

WORKING PAPER NO: 537

**Does Internet Search Intensity Predict House Prices in
Emerging Markets? A Case of India**

Ekta Jalan*

*Research Assistant,
Real Estate Research Initiative,
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
ekta.jalan02@gmail.com*

Venkatesh Panchapagesan*

*Associate Professor
Finance and Accounting
Head, Real Estate Research Initiative,
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
Ph: 080-26993349
venky.panchapagesan@iimb.ernet.in*

Madalasa Venkataraman*

*Lead Researcher,
Real Estate Research Initiative,
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
madalasa.venkataraman@iimb.ernet.in*

Year of Publication – January 2017

* All authors are with the Indian Institute of Management, Bangalore. We are solely responsible for all errors in this paper.

Does Internet Search Intensity Predict House Prices in Emerging Markets? A Case of India

Please don't circulate without permission.

Abstract

We examine whether internet search intensity, as captured by Google's Search Volume Index (SVI), predicts house price changes in an emerging market like India. Emerging markets have low internet penetration and high information asymmetry with a dominant unorganized real estate market. Like in developed markets such as the US, we find that internet search intensity strongly predicts future house price changes. A simple rebalancing strategy of buying a representative house in the city with the greatest change in search intensity and selling a representative house in the city with the smallest change in search intensity each quarter yields an annualized excess (over risk-free government t-bills) return of 4 percent.

Keywords: Google Trends Search Volume Index, House Price Index, Search Intensity, Housing price forecasting, India

1. Introduction

Much has been written about what drives asset prices, especially on the role of information and its ease of availability to market participants. With the explosion of internet as a medium to collect and disseminate information today, it is not hard to imagine its pivotal role in improving the price discovery process for assets ranging from consumer goods to financial and real assets. Recent literature suggests that prices in both primary (Da et al., 2011) and secondary markets (Joseph et al., 2011) of common stocks, and in foreign exchange markets (Goddard et al, 2015) are impacted by internet search volumes. Even in physical markets such as real estate, Beracha and Wintoki (2013) find that internet search intensity explains house price movements, not only over time, but also in the cross-section by capturing supply constraints well.¹

Most of these studies are set, however, in the context of developed countries with large internet penetration and low cost of technology. Little is known about the role of internet on asset prices in developing countries where information asymmetry is high and the cost and availability of technology limits its usage only to a privileged few. We address this gap in this study.

Using data from India, a fast growing developing country, we examine whether house prices are impacted by internet searches of prospective buyers and sellers. We use Google's Search Volume Index, or SVI, to capture internet search intensity. Our study covers *four* large metropolitan cities over 24 quarters between 2011 and 2016. Despite large structural differences compared to developed countries, especially in the levels of internet penetration and in the entrenched power of physical intermediaries such as brokers, we find that internet searches do impact house prices in India.

Our results are interesting as it sheds light on the nature and role of the internet as an infomediary even in emerging markets. Internet impacts asset prices in two distinct ways - by reducing transaction costs, and by reducing information asymmetry. Zumpano et al. (2003) shows that internet reduces across-period search costs leading to lower time to market, but also increases search intensity through lower within-period search costs (more properties are covered in a shorter duration due to lower transaction costs). In this case, the internet ***reduces transaction costs of search*** through technological innovations such as targeted locality searching, searching on

specific parameters, 3D view of the insides of the housing unit, neighborhood quality ratings, comparison shopping for financial, legal services as well as interiors, furnishings and other allied activities. Given high fixed costs of searching such as transportation costs, we expect that search costs could be significantly reduced for technology induced searches than physical searches in emerging markets.

The other role of the internet is that of *reducing information asymmetry*. Developed countries have readily available price information on previous sale prices and appraisal values in the public domain - in land registries as well as through informed broker intermediaries. However, in many developing countries including India, price transparency is abysmally low. Land register prices are usually below actual market prices as buyers underreport to avoid taxes. Similarly, buyers face high degree of adverse selection as sellers or the developers have much more information about the real estate product being sold like quality of construction and legality of approvals than the buyer. Lacunae in statutory clearances, municipal violations, and other seller specific risks can all be uncovered through a search algorithm or through online property portals where prior buyers share experiences and information.

The internet search process would, therefore, not only be a powerful "intention aggregator" (Batelle, 2005) capturing sentiments of buyer demand in the market, but also an "information aggregator" collating diverse information across a wide variety of sources. Our expectation would, therefore, be that search intensity could be a strong predictor of house prices even in an emerging market like India.

In contrast, it is possible that internet penetration is limited and the quality of information in the internet is dubious so as to make internet less relevant in the purchase or sale of the largest asset that a typical Indian household owns. In such cases, the dependence on entrenched market intermediaries such as brokers, who help in hyper-local demand and supply matching, could continue to be higher making internet searches less relevant in the setting of house prices.

Whether internet searches help predict house prices in emerging markets such as India is, therefore, an empirical question. Our study is an attempt to answer this question. The degree of such an impact could, however, be muted by the fact that a large part of

searches in India could be to “ping” for information with no real intent to transact. Both Beracha and Wintoki (2013) and Joseph et al. (2011) argue, using behavioral reasons, that buyers are more likely to seek information than sellers since the search process is led by buyers rather than by sellers who hold the asset. Information about buyers, and thereby, information about latent demand, may have direct impact on prices. In developing markets where price transparency is limited, both buyers and sellers may seek to cut their search costs by participating in the internet based search process - sellers, for benchmarking their prices and rents, and buyers, for seeking suitable properties. In this case, the impact of internet search intensity on housing prices is likely to be lesser since it does not only capture the demand-push for the real estate product.

Consistent with Beracha and Wintoki (2013), we find that online search intensity in the previous quarter, derived from terms suggesting intent to sell or buy, predicts house price changes in the following quarter in the four top metropolitan cities in India. However, the magnitude of the impact in India is almost *eight* times what was reported by them for the US market. We find that a 1% increase in the lagged search intensity for sale leads to a 0.132% abnormal quarterly return in house prices, compared to 0.0167% in their paper.

Our other results are also in line with Beracha and Wintoki (2013). For example, we find that the search intensity from quarters other than the immediately preceding quarter does not predict house price changes, suggesting that the time between searching and transacting is relatively short for prospective Indian home buyers. Also, we do not find any appreciable impact of rental search intensity on house price changes. Unlike in developed markets, there is sufficient anecdotal evidence that Indians have a cultural bias against renting and rental yields are quite low despite housing shortage.² The usual tradeoff between buying and renting is largely at a nascent stage in India, suggesting that rental demand may not help predict house prices.

In addition, to estimate the economic impact of our result, we construct a simple long-short portfolio where we buy a representative house in the city with highest search intensity and sell one in the city with the lowest search intensity in any given month. We show that this portfolio, which is immune to market wide effects on house prices,

generates a quarterly return of 2.93%, or 12.2% annually, over our sample period. To put this in perspective, the average risk-free return from investing in Indian government 91-day T-bills was around 8% between 2011 and 2016.³

The remainder of the paper is organized as follows: Section 2 provides a review of extant literature, while Section 3 describes internet usage in India along with details on its impact on the real estate industry. Section 4 raises the main empirical question that we are seeking to answer while Section 5 discusses data sources, details about the analysis and the policy recommendations that follow our findings. Section 6 concludes.

2. Extant Literature

Research on housing price changes falls broadly into two main categories – research on the impact of house price changes on the macro economy such as on economic output, employment and on inflation (see Forni et al., 2003, and Das et al., 2015), and research on factors that drive house price changes in the first place such as changes in local demand, credit conditions, property or locational attributes (see Hill, 2012, for a survey of hedonic model approaches to value residential housing). While the former focuses on housing being a leading indicator of the real sector of the economy, the latter tries to understand buyer behavior and what they are willing to pay for in a house. Our paper belongs in the latter category as we try to forecast house price changes using internet search intensity of home buyers and sellers.

Variables that have been used before to predict house price changes include valuation ratios (like rent-to-price and income-to-price ratios), local economic variables (like employment rate, income growth, construction costs), demographic trend variables (population growth), local space market variables (like housing starts, vacancy rates, transactions volume and proxies for zoning restrictions), and measures of leverage and monetary policy action and asset returns as REIT returns (see Ghysels et. al, 2012).

Studies on forecasting prices using buying and selling intent are prevalent in the financial asset pricing literature. Market makers infer information from buy and sell order flows in many market microstructure models in securities markets transactions. While order flows are a tangible action, intent is inferred from internet search models on buyer and seller pressures. Searching for information is considered a pre-cursor to

an action, and various studies have established search conditions to be significantly related to underlying phenomenon or to actions.

The first study to forecast using Internet searches was by Ettredge et al. (2005). Their study showed that the online searches extracted from Wordtracker's Top 500 keyword report helped predict the unemployment rate in the US. Choi and Varian (2009, 2012) used Google Search Insights data as a predictor of several economic indicators such as initial claim for unemployment, dynamics of retail sales, home sales, automobile demand and vacation destinations. Askitas and Zimmermann (2010), D'Amuri and Marcucci (2010) and Suhoy (2009) also examined unemployment in the US, Germany and Israel with Google Search Insights. Vosen and Schmidt (2011) created another index from online search query time series data provided by Google Trends on private consumption while Radinsky et al. (2009), Huang and Penna (2009) and Preis et al. (2013) used internet searchers to measure consumer sentiment.

Similarly, in financial markets, Da et al. (2011), predict stock returns and trading volume based on search intensity for stock tickers. Joseph et al. (2011) capture investors' sentiment from online searches and correlate it with stock volatility. Guzman (2011) forecasts inflation with a 12 months' lag period that outperformed three other popular survey-based inflation predictors. Vlastakis and Markellos (2012) model the fluctuation of stock prices using buy and sell search intent. According to their results search queries with information seeking attitude is positively correlated to the risk, i.e., the volatility of the stock. Smith (2012) studies the predictability of foreign currency market using internet search-volume indices and Preis et al., (2013) find a significant relationship between trading behavior and search-volumes.

In the real estate market, Wu and Brynjolfsson (2009) were the first to derive a housing search index from Google Trends data using appropriate keywords to predict housing sales and prices. They posit that a one percent increase in search frequency leads to 67,700 future quarterly additional house sales. They find a significant positive relationship between the FHFA-provided HPI (house price index) and current search intensity and volumes, but also state that lagged relationship between these variables is negative due to supply constraints. The authors apply a seasonal autoregressive model with Google Trends categorical data and record improvements of the mean square error

as in Choi and Varian (2012). They find the prediction error term reduces by a factor of four when search volume intensity is used in the calculations.

Beracha and Wintoki (2013) study about 200 cities in the US to predict whether abnormal house searches can predict future abnormal price changes. They use HPI data provided by FHFA in an autoregressive model with factors including the effects of city, time and population growth along with search volume data as an independent lagged variable. They also show correlation between search volume and prices turn around with higher lag period and the effect is more prominent in areas with a lower land supply. Also, house prices are more responsive towards an 'uptick' rather than a 'downtick' in searches, proving that prices are sticky. They find that 'hot' areas (areas with rapid price appreciation) are more sensitive to search intensity as compared to other areas.

In commercial real estate, Dietzel, et al., (2014) create a sentiment based indicator using Google SVI to forecast future prices of commercial real estate. Real-estate related search volume data is also used in studies that predict variables other than house prices. Das and Ziobrowski (2015) study the effects of online searches on stock price movements of listed Indian real estate companies using different estimation methods such as OLS and GARCH. They find that search volume index of certain keywords like 'Home Finance' is positively correlated to the volatility of the index comprising real estate stocks in India. Rochdi and Dietzel (2015) show in their study that online information seeking behavior can predict future price movements of US REITs.

3. Internet and the Indian Real Estate Market

3.1 Internet Usage in India

India is among the top three countries in the world when it comes to the absolute number of internet users, indicating the sheer size of its population. As of July 2016, India had about 462 million internet users, more than the population of the USA, and second only to China's total of 721 million users.⁴ The growth in internet usage has been phenomenal as more than 52 million users, or 17%, were got first time access in the initial 6 months of 2016.

However, the internet penetration (usage relative to the size of the population) still remains low when compared to other developed countries. According to the Internet Live Stats (www.internetlivestats.com), India ranks 127th among 201 countries in internet penetration. Only a little more than a third of its population (35%) have internet access, as compared to 88% in the US and 52% in China. In fact, it has the lowest penetration among all the BRICS countries that it gets compared with.⁵ Also, the average monthly online usage in India was only 10 hours, as compared to 34 hours in the US and 19 hours in China.⁶

3.2 Growth of Internet Property Portals

The Indian real estate market, both organized and unorganized, is estimated to be around \$126 billion in 2015, and is expected to grow to \$853 billion by 2028.⁷ Much of the market remains unorganized with a loose network of brokers and builders at the hyper local level. However, the official data on the exact share is more than a decade old.⁸ But even then, the share of estimated gross value added by economic activities attributed to the unorganized real estate sector was declining, from 77.25% in 1995-2000 to about 63.44% in 2004-2005, showing that the sector was getting organized rapidly.

Some of this structural change can be attributed to the role of internet in aiding and improving the search for and information about properties. Real estate developers have a large web-presence, and property portals have achieved spectacular success in India, with over 40 country-wide portals currently focusing on providing property specific information on renting, buying, leasing, selling and a host of other real estate related services besides. Since 2007, the listings of properties for rental and for sale on these websites have proliferated, benefiting consumers enormously through increased transparency on pricing and features of properties.

A recent Google study in 2014 estimates that real estate transactions worth about \$43 billion were being directly influenced by the internet in India.⁹ Moreover, the study finds that real estate related online searches have increased three-fold in India between 2011 and 2013, and as much as 62 percent of real estate related searches were done on aggregator websites such as the property portals. The study also estimates that the

trend is only going to exponentially increase as more than half a billion people in India will get access to the internet by 2017.

Table I lists the top 5 property portals in India in 2015. The internet-enabled property services sector has also been continuously innovating, with newer products and features being introduced at a very rapid pace. Competition within this sector has intensified, and venture capitalists have been flocking to provide more money muscle to the well-entrenched players.¹⁰ The Google study, cited earlier, indicates that as much as 60% of internet users want access to in-depth property information and market trends while 52% want ease of comparison across properties.

Table I: Top Indian Property Portals in 2015

S No	Names	Past 30 Day traffic (in million)	Alexa Traffic Network (Rank across all Indian websites)	Bounce Rate (Rate at which an user comes and leaves without exploring the website fully)
1	99acres	8.4	144	32.50%
2	India Property	8.1	173	26.30%
3	Magicbricks	7.9	150	41.70%
4	Commonfloor	6.4	176	30.70%
5	Housing.com	4.4	114	19.50%

Source: Alexa.com; authors' compilations.

4. Empirical Question

A low internet penetration would suggest that search intensity using Google may not adequately capture real estate demand, and hence predict house prices, in India. Moreover, unlike Da et al. (2011), one cannot make the assumption that internet searches are driven solely by individual investors in Indian real estate markets. Other stakeholders such as developers and real estate brokers are also known to use internet portals to promote or to market their product. Some of them also use internet to collect market intelligence and size up the demand. Therefore, not all searches are likely to convey information on actionable purchase intent, and hence about future price

changes.¹¹ All this could make search intensity a very noisy signal for demand, and less likely to predict house prices.

In contrast, there are strong reasons why search intensity may be highly correlated with prices in India. Internet plays a dual role in emerging countries – it is a channel for reducing transaction costs (thereby increasing depth as well as breadth of search (Zumpano et al., 2003) and for reducing information asymmetry. Search costs are reduced as buyers don't have to travel long distances or spend hours in traffic to see properties. Property portals with state-of-the-art technology provide prospective buyers with most of the information related to the property such as locality attributes, amenities, construction design choices etc. These portals also reduce 'information costs' by aggregating data related to the builder such as credit rating, past performance, construction completion track record etc.

So, unlike developed markets such as the US, it is not clear whether internet search intensity will be an important factor in predicting house prices in India. We seek to address this gap using data. Specifically, we study the relationship between internet search intensity using Google's SVI and abnormal house price changes. We focus on abnormal price changes to rule out changes that impact the broader real estate market in general.

Like Beracha and Wintoki (2013), we separate the effects of searches with an intent to buy or sell from searches with an intent to rent. High rental demand can indicate a future drop in prices as prospective buyers postpone buying in favor of renting. Our empirical tests controls for exogenous variables such as population and returns on the stock market. We also perform a Granger causality test to ensure that SVI is impacting house price changes and not vice versa.

To estimate the economic magnitude of the predictive power of SVI, we construct a long-short portfolio where we buy a representative house in the city with the highest SVI in a given quarter and sell a representative house in the city with the lowest SVI in the same quarter. We then track the performance of this portfolio in the following quarter assuming no transaction cost of buying and selling.

5. Data and Analysis

Our main data for this study are: Google Search Volume Index (SVI) to determine internet search intensity and *two* housing price indices - Housing Price Index, or HPI, computed by India's central bank (the Reserve Bank of India) and Residex, computed by India's mortgage regulator (the National Housing Bank). We use data for *four* of the largest cities in India - Mumbai, Delhi, Bangalore, and Chennai - for the period between June 2011 and September 2015, a total of 17 quarters. This is slightly less than 29 quarters used by Beracha and Wintoki (2013) for their US study.

Our sample cities account for almost 30 percent of India's population. They also have the highest population densities and have a substantial pent-up demand for housing.¹² Internet usage is also among the highest in the country with all 4 cities among the top 5 in internet penetration.¹³ Table II provides macroeconomic data about our sample cities.

Table II: Macroeconomic Data on Sample Cities

City Name	GSDP (2013-14)	% of India's GDP	Population (2011 census)	% of India's Population (2011 census)	Number of Households (2011 census)	% India's Households (2011 census)	Average increase in HPI (5 years)
Delhi (NCR)*	391125	3.73	167,87,941	1.39	34,35,999	1.46	5.1%
Mumbai	15,10,132	14.42	30,85,411	0.25	6,74,339	0.29	3.8%
Chennai	8,54,238	8.16	46,46,732	0.38	11,54,982	0.49	3.4%
Bengaluru	6,14,607	5.87	96,21,551	0.80	23,93,845	1.01	3.7%

Source: Census data from <http://censusindia.gov.in/2011-prov-results/indiaatglance.html>; GSDP data compiled from www.data.gov.in.

* NCR, or the National Capital Region, includes all the major suburban development around Delhi.

5.1 Google Search Volume Index (SVI)

Google Search dominates the Indian search engine market. Table III presents the market share for various search engines as of 2014.

Table III: India Search Engine Market Share, Usage Statistics - 2014

Rank	India	Market Share (%)
1	Google	95.3
2	Yahoo	1.8
3	Bing	1.4
4	Ask Jeeves	1.0
5	Webcrawler	0.1

Source: alexa.com

We use the search volume index data that is based on query engines derived from Google Trends. Google Trends provides the search volume index (SVI) as a 'query share' - a particular value between 1 to 100 for specific input keyword searches for a given time-period and at a particular location. The value is derived by taking the percentage of the keyword searches relative to the total volume of searches and then normalizing it by the highest query share of that keyword in the same time-period and location. The higher the value of the SVI, the greater is the intensity of the search among internet users. A decrease in SVI, however, does not imply that web-search for a particular term has decreased; it simply implies that the popularity (or query share) of a particular term has declined.

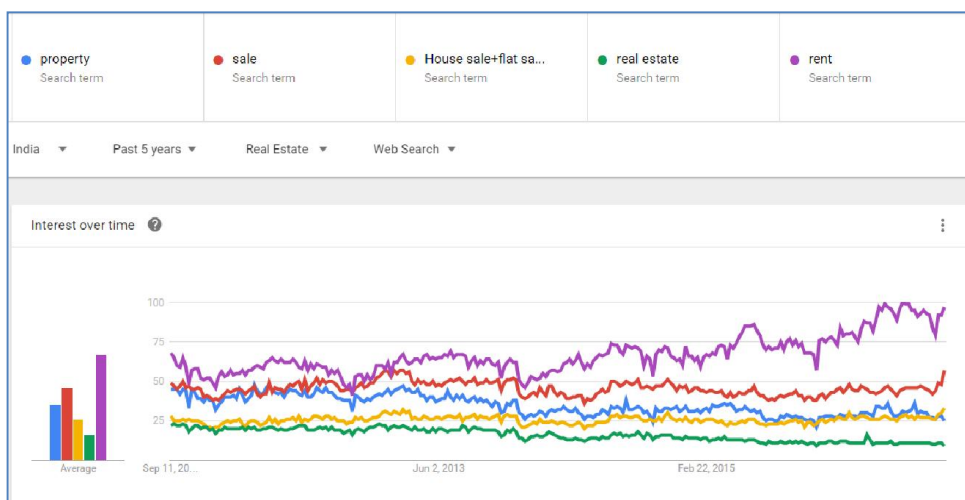
What keywords we use is hugely critical to the usefulness of the SVI measure for the analysis. So much of our effort was spent trying to mimic the common expressions that Indians use while searching for property online. We started with logical keywords like "flat" or "apartment" to capture the users' intent for buying houses. Unlike in Beracha and Wintoki (2013) where the index was built using "real estate+ MSA name", we observed that, in India, people were more likely to use the terms "house", "flat" or "apartment" rather than "real estate" when they are searching for a property online. We also used the term "sale" and "rent" along with the City name to capture trends in searches at a particular city. To get all the searches together for these several terms the '+' sign has been used as it signifies 'or' in Google trends to combine the results for all keywords.

In order to filter certain topics, Google provides certain categories as well as subcategories for specific searches. The search engine not only tracks single search query but also tracks queries that users conduct before and after (Dietzel et al., 2014).

In addition to building the search index using keywords we also use the subcategory “Real estate” provided by Google. We use popular single search terms and combined them into one search term to build the index.

Figure 1 presents search trends for different keywords published by Google Trends. The keywords, “rent” and “sale” dominate among real estate searches. We, therefore, chose “City Name Property” + “City Name +Sale” to compute sale SVI, and “City Name Property” + “City Name +Rent” to compute rent SVI.

Figure 1: Google Trends for Different Keywords



Source: Authors' compilation.

We obtain monthly SVI data for every combination of identified keywords between 2011 and 2016. This data is collected for the four cities that are included in our sample. We roll up the data at the quarterly level to coincide with the frequency of the Housing Price Indices available in India.

5.2 House Price Indices

There are *two* main house price indices in India. The House Price Index (HPI) is published by the Reserve Bank of India, India's central bank, and is quarterly in frequency. The HPI (base: 2010-11=100) is published for ten major cities, viz., Mumbai, Delhi, Chennai, Kolkata, Bengaluru, Lucknow, Ahmedabad, Jaipur, Kanpur and Kochi. An average house price index representing all-India house price movement is also compiled based on the weighted averages of these cities. The price represented in these

indices are based on the data of property price transactions collected from registration authorities (land registers) of respective the state governments, using a stratified weighted average measure, with administrative zones within cities being used for stratification.

An alternate house price index in India is the Residex that is published on a quarterly frequency by India's mortgage company regulator, the National Housing Bank of India (NHB). NHB is a subsidiary of the Reserve Bank of India. The Residex predates the HPI, with a base year of 2007 for the index and covers 26 cities as of date. In 2013, the NHB attempted to cover administrative zone wise/tax zone wise price changes in its computation of property values. The Residex is based on a weighted average of mortgage prices under a modified Laspeyers' approach. Unlike HPI, Residex does not provide a national average. Residex was temporarily suspended in September 2015 to provide for a methodology revamp.

We use HPI as our main index to track house prices in India for *three* reasons. First, since the Residex was discontinued in September 2015, we would have to give up relevant, current data on analysis of SVI, which is likely to lead to bias especially considering the short time period of analysis. Second, aggregating Residex to a national index would be difficult since the base number of properties for each city is not known, whereas this information is readily available for national HPI. Third, it is commonly accepted that the Residex has certain methodological shortcomings (and hence its suspension pending revamp) which have been addressed in HPI.

Returns in city level HPI are differenced from the returns in national level HPI to calculate the abnormal city HPI, and the same process is performed for various lags of abnormal city level HPI.

Equation 1

$$R_{i,t} = \left(\frac{\text{Price}_{i,t} - \text{Price}_{i,t-1}}{\text{Price}_{i,t}} \right) - \left(\frac{\text{Market}_{i,t} - \text{Market}_{i,t-1}}{\text{Market}_{i,t}} \right)$$

5.3 Abnormal SVI calculation

Since HPI is quarterly in frequency, we roll-up the SVI data which is obtained weekly, to a quarterly frequency. We define abnormal SVI as the difference between SVI growth in the particular city and the national level SVI Growth, similar to how we define our abnormal HPI variable.

Equation 2

$$SaleSVI_Ab_{i,t} = \left(\frac{SaleSVI_{i,t} - SaleSVI_{i,t-1}}{SaleSVI_{i,t}} \right) - \left(\frac{SaleSVI_{national,t} - SaleSVI_{national,t-1}}{SaleSVI_{national,t}} \right)$$

The choice of time period for the study is determined by the availability of the HPI and by the maturity of the online housing market in India. Google only reports SVI if the search volume for the term is non-trivial, and similar to Beracha and Wintoki (2013) we find that the cities that we focus on have SVI of adequate quantum and frequency to undertake an analysis.

We use commonly used variables that influence housing prices as controls in our estimation. These variables include the population from the Census of India, (published latest in 2011), the stock market returns at a quarterly level (NIFTY index returns, 2011-2016, quarterly), construction costs (sourced from the Construction Industry Development Council, India, city-wise) as well as unsold housing inventory with builders (sourced from data vendors).¹⁴ Unsold inventory index is the number of months that it would take to dispose of all inventory at current absorption rates. High inventory should mean lower future prices, everything else remaining constant. We include this measure despite the fact that, in Indian real estate, prices are very sticky downwards, and price drops are rare as builders prefer keeping properties vacant for a longer time in the market instead.

We expect HPI to be positively impacted by population and construction costs and negatively impacted by unsold inventory. On stock returns, our prior is mixed. On the one hand, real estate and stocks compete for savings among investors and could therefore be negatively related. On the other hand, increase in stock returns could provide more liquidity in the hands of investors who could use them as down payment for a house purchase. This would suggest that they would be positively related.

5.4 Empirical Results

5.4.1 Search Intensity and Returns

The first results from our empirical analysis consider the impact of abnormal city-wise SVI and lags of abnormal city-wise HPI on the HPI of the specific city in the succeeding quarter. This ties in with our hypothesis that Google SVI is an indicator of transaction intensity and higher abnormal searches indicate a future demand for housing, leading to higher HPI in the succeeding quarter.

One significant deviation from Beracha and Wintoki (2013) is the incorporation of now-casting (contemporaneous relationship) of HPI based on current period abnormal Sale and Rental SVIs. We also include the lagged abnormal SVI for sale and rent, as well as lagged abnormal HPI to account for time-series impact of HPI (momentum in HPI). The argument is similar to Beracha and Wintoki (2013): both housing price momentum, and SVI are driven by past abnormal returns on housing, and buyer sentiment needs to be captured in the model. We include city-level fixed effects to account for unobserved factors. The Hausmann test is performed to ascertain that the model is correctly specified. Quarter dummies account for the impact of time and seasonality, if any.

Our basic regression equation is as follows:

Equation 3

$$\begin{aligned} HPIRet_Ab_{i,t} = & \alpha + \sum_{p=1}^{p=4} \beta_p HPIRet_Ab_{i,t-p} + \sum_{p=0}^{p=3} \gamma_p SaleSVI_Ab_{i,t-p} \\ & + \sum_{p=0}^{p=3} \varphi_p RentSVI_Ab_{i,t-p} + \delta Pop_Growth_{i,t} + \rho Change_CCI_{i,t} \\ & + \tau StkRet_{i,t-1} + \sum_{p=3}^{p=2} \omega_p REUnsold_mths_{i,t-p} + \eta_t + d_t + \varepsilon_{i,t} \end{aligned}$$

where $HPIRet_Ab_{i,t}$ is the abnormal return on the city i during the quarter t and other subscripts have the usual interpretations; $SaleSVI_Ab$ and $RentSVI_Ab$ refer to the

abnormal Sale and Rent searches for the keywords "Sale" and "Rent" for city i with p quarterly lags respectively; Pop_Growth is annual population growth calculated from the census population growth between 2001 and 2010 successive census; $Change_CCI$ is the prior period increase in construction cost index for city i , $StkRet$ refers to the stock returns over the previous period p , and $REUnsold_mths$ stands for the number of vacancy months in the past two quarters for city i ; η_i is the fixed effect parameter for each city; d is the time (quarter) dummy and $\varepsilon_{i,t}$ denotes the error term.

Since we have fixed-effect estimates with a short time span of 17 quarters, it is likely that any estimates that we obtain may be biased due to the presence of correlated error terms – lagged returns are mechanically correlated with errors due to the autoregressive nature of the variables in the model. The bias-corrected fixed effect estimator provided by Kiviet (1995), and Bruno (2005) is used to estimate the coefficient values. The results of the estimation of equation 3 with bias correction are shown in Table IV below.

Table IV – Estimated Coefficients for the Model that Predicts HPI with SVI

HPIRet_Ab	Coefficients (Std. Error) with bias correction
HPIRet_Ab_Lag1	-0.301 (0.132) **
SaleSVI_Ab	0.078 (0.049)
SaleSVI_Ab_Lag1	0.132 (0.067) **
SaleSVI_Ab_Lag2	0.067 (0.074)
SaleSVI_Ab_Lag3	0.064 (0.051)
RentSVI_Ab	0.01 (0.082)
RentSVI_Ab_Lag1	0.105 (0.101)
RentSVI_Ab_Lag2	0.024 (0.092)
RentSVI_Ab_Lag3	-0.007 (0.058)
Pop_Growth	0.627 (0.288) **
Q_D1	-0.007 (0.022)
Q_D2	-0.001 (0.02)
Q_D3	-0.004 (0.02)
Change_CCI	-0.53 (0.698)
HPIRet_Ab_Lag2	-0.289 (0.126) **
HPIRet_Ab_Lag3	-0.145 (0.125)
StkRet	-0.001 (0.142)

REUnsold_mths	-0.001 (0.001)
REUnsold_mths_Lag1	0.001 (0.001)

** significant at 5% confidence level; * significant at 10% confidence level

R-square:

Fixed effects (within) regression = 0.3365

Fixed effects (between) regression = 0.2415

As expected, population growth is statistically significant and positively associated with rise in housing prices. Though we expected construction cost index to be important to house prices, it is not the case from the data. One of the reasons could be that construction cost has a lagged impact on house prices due to the long lead time between construction and a transaction. However, longer lags for construction costs also turn out to be insignificant. Similarly, unsold months' inventory is expected to have a negative impact on house prices from literature, and we find that the direction of impact is correct, though the variable is insignificant statistically, both at current and at lagged levels. Stock return, either as a direct growth variable, or in its log transformation, was not found to significantly influence house price indices; the directionality between Abnormal HPI returns and stock returns is negative, indicating that as stock returns rise, the investment in housing is diverted to financial markets, easing up demand pressure on the housing market, leading to lower HPI.

Our main variable of interest, the lagged indicator of abnormal searches from Google trends (lagged abnormal Sale SVI), is significantly and positively correlated to the abnormal House Price Index Returns for the city, after accounting for momentum in house prices as well as momentum in SVI, and adjusting for factors that are likely to lead to a secular increase in house prices such as population, construction costs, etc. Our results mirror those of Beracha and Wintoki (2013) who find that the abnormal search intensity for the term "real estate" is statistically and positively correlated with next quarter abnormal home price. None of the other Abnormal Sale SVI terms are correlated to the Abnormal HPI Returns in this study, though all the terms are in the right direction.

Recognizing that the SaleSVI_Ab is a normalized variable (between 0 and 100 by design), the positive coefficient $\gamma_1 = 0.132$, associated with the first lag of the abnormal Sale SVI, indicates that a 1% change in the lagged search intensity for sale leads to a

0.132% abnormal increase in house prices in the following quarter. This means that a city that moves from the 25th to 75th percentile in search intensity will experience a 29% increase in house prices, after accounting for changes in population and other factors. Beracha and Wintoki (2013) find a more subdued increase of 3.4% suggesting that information about latent demand in developed markets is less asymmetric and that internet usage differences is just one of the many ways to resolve such asymmetry. However, in India, where information asymmetry in housing markets are very high, internet search intensity plays a far more important role in providing information about the latent housing demand.

Abnormal Rent SVI is expected to have a negative sign for two reasons. It is expected that since renting is a substitute to home ownership, rental activity would depress house prices by substituting for house purchase demand. Rental search, may, in some cases, also be a leading indicator of increasing house prices. This is because abnormal rent search is also a measure of buoyancy of the economy; where economic activity goes up, rental searches go up since most workers in the short term would not look for home-ownership. We find that abnormal rent search is insignificant in all lags, and only the third lag in the bias corrected model indicates the right sign. Here too, our results are consistent with Beracha and Wintoki (2013) who find noisiness in rental SVI's forecasting ability.

An interesting result is that the lagged HPI abnormal return (prior period abnormal return on HPI) is negative and significant, leading to a mean-reverting behavior of HPI. Mean-reverting stock returns is well-established in mainstream finance literature. In the housing market, literature suggests that while price momentum is quite marked in the short term with a \$1 real, constant quality, housing price change leading to a 70-80 cent increase in the next year (Case and Shiller, 1989; Cutler et al., 1991), it is likely to mean revert in the next five years (Capozza et al, 2002; Glaeser, 2014). We find support for the mean-reversion, however, in Nathanson (2014), who shows that mean reversion occurs if there is positive one-time shock, leading to increased population, construction costs and prices. Mean reversion occurs when there is a sudden spurt in construction and prices, followed by a dampened effect of prices, exacerbated by construction finance availability.

To further test the economic significance of our results, we create a long-short portfolio of the 'winners' and the 'losers', where cities with high search intensity growth are considered winners and cities with low search intensity growth are termed losers. That is, we buy a representative house in the city with the highest search intensity growth and sell a representative house in the city with the lowest search intensity growth. We do both "trades" in the same quarter and rebalance it in the next quarter. If the return on such a portfolio measured in terms of HPI is 'abnormally' good, then the SVI can be taken as a good indicator of HPI returns. We use a quarterly position on the trade to analyse the returns of this portfolio. Note that this approach is just to provide broad indications of the forecasting ability of SVI and is not representative of actual trading strategy as it does not consider transaction costs.

We find that our long-short portfolio provides a quarterly average return of 2.93%, or 12.2% annually. A naïve strategy where each of the markets get a 1/4 weight yields only 0.4% per quarter. Our long-short portfolio return provides an excess return of more than 4% over the risk-free government treasury bills, indicating that lagged SVI does indeed predict house prices well.

5.4.2 Granger Causality Test

As a test for robustness, we confirm the direction of causality using the Granger Causality Test (Granger, 1969). A time-series variable X *Granger-causes* another time-dependent variable Y if predictions of the value of Y based on lagged values of X and Y outperform predictions based on lagged values of Y alone. The Granger-causality test therefore incorporates elements of directionality (X occurs before Y) and predictability (lagged values of X statistically influence Y).

The formal Granger-causality tests are based on a panel Vector-Auto-regression (VAR) modeled on the earlier equation.

Equation 4

$$\begin{aligned} HPIRet_{Ab_{i,t}} = & \alpha + \sum_{p=1}^{p=4} \beta_p HPIRet_{Ab_{i,t-p}} + \sum_{p=0}^{p=3} \gamma_p SaleSVI_{Ab_{i,t-p}} + \delta Pop_{Growth_{i,t}} \\ & + \rho Change_{CCI_{i,t}} + \tau StkRet_{i,t-1} + \sum_{p=3}^{p=2} \omega_p REUnsold_{mths_{i,t-p}} + \eta_t + d_t \\ & + \varepsilon_{i,t} \end{aligned}$$

Equation 5

$$\begin{aligned} SaleSVI_{Ab_{i,t}} = & \alpha + \sum_{p=0}^{p=3} \gamma_p SaleSVI_{Ab_{i,t-p}} + \sum_{p=1}^{p=4} \beta_p HPIRet_{Ab_{i,t-p}} + \delta Pop_{Growth_{i,t}} \\ & + \rho Change_{CCI_{i,t}} + \tau StkRet_{i,t-1} + \sum_{p=3}^{p=2} \omega_p REUnsold_{mths_{i,t-p}} + \eta_t + d_t \\ & + \varepsilon_{i,t} \end{aligned}$$

where the variables are as defined earlier in equation (3). The test is conducted using the bias-correction approximation as described earlier in this paper, as well as in Beracha and Wintoki (2013).

Our results, shown in the Table V below, confirm that the search intensity Granger-causes the abnormal HPI returns, and the direction is clearly from the search intensity towards the abnormal HPI. Columns (1) and (2) show the F -statistics (p-value in brackets) results of the hypotheses tests that the coefficients of equation (4) and equation (5) are zero respectively. The null hypothesis for equation (4) is that the lags of search intensity in the return VAR is zero. The corresponding null hypothesis for equation (5) is that the lags of abnormal return on the search intensity VAR is zero. Based on the Column (1), it is clear that we can reject (at a one percent level of significance) the null that lags of search intensity do not predict current abnormal returns.

There is a weak feedback loop that is apparent in terms of Abnormal HPI Granger-causing the search volume index too, however, in all cases, this relationship is overshadowed by the SVI's impact on the HPI. Therefore, we can conclude that the

direction of predictive causality proceeds from SVI to Abnormal HPI returns of the city and not vice-versa.

Table V: Granger Causality Tests

Lags (k)	<i>H₀</i> : Search intensity does not cause abnormal returns (Equation 4)	<i>H₀</i> : Search intensity does not cause abnormal returns (Equation 5)
	F-Stat (p-val)	F-Stat (p-val)
1	77.8 ***	3.2**
2	13.78**	2.67*
3	7.62*	0.96 (0.67)
4	3.94**	0.004 (0.9)

Note: *significant at 10% level, **significant at 5% level, ***significant at 1% level

6. Conclusion

This paper empirically addresses the question whether changes in internet search intensity predicts house prices even in developing countries like India. Unlike in developed countries where this relationship has been proven to be strong (see Beracha and Wintoki, 2013), it is not clear such a relationship would hold in countries where internet penetration is low and information asymmetry and the role of unorganised participants are high. Using data on Google’s Search Volume Index (SVI) for four Indian cities and their corresponding House Price Index values, we examine whether abnormal SVI (growth in search intensity normalized by the national average) impacts abnormal house prices (house price change normalized by the national average).

Our results confirm existing literature for developed markets that internet search intensity does predict house price changes in India. A simple rebalancing strategy of buying a representative house in the city with the greatest change in search intensity and selling a representative house in the city with the smallest change in search intensity each quarter yields an annualized return of 12 percent, or 4 percent in excess of the risk-free government treasury bills.

References

- Askitas, N. & Klaus F. Zimmermann. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, vol. 55(2), pages 107-120.
- Batelle, J. (2005). The search. *How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Portfolio.
- Beracha, E., & Johnson, K. H. (2012). Lessons from over 30 years of buy versus rent decisions: is the American dream always wise? *Real Estate Economics*, 40(2), 217-247.
- Beracha, E., & Wintoki, M. B. (2013). Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research*.
- Bruno, G. (2005). *Estimation and inference in dynamic unbalanced panel data models with a small number of individuals*. Università commerciale Luigi Bocconi.
- Capozza, D. R., Hendershott, P. H., Mack, C., & Mayer, C. J. (2002). *Determinants of real house price dynamics* (No. w9262). National Bureau of Economic Research.
- Case, K. E. and Shiller, R. J., (1989). The Efficiency of the Market for Single-Family Homes. *American Economic Review*, 79(1), 125 – 137.
- Choi, Hyunyoung, and Hal Varian. (2009). Predicting initial claims for unemployment benefits. *Working paper*.
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1), 2-9.
- Contribution of the Unorganized sector to GDP Report of the Sub Committee of a NCEUS Task Force Working Paper No 2 share of labour in unorganised RE is 81.4% (http://mospi.nic.in/mospi_new/upload/nsc_report_un_sec_14mar12.pdf).
- Cutler, David M., James M. Poterba, & Lawrence H. Summers. (1991). Speculative Dynamics. *Review of Economic Studies*, 58(3): 529–546.
- Da, Z.H.I., Engelberg, J. & Gao, P., (2011). In Search of Attention. *The Journal of Finance*, 66, 1461-1499.
- Das, P., & Ziobrowski, A. (2015). The Relationship between Indian Realty Stocks and Online Searches. *Journal of Emerging Market Finance*, 14(1), 1-19.
- D'Amuri, F., & Marcucci, J. (2010). 'Google it!' Forecasting the US Unemployment Rate with a Google Job Search Index.
- Dietzel, M.A., Braun, N., Schäfers, W., (2014). Sentiment-based commercial real estate forecasting with Google Search volume data. *Journal of Property Investment & Finance* 36 (6) 540–569.

-
- Dillman, D. A. (2011). Mail and Internet surveys: The tailored design method--2007 Update with new Internet, visual, and mixed-mode guide. *John Wiley & Sons*.
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11), 87-92.
- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*, 50(6), 1243-1255.
- Ghysels, E., Plazzi, A., Torous, W. N., & Valkanov, R. I. (2012). Forecasting real estate prices. *Handbook of Economic Forecasting*.
- Glaeser, Edward L., Joseph Gyourko, Eduardo Morales, & Charles G. Nathanson. (2014). "Housing Dynamics: An Urban Approach." *Journal of Urban Economics*, 81: 45-56.
- Goddard, J., Kita, A., Wang, Q. (2015). Investor Attention and FX Market Volatility. *Journal of International Financial Markets, Institutions and Money* 38, September, 79-96.
- Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, Vol. 37, No. 3, pp. 424-438.
- Guzman, G.C. (2011), "Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations", *Journal of Economic and Social Measurement*, Vol. 36 No. 3, pp. 119-167.
- Hill, Robert J. (2012). Hedonic Price Indexes For Residential Housing: A Survey, Evaluation and Taxonomy. *Journal of Economic Surveys*.
- Huang, H., & Penna, N. D. (2009). Constructing consumer sentiment index for US using Google searches. *University of Alberta, Department of Economics*.
- Joseph, K., Wintoki, M. B. and Zhang, Z., (2011). Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search. *International Journal of Forecasting*, 27, 1116 - 1127.
- Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics*, 68(1), 53-78.
- The IAMA & KPMG Report "India On The Go - Mobile Internet Vision Report 2017".
- Nathanson, C. G. (2014). *Mean Reversion in Housing Markets* (Doctoral dissertation), (https://dash.harvard.edu/bitstream/handle/1/12274618/Nathanson_gsas.harvard_0084L_11404.pdf?sequence=1)
- Preis, Tobias, Helen Susannah Moat, and H. Eugene Stanley. (2013). "Quantifying trading behavior in financial markets using Google Trends." *Scientific reports*.
- Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions*

of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368(1933), 5707-5719.

Radinsky, K., Sagie Davidovich, and Shaul Markovitch. 2009. Predicting the news of tomorrow using patterns in web search queries. Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WI08), <http://portal.acm.org/citation.cfm?id=1487070>

Rochdi, K., & Dietzel, M. (2015). Outperforming the benchmark: online information demand and REIT market performance. *Journal of Property Investment & Finance*, 33(2), 169-195.

Smith, G. P. (2012). Google Internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, 9(2), 103-110.

Suhoy, Tanya. 2009. Query indices and a 2008 downturn: Israeli data. Technical report, Bank of Israel, <http://www.bankisrael.gov.il/deptdata/mehkar/papers/dp0906e.pdf>

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, 36(6), 1808-1821.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6), 565-578.

Wu, L., & Brynjolfsson, E. (2009). The future of prediction: how Google searches foreshadow housing prices and quantities. *ICIS 2009 Proceedings*, 147.

Zumpano, L. V., Johnson, K. H., & Anderson, R. I. (2003). Internet use and real estate brokerage market intermediation. *Journal of Housing Economics*, 12(2), 134-150.

End Notes

- ¹ Measured by the abnormal search volume index (SVI) data from Google.
- ² <http://www.firstpost.com/business/economy/the-growing-pains-of-indias-rental-market-550252.html>
- ³ <http://www.tradingeconomics.com/india/interbank-rate>
- ⁴ <http://www.internetlivestats.com/internet-users-by-country/>
- ⁵ Ibid. Internet penetration is 66% in Brazil, 71% in Russia and 52% in South Africa.
- ⁶ <https://www.statista.com/statistics/239007/duration-of-monthly-internet-usage-in-selected-countries/>
- ⁷ <http://www.ibef.org/industry/real-estate-india.aspx>
- ⁸ http://nceuis.nic.in/Final_Booklet_Working_Paper_2.pdf
- ⁹ “Real estate in India – The Digital Influence,” a study by Google India and Zinnov in 2014 (cited in <http://www.financialexpress.com/archive/internet-influences-over-50-pct-real-estate-buying-decisions-google-study/1262089/>)
- ¹⁰ <http://www.thehindubusinessline.com/info-tech/venture-capitalist-pe-firms-log-on-to-realty-portals/article6462486.ece>
- ¹¹ Ibid. Google and Zinnov (2014), in its report on the role of internet in India's evolution, estimates that 53% of all internet searches have a clear purchase intention.
- ¹² The official data on housing shortage issued by the Ministry of Housing and Urban Poverty Alleviation in 2012 is for states and not cities. Anecdotal evidence and discussions with developers and property consultants suggest that these 4 cities represent the largest markets for housing in India.
- ¹³ <http://www.livemint.com/Home-Page/ZdH5Yvy72it4jHvngbZR6N/Mumbai-has-highest-number-of-Internet-users-in-country.html>
- ¹⁴ Nifty index represents the top 50 stocks in India's largest stock exchange, the National Stock Exchange (NSE).