shiller index because the data are readily available on the S&P website.

With all these indices, then, why do we need yet another index? If the buying thoughts of prospective homebuyers or the selling thoughts of prospective homesellers can in some way be captured in real time, we should be able to monitor housing market conditions regardless of the fundamentals that may be driving the housing market. The result would be a behavioral housing market index. A simple ratio of the number of buying thoughts to the number of selling thoughts should indicate something about the formation of upcoming home prices. Underwriting standards, interest rates, mortgage rates, lending trends and practices, the inflow of foreign capital, the prevalence of securitization of mortgages, government programs for affordable housing, tax incentives for homebuyers, labor market conditions: whatever the fundamentals are at each point in time, they should shape and, in fact, be captured by that buy-to-sell ratio. The answer to the question, “Why yet another index?” consequently is that I seek to construct an index that is “buzz based” (that is, based on search intensities for “buy” and “sell”) and contains the fundamentals of what later becomes price. It is also for this reason that I think arm’s length transactions are better suited as a target variable.

---

<sup>2</sup> The information here largely comes from the Federal Reserve Bank of St. Louis website (<https://www.stlouisfed.org/on-the-economy/2015/january/the-differences-between-house-price-indexes>).

To construct such an index, however, raises another question: 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 in each era, although 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” (Shiller, 2015: 101), and the telephone played a role in the “volatile stock market of the 1920s” (Shiller, 2015: 181). The stock market boom of the 1990s was similarly accompanied by another technological revolution: the advent of the Internet. Social media, as we know it today, has brought on the era of a more intensive “interpersonal contagion of ideas” (Shiller, 2015: 182). Therefore, it is not farfetched to search for market clues in Internet data, especially at a time when virtually every market has an online component (Askitas and Zimmermann, 2015). Askitas and Zimmermann (2011) showed how mortgage delinquency rates in the U.S. housing market might be usefully nowcast around the most recent economic crisis by looking at Google Search intensities for the term *hardship letter*. In this article, 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 searches are the buying and selling thoughts of prospective market participants. I thus obtain an index, which I call the Google U.S. Housing Market BUSE Index, or simply the BUSE index. It captures the relative proliferation of prospective homebuyers to prospective homesellers in the pool of prospective buyers and sellers (that is, the pool of all prospective housing market participants).<sup>3</sup>

Because, eventually, the efficient market hypothesis weighs in on asset prices, predicting the (far) future is a futile exercise. Hence, I do not claim to predict future prices but rather simply nowcast their formation in the present in a way that has behavioral underpinnings and enables a better understanding of market behavior. In the literature using 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 Varian and Choi (2012). The novelty of this article is that I use the Google Trends category Real Estate but take the ratio of buy-to-sell searches therein (the BUSE index). This technique has previously been used only in Askitas (2015), with very good results. In appendix A, I discuss the data and provide support for my identification strategy.

In the post-2006 bubble-burst U.S. housing market, the BUSE index strongly and negatively correlates with housing prices as expressed by the Case-Shiller index, reflecting the main result of this article (see exhibit 1 in the next section). As prices increase in the boom phase (which can unfortunately be observed only since January 2004), prospective homebuyers are increasingly diluted in a pool in which homesellers proliferate, thus setting the stage for a downturn. When the conditional probability that a house for sale will be sold reaches a trigger threshold (which I estimate around 15 percent), 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 accelerate their entry in the pool only after buyers start slowing down due to the

<sup>3</sup> Adding “build” searches appears to sharpen the results, which I believe is because builders may be former buyers, and data on building permits show that the building of new homes is currently on the rise.

high prices, whereas in a bust, sellers leave at accelerating rates after buyers start slowing down their exit. A more technical way of stating this is that the percentage change of buyers and sellers is related, whereby, when one reaches its local extremum, the other changes concavity.

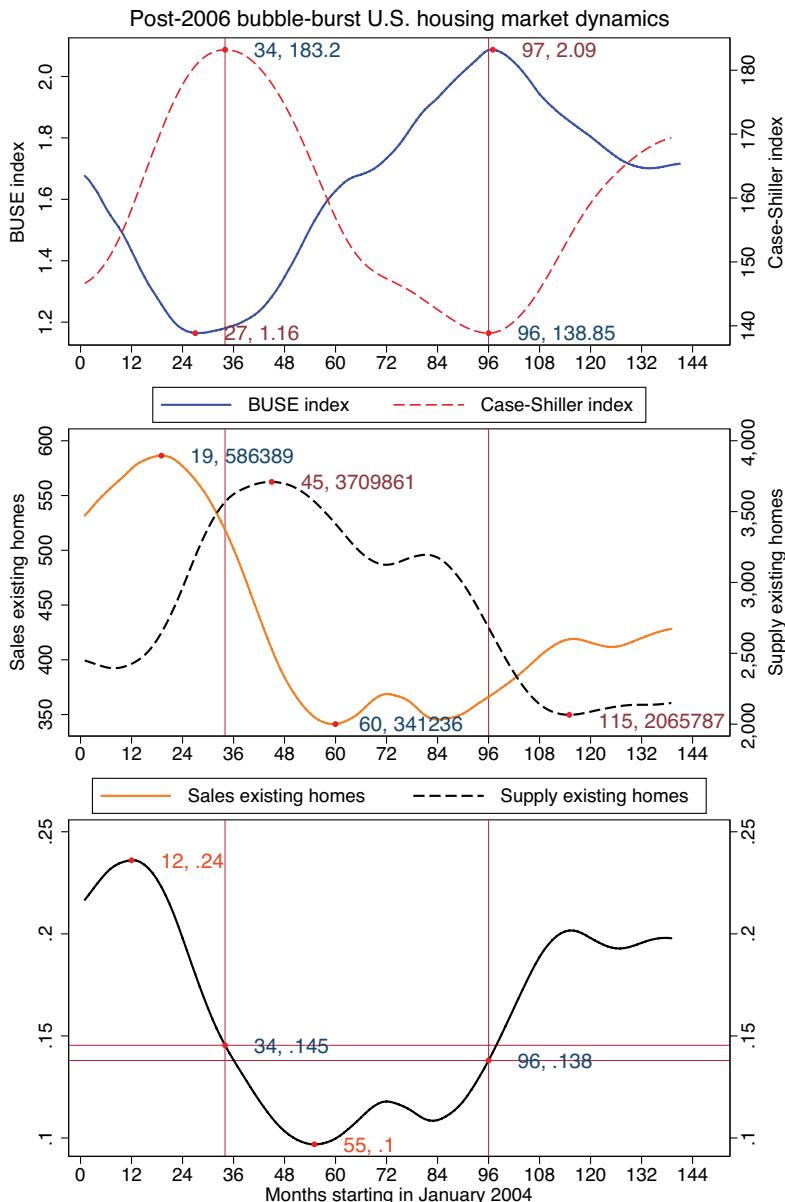
Considering the seasonal properties of this ratio, I observe that, although the relative intensities of both buy and sell searches diminish at the end of the year, the ratio of homebuyers to homesellers (as expressed by the ratio of corresponding searches) exhibits a peak, a phenomenon that I believe to be consistent with prospect theory, which postulates that—all other things kept equal—a loss hurts more than a comparable gain gratifies. The trough in both buy and sell searches means that housing market participation is viewed or felt as incompatible with the family-centered, hedonic bliss of the holiday season. The fact now that the buy-to-sell ratio spikes indicates 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, thus strengthening my point that this observation is indeed an aggregate form of prospect theory in action (Kahneman and Tversky, 1979).

I also examine the dynamics of sales and inventories of existing homes, the housing prices, and the BUSE index, finding a narrative that sheds light on the post-bubble-burst dynamics. To better allow for the dynamics to emerge, I apply a certain smoothing technique explained in appendix B. A certain pork-cycle-like pattern emerges (Hanau, 1928) among sales of existing homes (which I think of as a proxy for demand), the inventory of homes for sale (supply), the Case-Shiller index, 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 homebuyers faster compared with prospective homesellers. In the first phase, the sales peak first, before the prices subsequently peak in tandem with the bottoming out of the BUSE index, while the inventory peaks last and the market busts. 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 booming again. At the end of this phase, it would appear as though the market is getting ready for a bust.

The rest of this article is structured as follows. In the next section, I describe the dynamics of the post-bubble-burst U.S. housing market, showing how the BUSE index may be used to create a tight narrative 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 the Nowcasting section, I undertake some forecasting exercises informed by the dynamics section, before I close with conclusions in the final section.

## Describing Housing Market Dynamics With Google Search

One admittedly would have had to wait a very long time to see variable dynamics in the U.S. housing market, as the Case-Shiller index rose more than sevenfold from 25.18 points in February 1975 to a peak of 184.62 points in August 2006. We are in the post-bubble-burst era, however, and the dynamics are now there, depicting remarkable regularity, as organized in exhibit 1, which I describe in the next paragraph.

**Exhibit 1****The Dynamics of Prices and Market Participation**

BUSE index = buyers sellers index. Case-Shiller index = Standard & Poor's/Case-Shiller® U.S. National Home Price Index.

Notes: The series have been smoothed to better recognize the cyclical pattern by a method described in appendix B.

Series used: Case-Shiller index and Google BUSE index (top), the inventory of existing homes for sale and the sales of existing homes (middle), and the probability of a house for sale being sold (bottom). The latter probability is simply the ratio of sales to inventory.

Data sources: Google Trends (<http://www.google.com/trends>); FRED (<http://www.research.stlouisfed.org>); S&P Dow Jones Indices (<http://www.us.spindices.com>); author's calculations

Notice how the Google BUSE index moves counter to price (that is, prospective homebuyers are being diluted in the pool of prospective market participants during price hikes and their concentration increases on falling prices) and also how the two indices reach their (opposite) local extrema simultaneously and, of course, turn around in tandem. Notice that when the sale probability is below .138 to .145, prices are decreasing and the BUSE index is increasing, but when the probability is above .138 to .145, the opposite occurs: a remarkably consistent and regular picture.

I distinguish three phases. In the first phase, prices are increasing, with the Case-Shiller index peaking at 183.2 points in October 2006. In the second phase, decreasing prices are observed, with the Case-Shiller 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 Case-Shiller index is preceded by the turnaround of sales and succeeded by the turnaround of supply, but the threshold for a house price turnaround appears to be around a probability of sale of .138 to .145 (bottom of exhibit 1). Finally, increasing prices are accompanied by a decreasing share of prospective homebuyers in the pool of prospective market participants (top of exhibit 1), and 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 also that the BUSE index is measured before the Case-Shiller index is made known (2-month lag) and thus the BUSE index shapes the prices, rather than the other way around. In fact, the BUSE index is the aggregate expression of real-time market dynamics whose expression in the Case-Shiller index is made known with a delay of 2 months (Shiller and Case, 2012). Of course, market participants (at least those who have just bought a home) are known to actually know current price trends (Shiller and Case, 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 an increasing supply of houses for sale. At the same time, prospective homebuyers become increasingly rare in the pool of prospective market participants, which sets the stage for a price bust. The bust subsequently comes when the market has 1.23 prospective homebuyers for every homeseller. When prospective buyers become sufficiently rare (that is, the BUSE index reaches a minimum) the market can no longer sustain its price level. When the prices are dropping, the supply of houses for sale decreases, as do the sales. Prospective buyers proliferate, setting the stage for a price stabilization and turnaround. First, the sales turn around and, when the market has two prospective buyers per seller or builder, the prices climb again. Furthermore, this is also the point at which the probability of a sale breaks through its critical threshold. This point means that real market conditions are such that one senses an improvement in the chances of selling a house for sale and also senses that the number of prospective buyers is increasing; hence, sellers start to become more demanding. In the third phase, prices are climbing again; sales and also (and in particular) supply of houses for sale are recovering extremely slowly. This slowness may be due to the fact that many of the owners who would like to sell remain under water; that is, they have mortgage loans with balances that are higher than the fair market value of the property.

## Nowcasting

Several approaches can be taken to deal with mixed-frequency data—in this case, weekly Google Trends data and monthly home sales, home supply, and the Case-Shiller index. I choose the simplest one by reducing the higher-frequency data to the lowest one, by taking the weekly Google Trends series and averaging out by month. This method is also viable for a forecasting practitioner, who can have a month's preliminary measurement as soon as he or she has at least one weekly measurement within it. Exhibit 1 suggests that one should estimate at least two models.

If  $P_i$  is the monthly Case-Shiller index (and  $p_i$  is its smoothing),  $B_i$  is the monthly BUSE index<sup>4</sup> (and  $b_i$  is its smoothing), and  $Q_i$  is the monthly probability that a house will be sold conditional on its being for sale (and  $q_i$  is its smoothing), then one should write down and estimate two equations,

$$P_i = \alpha B_i + \beta \quad (1)$$

and

$$\Delta P_i = \gamma q_i + \delta \quad (2)$$

Equation (1) is based on the observation that the Case-Shiller index is strongly and inversely correlated with the BUSE index, and equation (2) expresses what one observes in exhibit 1, namely, that, analytically expressed—

$$(dP/dt)(Q-m) > 0 \quad (3)$$

for some  $m$  close to 0.14 or so. I estimate the equations once for the 3-month moving averages and once for the smoothed series, whereby the results of these regressions are listed in exhibit 2.

### Exhibit 2

#### Nowcasting the Standard & Poor's/Case-Shiller® U.S. National Home Price Index Using Google Trends

	$P$ Coefficient/p-Value	$p$ Coefficient/p-Value	$\Delta P$ Coefficient/p-Value	$\Delta p$ Coefficient/p-Value
$B$	– 37.269*** (.000)			
$b$		– 44.064*** (.000)		
$Q$			21.321*** (.000)	
$q$				18.769*** (.000)
Constant	219.719*** (.000)	230.921*** (.000)	– 3.149*** (.000)	– 2.821*** (.000)
Adjusted $R^2$	.658***	.798***	.592***	.872***
Number of cases	139	139	138	138

\*\*\*  $p < 0.001$ .

$b$  = BUSE smoothing.  $B$  = BUSE index.  $p$  = Case-Shiller smoothing.  $P$  = Case-Shiller index.  $q$  = smoothing of  $Q$ .  $Q$  = probability of a house to be sold conditional on its being for sale.

<sup>4</sup> More precisely, I first reduce the buy and sell Google series to monthly ones and then take 3-month moving averages because the Case-Shiller index is also a 3-month moving average.

Notice that in the third model, I estimate an equation  $\Delta P = 21.374 Q - 3.155$ , which can be rewritten as  $\Delta P = 21.374(Q - .14760924)$ , allowing me to recover the turnaround threshold seen in exhibit 1. It is interesting that about 60 percent of the variance of  $\Delta P$  is explained by  $Q - .148$ . In the smooth version (fourth column), about 87 percent of the variance of  $\Delta p$  is explained by  $q - .15$ . As far as I know, this observation is new. The first two models are those convincing me that the BUSE index will be interesting to monitor at least in the years ahead.

## 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 Case-Shiller index by means of the BUSE index. The index can also be used to better understand the dynamics of supply and demand in the U.S. housing market. Prices are formed based on beliefs, expectations, and a host of intangibles, which, in a highly connected world, often spread in an epidemiological manner and 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-term interest rates, and so on, also influence the market, of course, although, ultimately, for any values of these, one can observe how prices create ambient sentiment and also how the latter feeds into the market and its price-formation processes. This article can also be seen as using Internet data to study the effect of policy on market behavior and its endogeneity. Of course, I simply aim to portray a macro picture because I have access to only aggregate data, although one can only imagine the deep and profound insights into market behavior that could be gained with access to search micro data, in which such techniques as mentioned in Varian (2014) could come to use.

The Google BUSE index explains about 70 percent of the housing price variation, and I am aware that, if it becomes part of the toolkit of market participants, it will simply become another factor for shaping strategic behavior in that market. Although using the BUSE index for shaping strategic market behavior may well change its effectiveness, it will certainly provide a more informed understanding of market dynamics and applicable strategies and may help homebuyers and homesellers better understand the often seemingly puzzling market dynamics.

## Appendix A. Google Trends Data

Google Trends<sup>5</sup> data<sup>6</sup> are relative data. Within an aggregation time unit  $I$  (which can be an hour, a day, or a week), I take the number  $x_i$  of searches that include the keyword of interest  $x$  and divide it by the total number of searches  $T_i$  in the same aggregation time unit  $i$ , whereby I form  $x_i/T_i$ . Moreover, if observing a certain time period (which can be 7 days for hourly data, 3 months for daily

---

<sup>5</sup> Google Trends data are available at <https://www.google.com/trends/>.

<sup>6</sup> The general description of the data in this section draws heavily from the data section of Askitas (2015).

data, and everything since 2004 for weekly data), then  $i=1 \dots n$  for some  $n$  ( $n=7\times24$  in case of hourly data, about  $3\times30$  in case of daily data, or the number of weeks since 2004 in case of weekly data). If  $M_n=\max\{x_i/T_i; i=1 \dots n\}$ , then the time series obtained from Google is—

$$G_i(x)=(100\times x_i)/(T_i\times M_n) \quad (\text{A-1})$$

Or, setting  $c_n=100/M_n$ ,

$$G_i(x)=(x_i c_n)/T_i. \quad (\text{A-2})$$

Google uses undisclosed, proprietary algorithms to classify and group searches into categories such as Travel, Real Estate, Business, and Health. The final piece of Google Trends' nomenclature that I need to explain to proceed with the description of the data is the exclusion mechanism. One can ask for all searches containing a certain keyword without searches that contain certain other words, whereby up to 30 keywords can be excluded. For example, drawing the time series for  $x-y_1, \dots -y_{30}$  will produce the relative volumes of all searches that contain the word  $x$  without those that also contain any of  $y_1, \dots y_{30}$ .

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

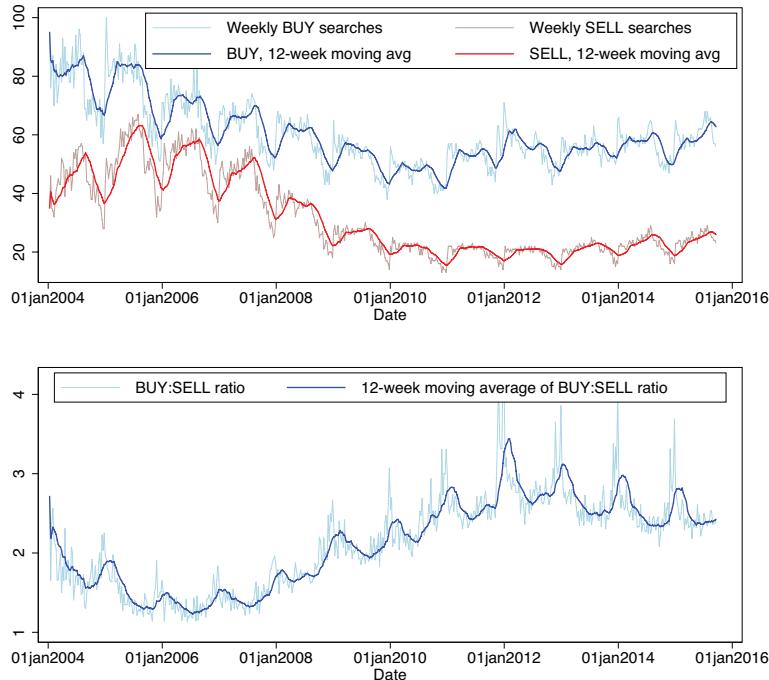
“buy -sell” and “sell -buy.”

I thus obtain two time series that may be thought of as the buy and sell in the Google category Real Estate. These time series look as depicted at the top of exhibit A-1.

Search intensities are vulnerable to ambient search noise and shocks from irrelevant keywords: in other words, from random variation of the denominator in the equation that defines  $G_i$ ; hence, I will be looking at the buy-to-sell ratio just like I did with the no-to-yes ratio in Askitas (2015). Put differently, the series that I will form is the point-wise ratio of the BUY and the SELL series. This series has the advantage that it equals the ratio of the absolute number of buy searches to the absolute number of sell searches as thus it is no longer vulnerable to the denominator  $T_i$ . The series and its 12-week moving average are depicted at the bottom of exhibit A-1.

Notice that although seasonal year-end lows occur in both the sell and buy searches, in exhibit A-1, the ratio peaks: in other words, during low relative volumes for buy and sell, we have more prospective homebuyers than homesellers.

Another version of what I have discussed thus 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 Standard & Poor's/Case-Shiller® U.S. National Home Price Index, and I leave this alternative specification as an exercise for the interested reader.

**Exhibit A-1****Buy and Sell Buzz in the Google Category Real Estate**

Notes: Buy and sell search intensities (top) and their ratios (bottom). Buy searches are "buy -sell" and sell searches are "sell -buy." To better depict the trends, 12-week moving averages are also displayed. The time series is aggregated and published weekly.

Data sources: Google Trends (<http://www.google.com/trends>); author's calculations

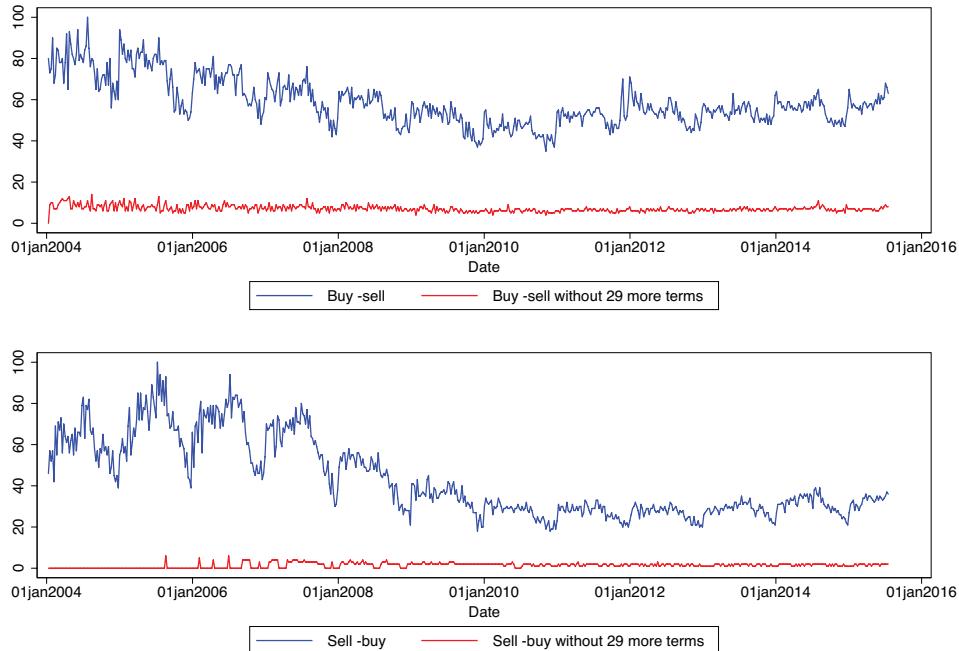
I now use the 30 keywords exclusion option in Google Trends to provide support for the plausibility of my identification strategy, although comparing the buy-to-sell ratio with housing prices will be the ultimate test. By successively excluding terms, a good picture emerges of the type of searches that contain the terms *buy* or *sell* in the Real Estate category. The results are presented in exhibit A-2. Through the additional keywords, it can be seen that it is reasonable to hope that buy searches broadly identify (home) buyers and that sell searches broadly identify (home) sellers. The order in which keywords are subtracted is significant, given that earlier terms have a larger share in the respective searches.

Finally, notice that by using the buy-to-sell ratio  $q$ , the shares of homebuyers and homesellers are observable, as in Askitas (2015), in the space of the buy and sell searches as follows. The percentage of buyers is  $100/q(1+q)$  and the percentage of sellers is  $100/(1+q)$ . This observation is simple yet important, and it is applicable whenever the topic of interest is a share (as in buyers versus sellers). In such cases, point-wise dividing Google Trends data avoids the disadvantage of these data; that is, the fact that they yield not absolute volumes but rather relative ones.

In conclusion, this article'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 before

**Exhibit A-2**

The Breakdown of the Buy (top) and the Sell (bottom) Searches by Additional Keyword



Notes: 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, -time-shares, -apartments, -you, -rent, -online, -u, -shares.

Data sources: Google Trends (<http://www.google.com/trends>); author's calculations

subsequently looking at buy and sell searches therein and by excluding terms establishing that most—if not all—of the searches are made by homebuyers and homesellers, respectively. The ratio is now a ratio of prospective buyers to sellers.<sup>7</sup>

## Appendix B. Smoothing Technique

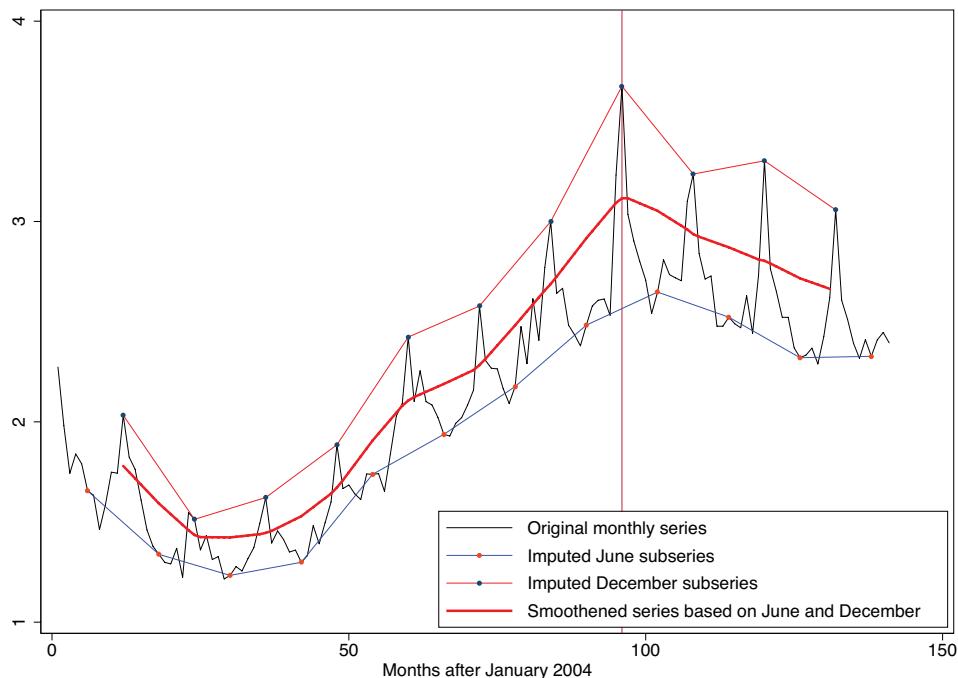
To eliminate seasonal variation and random noise, I apply a certain smoothing to all series, which captures the intuition that a trained eye applies to such series by ignoring seasonal variations to observe the trend. This method—which breaks each series into 12 month-based annual series, imputing missing values linearly in between, and taking point-wise averages of all 12 series—returns better results than undertaking month fixed-effects smoothing. I demonstrate the method in the

<sup>7</sup> According to <http://www.internetsociety.org>, as of 2013, the Internet penetration in the United States (that is, the share of U.S. residents with Internet connectivity) equaled 84 percent. The high penetration ensures that the sample of Internet users has very little space for being biased compared with the entire population. This assertion is reinforced if I am allowed to hypothesize that those who either own or can afford a home can also afford Internet connectivity.

remainder of this paragraph. For each series,  $S=(S_i: \{i=1, \dots n\})$ , I create 12 subseries  $S^j=(S_i^j: \{i=1, \dots n\})$ , one for each month  $j=1, \dots 12$ . Each  $S^j$  is formed from  $S$  as follows. First, I restrict  $S$  to the  $j$ th month with missing values everywhere else. I subsequently fill in the missing values by linear imputation between border values. Finally, I take point-wise averages to form the smoothed series. In exhibit B-1, I demonstrate this smoothing technique for the ratio of buy-to-sell searches and the months  $j=6, 12$ ; that is, I smoothen using only 2 rather than 12 months to avoid cluttering the graph.

### **Exhibit B-1**

#### **Smoothing the Monthly BUSE Index Using June and December Values**



Data sources: Google Trends (<http://www.google.com/trends>); author's calculations.

### **Acknowledgments**

The author thanks Konstantinos Tatsiramos and Christian Zimmermann for critical reading and helpful communication.

### **Author**

Nikos Askitas is the Director of Data and Technology at the Institute for the Study of Labor in Bonn, Germany.

## References

- Askitas, Nikolaos, and Klaus F. Zimmermann. 2015. "The Internet as a Data Source for Advancement in Social Sciences," *International Journal of Manpower* 36 (1): 2–12.
- Askitas, Nikos. 2015. Calling the Greek Referendum on the Nose With Google Trends. IZA Discussion Paper Series. Bonn, Germany: Institute for the Study of Labor.
- Askitas, Nikos, and Klaus F. Zimmermann. 2011. Detecting Mortgage Delinquencies. IZA Discussion Paper Series. Bonn, Germany: Institute for the Study of Labor.
- Hanau, Arthur. 1928. "Die Prognose der Schweinepreise." Dissertation published in the series *Vierteljahrsshefte zur Konjunkturforschung*. Berlin, Germany: German Institute for Economic Research.
- Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47 (2): 263–292.
- Shiller, Robert J., and Karl E. Case. 2012. What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets. Cowles Foundation discussion paper. New Haven, CT: Yale University, Cowles Foundation for Research in Economics.
- Varian, Hal, and Hyunyoung Choi. 2012. "Predicting the Present With Google Trends," *Economic Record* 88 (s1): 2–9.
- Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives* 28 (2): 3–18.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.