Forecasting Residential Real Estate Price Changes from Online Search Activity

Eli Beracha
Assistant Professor
College of Business
University of Wyoming
1000 E University Ave.
Laramie, WY 82071
Email: eberacha@uwyo.edu

M. Babajide Wintoki
Assistant Professor
The School of Business
University of Kansas
1300 Sunnyside Ave.
Lawrence, KS, 66045-7585
Email: jwintoki@ku.edu

This draft: June 2012

Abstract: The intention of buying a home is revealed by many potential home buyers when they turn to the internet to search for their future residence. This paper examines the extent to which future cross sectional differences in home price changes are predicted by online search intensity in prior periods. Our findings are economically meaningful and suggest that abnormal search intensity for real estate in a particular city can help predict the city's future abnormal housing price change. On average, cities associated with abnormally high real estate search intensity consistently outperform cities with abnormally low real estate search volume by as much as 8.5% over a two-year period.

Keywords: Predictability, Google Search, Residential Real Estate, Housing Prices

1. Introduction

As individuals commonly use the web to search for information, the value of their aggregated search data as a predictive tool is increasingly recognized by academics and practitioners in different fields. Rangaswamy, Giles and Seres (2009) interpret the search trails left by individuals as "what we collectively think" and "what might happen in the future". Similarly, Batelle (2005) refers to the aggregated search data as a database of intentions.

Google, the most popular search engine in the U.S.[i], publicly provides data on search intensity of many keywords. This relatively new source of information creates a window to witness the collective thoughts or intentions of individuals in ways we were unable to do just a few years ago. Tuna (2010) notes that according to Google's chief economist, search queries such as "unemployment office" and "jobs" help predict initial jobless claims. Choi and Varian (2009) provide evidence that search behavior forecasts automobile sales and tourism and Ginsberg et al. (2009) findings suggest that influenza related search terms predict the proportion of patients visiting health professionals with related symptoms. More recently, Da, Engleberg and Gao (2011), and Joseph, Wintoki and Zhang (2011), find that the search intensity for stock tickers predicts future abnormal stock returns and trading volume.

This paper explores whether online search intensity for queries that include the words "real estate" or "rent" help predict home prices. Particularly, we examine the extent to which future cross sectional differences in home price changes are predicted by cross sectional differences in current online search intensity. We argue that search intensity for real estate terms for a particular city is a proxy for buyer sentiment for that city. Therefore, we posit that abnormal search intensity for real estate related terms for a particular city signals a future change in demand for housing for that location, which is likely to result in a future abnormal price move.

Moreover, because cities with lower land supply elasticity are less capable of absorbing shocks in housing demand, we expect that home price changes in these cities exhibit greater sensitivity to online search intensity. To our knowledge, this is the first paper that directly relates aggregated internet search data to abnormal cross sectional home prices changes.

Our findings suggest that data on real estate related inquiries indeed predict cross sectional differences in home price changes. On average, higher or lower abnormal search intensity for real estate in a particular city precedes positive and negative abnormal home price change for that city relative to the overall U.S. housing market, respectively. Our results are economically meaningful and are robust to momentum in housing prices and area-specific population growth. On average, cities associated with abnormally high real estate search volume consistently outperform cities with abnormally low real estate search volume by as much as 8.5% over a two-year period. Further tests, using panel vector auto regressions, provide strong evidence that search intensity *Granger causes* abnormal returns. We find that while search-related outperformance is independent of the size of the city, it is stronger for cities with lower land supply elasticity. Additionally, the predictive power of real estate related search terms is asymmetric in the sense that they are more prominent for "uptick" changes in home prices than to a price "downtick". This upward "stickiness" characteristic in housing prices is consistent with the mental accounting and false price reference theories. Finally, we provide evidence that following a period of outperformance, cities with intensive real estate search activity exhibit a subsequent relative performance reversal that is economically significant.

Our study extends recent work by Hohenstatt, Kasbauer and Schaffers (2011). While they find that online search activity predicts *aggregate* price changes in 20 major cities across the

U.S., we find that *cross-sectional* differences in search intensity predicts *cross-sectional* differences in price changes across more than 200 cities in the U.S.

The rest of the paper is organized as follows. The next section reviews the relevant literature. Section 3 describes the data used in the analysis and the methodology employed. Section 4 presents and discusses our finding. In Section 5, we conduct additional Granger causality tests to further asses the causal direction of the relationship between search intensity and house prices. We conclude in Section 6 and suggest future avenues of related research.

2. Literature review and hypothesis development

The theory of buyer behavior posits that consumer's search for information precedes his or her purchase decision (Beatty and Smith, 1987). However, in order for a consumer to search for information, the cost of searching must not exceed its benefit (Stigler, 1961; Baryla, Zumpano and Elder, 2000; Klein and Ford, 2003; Gwin, 2004). Over the last two decades, the rapid growth in access to computer devices and the internet significantly reduced the cost of search, which caused a surge in the amounts of digital searches and simultaneously created a comprehensive database of intentions. Choi and Varian (2009) show that trends of search for keywords can help predict retail and automotive sales as well as travel traffic. Their results are consistent with the theory presented by Beatty and Smith (1987) and suggest that consumers' intentions can be directly viewed and studied in a way that was not possible just a few years ago.

There is also evidence that trends in search queries can help predict price changes in the financial markets. Da, Engelberg and Gao (2011) show that ticker symbol search trends of U.S. stocks predict future stock prices and first-day return of IPOs. Joseph, Wintoki and Zhang (2011)

confirm Da, Engelberg and Gao's (2011) general finding and illustrate that volatile stocks, which are more difficult to arbitrage, are more sensitive to search intensity than less volatile stocks. Their results also appear to be robust to the traditional risk factors commonly used in the finance literature.[ii] The evidence provided by these two studies is particularly intriguing because U.S. financial markets are viewed by many to be very efficient and therefore any abnormal attention to a particular stock should instantaneously be incorporated in its price. The fact that security prices do not incorporate all such information, however, is consistent with the notion that attention is a scarce cognitive resource (Kahneman, 1973) and that investors only have limited attention.

The housing market is generally less efficient than traditional financial markets (Case and Shiller, 1989; Clayton, 1998), which implies that investors have some ability to predict future price changes. Among the factors that contribute to housing price forecastability are low liquidity, high transaction cost, limited pricing information, lack of professional traders and short sale constraints (Gau, 1984; Fu and Ng, 2001; Case and Shiller, 1989;Shiller, 2007, among others). As in traditional financial markets, behavioral characteristics of investors may also cause the price of housing to move in a predictable fashion. Daniel, Hirshleifer and Subrahmanyam's (1998) model relates investors' overconfidence and self-attribution biases to price momentum. These characteristics are more likely to dominate the housing market because, as argued by Joseph, Wintoki and Zhang (2011), the bulk of online search queries are likely to come from individual retail buyers – who have limited and infrequent home-buying transaction experience. Professional or institutional real estate investors often have their own proprietary databases for making investing decisions. Even when professional investors use search engines, their numbers are likely to be dwarfed by the sheer number of individual home buyers who use search engines

like Google. Finally, real estate is particularly difficult to short. Even if some investors rationally recognize that individual buyer sentiment may be pushing home values beyond fundamentals, short-sale constraints in real estate markets make it difficult for them to step in and correct prices (Harrison and Kreps, 1978; Scheinkman and Xiong, 2003).

Case and Shiller (1990) show that, in four large U.S. cities, excess change in home price tend to continue for more than a year and can be predicted by adult population trends, construction cost and real per-capita income. Beracha and Skiba (2011) confirmed this momentum effect with a sample of over 380 cities. The authors document that, going forward, cities associated with recent high price appreciation outperform cities that recently experienced low price appreciation by as much as 10% annually, on average. The inefficiencies in home prices differ among locations. Glaeser, Gyourko and Saiz (2008) show that housing price bubble is more likely to be created in cities where less land is available for development. In their analysis, the authors use Saiz (2010) land supply elasticity estimates for nearly 100 major U.S. cities, which consider land elevation and presence water bodies. Glaeser, Gyourko and Saiz (2008) base their findings on the hypothesis that markets with less land that can be developed are unable to absorb demand shocks in the form of new construction and therefore react with rapid price changes. The evidence of relative pricing inefficiencies in the housing market suggests that prices negotiated between buyers and sellers and amount of new construction created by builders do not effectively incorporate currently available information about future price projections. Therefore, it is reasonable to expect that future demand for housing in a particular market, proxied by current real estate related search volume, can help predict housing price change.

The finance literature recognizes that investors are reluctant to sell securities at a loss and therefore tend to hold losers for too long and sell winners too early (Shefrin and Statman, 1985).

This behavior is rooted in mental accounting and false reference point theories (Kahneman and Tversky, 1982; Thaler, 1985). The same kind of behavior is even more noticeable among homeowners because they are more likely to be emotionally connected to their homes and in some cases are also unable to sell due to an "underwater" situation.[iii] Genesove and Mayer (2001) show that sellers that purchased their homes at a higher price set a higher asking price and tend to have their home on the market for longer compared with sellers that purchased their homes at a lower price. Consistent with Genesove and Mayer's (2001) finding, Engelhardt (2003) examines household mobility and shows that a positive price-volume correlation in housing is in part attributed to loss aversion. This type of irrational, yet natural, behavior is especially intriguing given that most homeowners move between cities of highly co-varying home prices (Paciorek and Sinai, 2011). Seiler, Seiler, Traub, and Harrison (2008) provide evidence that behavior in line with the false reference point theory also exists among real estate investors who fixate not only on the purchase price, but also on "an all-time high" price. Seiler, Seiler and Lane (2010) also illustrate that investors' willingness to sell increases most when their investment enters the profitable territory. Moreover, evidence presented by Bokhari and Geltner (2011) suggests that loss aversion behavior is at least as pronounced among large and sophisticated investors as it is with their smaller and less sophisticated counterparts. Based on the evidence on the upward "stickiness" in home prices it is reasonable to assume that higher demand for housing will be followed by a more noticeable price "uptick" compared with the "downtick" in price that is likely to follow decline in demand for housing.

In summary, based on the review of the literature, we make four broad predictions which we set out to test in this paper. First, we predict that abnormal online search intensity for real-estate related terms for a particular city will be positively related to subsequent abnormal house

price changes in that city, at least in the short-term. Second, we predict that this positive relation will be stronger in cities with exogenous land supply constraints. Third, we predict that ability of online search intensity to predict abnormal house price changes will be stronger when there is an "uptick" in buyer sentiment compared with a "downtick" in buyer sentiment. Finally, we predict that housing price changes in "hot" cities in term of housing appreciation are more sensitive to search intensity compared to "cold" cities due to the price chasing behavior that is more likely to be present in "hot" cities.

3. Data and sample selection

The main data source used for the analysis performed in this paper is obtained from Google Insights. Google publicly provides, at a weekly or monthly frequency, the relative search volume of words or phrases that are searched from its website.[iv] For each search term, Google normalizes the weekly search data to a number between 0 and 100, where 0 represents a week with no discernible search activity and 100 is assigned to the week with the most search activity. We collect data on the search volume associated with the keywords "real estate *i*" and "rent *i*" where *i* represents each of the current 384 U.S. metropolitan statistical areas (MSA) for the period spanning January 2004 till the end of June 2011. An example of the raw data from Google Insights is shown in Figure 1. The figure illustrates the time series relative search intensity for the terms "real estate Miami" and "rent Miami" as reported by Google Insights. A glance at the panel reveals that the term "real estate Miami" was searched with much higher relative intensity during the earlier period of the data, which coincides with the recent housing bubble. On the contrary, the intensity of the search for the term "rent Miami" is higher during the latter part of

the series, which is associated with the housing bust of the "post bubble" period. It is important to note, however, that in this paper we focus on the cross sectional differences in search intensities among different MSAs and not on search intensity in time series.

Before proceeding, it is worth pausing to examine whether a search term like "real estate *i*" (where *i* is the name of a city) is the best search term to capture real estate related searches that may also capture a buying intention. To an extent, this choice is somewhat subjective. One can argue that there exists a myriad of search terms that potential home buyers may type into Google to search for real estate online. On the other hand, not everybody who types in "real estate *i*" is actually looking to buy a home. Google has no way of divining users intention, it can only track what users type in. Thus, deciding what search term to use is not a trivial problem. Indeed, this was a problem faced by Da et al (2011) and Joseph et al (2011) in the studies of the relationship between online searches for company names, and subsequent stock returns. In their studies, they get around this by using stock tickers. While searches for the term "Apple" may include users searching for a fruit or juice, searches for the stock ticker "AAPL" are more likely to be from users actually seeking out financial information about the company. Unfortunately, there are no easy to use stock tickers for real estate related searches.

Given this fundamental problem, in deciding what search terms to focus on, we are guided by two broad factors. First, the term has to be one that is broad enough to be used by people and picked up by the Google database. We experimented with extremely specific search terms. For example, we tried a term as specific as "adjustable rate mortgage Miami" and we found that the database did not pick up enough of searches for this term to report any data. This is not to say that users are not searching for adjustable rate mortgages in Miami, there just are not

enough of them specifically associating the term with the name of a city, which is something we need for our analysis.

The search term that we use also has to be associated with the intention to buy a home. This condition is a little harder to meet but the Google Insights database may give some useful clues. Along with providing data and showing a trend (as pictured in Figure 1), Google also provides the top searches associated with a particular search term. One of the terms we experimented with was "housing <*name of city*>". As an example, when we entered the term "housing Miami", the top eight searches associated the term included "Miami university housing", "Miami housing authority", "Miami apartments" and "section 8 Miami". It is unlikely that any of these search terms will be associated with intention to purchase a home. In contrast, when you enter "real estate Miami", we do not get terms that are obviously not associated with purchase intentions.[v] In this case, the term "real estate Miami" is clearly a better gauge of buyer intentions that the term "housing Miami".

Another potential criticism of using any specific search term is that it is noisy. Not everybody who searches "real estate Cleveland" plans to buy a house in Cleveland. However, the implication here is that this noise severely biases us *against* finding any statistically significant relationships. To illustrate, imagine a researcher decides to use the search term "walk way <*name of city*>". While "walk away" might have some economic justification as a negative real estate related sentiment indicator, a quick check on the Google Insights database will reveal (from the list of "top searches") that it is also a popular song. Any attempt to use this in any kind of empirical analysis is unlikely to yield powerful results since there is no inherent cross-sectional variation in searches for this term that is in any way connected with house prices.

The upshot of all this is that the use of any data from Google Insights is subject to plenty of caveats, especially the fact that it can be noisy. Nevertheless, we believe that the search terms that we have chosen to focus on: "real estate <*name of city*>" and rent <*name of city*> meet the two conditions of being general enough and not so noisy as to include mostly non real estate related searches.

Although we start with 384 MSAs, in order to make sure that the search data is indeed associated with the intended MSA, we eliminate from our sample all MSAs that carry a name duplicate. For example, Columbus is an MSA located in Indiana, Ohio and on the border between Georgia and Alabama. Similarly, there are four MSAs named Springfield located in Illinois, Massachusetts, Missouri and Ohio. Eliminating all MSAs with name duplicate reduces our sample of cities to 314. Next, we restrict our data to cities for which Google has enough data to report weekly search volume. While Google only reports data for which there is more than a trivial amount of search it defaults to reporting monthly search frequency if the frequency of search is too small to report weekly search frequency. This restriction ensures that our sample consists of cities for which there is some regularity in online search frequency for real estate related terms. Thus, our final sample in this study consist of 245 cities (MSAs) for the search term "real estate *i*" and 219 cities for the search term "rent *i*".

We obtain the housing price data from the housing price index (HPI) provided by Federal Housing Finance Agency (FHFA). The HPI is based on repeated home sales and available with quarterly frequency. Using the HPI we calculate home price changes for each MSA on a quarterly basis.[vi] Since price index data is available at a quarterly frequency, we calculate quarterly search frequency as the average weekly search frequency for the entire quarter. In our

analysis, we generally use abnormal returns, where $R_{i,t}$, is the abnormal return for city i in quarter t and is calculated as the following:

$$R_{i,t} = \left(\frac{Price_{i,t} - Price_{i,t-1}}{Price_{i,t-1}}\right) - \left(\frac{Market_{i,t} - Market_{i,t-1}}{Market_{i,t-1}}\right) \quad (1)$$

$Price_{i,t}$ is the price index for city i in quarter t and $Market_{i,t}$ is the average national price index according to the FHFA database.

Other datasets used in our analysis are MSA population estimates from the U.S. Census Bureau and land supply elasticity values from by Saiz (2010). Saiz (2010) provides land supply elasticity estimates only for 94 U.S. cities with population that exceeds 500,000. As a result, the part of our analysis that considers land supply elasticity is limited to a smaller sample of cities.

To get an idea of how house prices and search intensity have trended over time, we plot in Figure 2, the house price index and search intensity for the term "real estate <city name>" in a number of selected cities. The selected cities (which include one in each major region of the country) are Boise, ID (West), Boston, MA (Northeast), Des Moines, IA (Midwest) and Miami (South). The figures suggest that, within city, the correlation between search queries for real estate related terms and house prices is positive or negative depending on the city and the time period. Therefore, the figures provide an interesting backdrop for our cross-sectional analysis to which we turn in the next section.

4. Empirical methodology and results

4.1 Search intensity and short horizon returns

We begin our empirical analysis by investigating the ability of abnormal search intensity to predict house price changes (abnormal returns) in the following quarter. To examine the

ability of Google volume search to predict home price changes we regress the abnormal return of each MSA on the lagged abnormal search volume of the keywords "real estate" and "rent" for each of the MSAs in our sample.[vii] Our assumption is that potential home buyers' searches are associated with the keyword "real estate" and hence higher abnormal search for this keyword is a positive sentiment for housing and signals higher future demand for single family homes. On the other hand, we postulate that positive abnormal amount of searches associated with the keyword "rent" is a sign of negative sentiment for single family homes and may signal lower future housing demand.

In addition to the abnormal search variables (RE and $Rent$), we control for lagged abnormal returns. Controlling for lagged abnormal return is important because housing return is known to exhibit meaningful price momentum as first documented by Case and Shiller (1989, 1990) and more recently by Beracha and Skiba (2011). Coupled with housing price momentum is the fact that search intensity itself may be driven by past abnormal returns; it is possible that buyer sentiment may be stoked by many quarters of above average abnormal returns to housing in a particular MSA. Furthermore, because population growth may be correlated with housing price appreciation and increased search intensity, we control for the abnormal population growth in each MSA. We define abnormal population growth as the difference between population growth in the particular MSA and national population growth. We include the MSA fixed effect to control for unobservable factors (local long term economic factors, local real estate buying patterns, geography etc.) that may have been correlated with both abnormal returns and abnormal search intensity over our sample period. Finally, we include quarter dummies (quarter fixed effect). This serves two purposes. First, it accounts for any factors that are common across all cities at any particular point in time. Second, it allows us to account for seasonality.

Formally, our basic regression specification model is the following:

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} \quad (2)$$
$$+\theta Pop.Growth_{i,t} + \eta_i + d_t + \varepsilon_{i,t}$$

$R_{i,t}$ is the abnormal return on MSA i during quarter t; $RE_{i,t-p}$ and $Rent_{i,t-p}$ are the abnormal search for the keywords "real estate i" and "rent i" with p quarterly lags, respectively; $Pop.Growth_i$ is the previous year abnormal population growth for MSA i; η_i is the MSA fixed effect and d_t is a time (quarter) dummy variable. In addition to the specification described in equation (2), we also use a regression specification that includes a "net sentiment" factor generated by taking the difference between the search intensity of "real estate i" and "rent i".

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \mu_p (RE_{i,t-p} - Rent_{i,t-p}) \quad (3)$$
$$+\theta Pop.Growth_{i,t} + \eta_i + d_t + \varepsilon_{i,t}$$

While we have a relatively long time series of search and price (29 quarters), the inclusion of MSA fixed effects means that traditional fixed-effects "within" estimation may be biased since the lagged dependent variable, $RE_{i,t-p}$, is mechanically correlated with $\varepsilon_{i,s}$ for $s < t$ (Nerlove, 1967). To illustrate the bias, consider an empirical estimation model of the form:

$$y_{it} = \alpha + \rho y_{it-1} + \beta X_{it} + \eta_i + \epsilon_{it}$$

A fixed effects estimate to estimate the unobservable heterogeneity (η_i) is equivalent to time-demeaning all the variables to yield:

$$\ddot{y}_{it} = \alpha + \rho \ddot{y}_{it-1} + \beta \ddot{X}_{it} + \ddot{\epsilon}_{it}$$

Nickell (1981) shows that the bias from estimating the fixed effect specification will be given by:

$$\beta - \hat{\beta} = -(\ddot{X}_{it}' \ddot{X}_{it})^{-1} \ddot{X}_{it}' \ddot{y}_{t-1} (\rho - \hat{\rho}) + (\ddot{X}_{it}' \ddot{X}_{it})^{-1} \ddot{X}_{it} \ddot{\epsilon}_{it}$$

This expression will not be zero since

$$\rho - \hat{\rho} = -(\ddot{y}_{t-1} M \ddot{y}_{t-1})^{-1} \ddot{y}_{t-1} M \ddot{\epsilon}_{it}$$

and $E(\ddot{\epsilon}_{it} | \ddot{y}_{t-1}) \neq 0$.

To account for, and correct this potential bias, we estimate (2) and (3) using the bias-corrected fixed-effects estimator proposed by Kiviet (1995) and implemented by Bruno (2005a, 2005b).

In Table 1, we present the results of different regression specifications based on equations (2) and (3). In line with economic intuition and the extant literature, all three specifications show that population growth is positively related to abnormal change in home prices with statistical significance. Similarly, the coefficients on lagged housing price return are also positive and statistically significant, which confirms the literature evidence on housing price momentum. The new evidence the table presents is the relation between the lagged search intensity for the relevant keywords and housing price change. The first specification suggests that abnormal search intensity for the keyword "real estate" is positively related with future (next quarter) abnormal home price changes in cross section; the estimated coefficient, γ = 0.0166 is statistically significant (t = 5.60). The positive γ coefficient remains statistically significant and carry similar magnitude (γ = 0.0167 with t = 4.16) in the third specification where the negative sentiment from the search for the keyword "rent" is included. On the other hand, the coefficient for abnormal "rent" search is insignificant in the second as well as the third specification. The "net sentiment" factor is included in the fourth specification. The coefficient of the "net sentiment" is positive, statistically significant, but smaller in magnitude compared with the positive sentiment coefficient in the other specifications. This finding suggests that while the positive sentiment for home purchase proxied by search volume for the keyword "real estate" (RE) is predictive of future home prices, the negative sentiment proxied by the keyword "rent"

($Rent$) has no observable ability to predict future home price movements. It is possible that the inability of *Rent* to predict home price movement is due to two competing forces that derail the purity of information captured by this factor. While, on one hand, *Rent* serves as a negative housing sentiment because it captures the willingness to rent *instead* of buying a home, rent activity can also be positively correlated with housing demand when it is driven by an overall improving economy and job creation.[viii] This is consistent with Beracha and Johnson's (2012) finding that homeownership rate remains largely stable through time regardless of the meaningful changes in the ex-ante probability that buying is preferred to renting. Finally, the search term "rent *i*" also captures some searches that are not real estate related. While "rent *i*" searches that are not real estate related are only a small portion of all "rent *i*" searches they nevertheless contribute to the impurity of the *Rent* factor.[ix]

Specifications five through eight replicate specifications one through four, but exclude the population growth factor from the regression. From these specifications it appears that the magnitude and significance of the variables of interest (*RE*, *Rent*, and *RE-Rent*) remain practically unchanged when the population growth variable is left out.

To highlight the economic significance of these results, we note that the abnormal search variables (RE and $Rent$) are normalized from 0 to 100. This means that a 1% change in search intensity is associated with 0.0167% abnormal return. For example, a city that moves from the 25th percentile of search intensity to the 75th percentile (for the search term "real estate *i*") will, on average, in the next quarter, experience an increase in house prices that is 0.84% higher relative to the national average. This translates to 3.4% on annualized basis. While abnormal housing returns of such magnitude are, by themselves, economically meaningful it is important

to emphasize that these figures are in excess to housing appreciation that can be attributed to price momentum or population growth, which are independently significant.

One possibility is that the results in Table 1 represent a "big city" or "small city" effect. In order to examine whether the predictive ability of search is dependent on the size of the MSA, we also apply equation (2) to quadrants of the sample segmented by MSA's population at each point in time. A positive and negative coefficients for the *RE* and *Rent* variables, respectively, would suggest that Google search volume can help predict future home price changes in cross section.

In Table 2, we sort the MSAs into quartiles (Q) by population and apply the specification in equation (2) to each quartile. Q1 represent cities with the lowest population and Q4 represents the cities with the highest population. The estimated coefficients on the abnormal search variable RE are 0.0175 (t = 2.20), 0.0231 (t = 2.90), 0.0160, (t = 1.80) and 0.0232 (t = 1.96) for quartiles Q1 through Q4, respectively. While the estimated coefficients are all statistically significant at conventional levels, they are *not* significantly different from each other. The fact that the RE coefficient is stable across different population levels suggests that the ability of search volume to predict home price changes is similar across different city sizes, measured by its population.

4.2 Search intensity, home price changes and land supply elasticity

Next, we explore whether the ability of abnormal online search intensity to predict abnormal housing return is dependent on the MSA's land supply elasticity. MSAs associated with less available land are likely to be more sensitive to demand shocks because the amount of possible new construction is limited. Therefore, we hypothesize that abnormal online search

intensity will result in a higher magnitude of home price changes in areas that are associated with lower land supply elasticity. To test our hypothesis, we add two interaction terms to the regression specification defined in equations (2). These interaction terms are between the abnormal volume search for "real estate" and "rent", respectively, and the supply land elasticity associated with each MSA. More formally, the regression specification that considers the MSAs' land supply elasticity is the following:

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} +$$

$$\sum_{p=1}^{p=4} \sigma_p RE_{i,t-p} * LE_i + \sum_{p=1}^{p=4} \omega_p Rent_{i,t-p} * LE_i + \theta Pop.Growth_t + \eta_i + d_t + \varepsilon_i \qquad (4)$$

where LE_i is the supply land elasticity value from Saiz (2010) for MSA i. We interpret a negative coefficient on the interaction terms as an indication that housing price response for abnormal search volume is larger in MSAs associated with lower values of supply land elasticity (such as San-Francisco, CA or NYC, NY) compared with MSA of high supply land elasticity (such as Wichita, KS or Tulsa, OK).

The results of this analysis are shown in Table 3 and are consistent with our expectations that MSAs with less available land are more sensitive to demand shocks. We find that future change in home prices in MSAs associated with inelastic land supply are more sensitive to abnormal search volume. This is suggested by the negative and statistically significant coefficient of the interaction term between the keyword "real estate" and the value of the supply land elasticity. These results are reconfirmed when we run the regression from equation (2) on the higher and lower halves of the sample in terms of supply land elasticity. The coefficient of the "real estate" volume search is insignificant for MSAs with high supply land elasticity and positive and significant (in spite of the small sample size) for MSAs from the bottom half in terms of supply land elasticity. This indicates that MSAs with inelastic land supply are more

sensitive to demand shocks and therefore price movements in such MSAs are larger and easier to detect.

4.3 Search intensity and the direction of buyer sentiment

The finance and the real estate literatures provide evidence that investors and homeowners are more reluctant to sell an asset that has experienced nominal price decline relative to an all-time high or relative to the price for which it was acquired (Genesove and Mayer, 2001; Seiler, Seiler, Traub, and Harrison, 2008). This may suggest that home sellers are more likely to respond to an "uptick" in demand by increasing their asking and reservation price. On the other hand, homeowners would be reluctant to decrease their reservation price when the demand for housing declines. If this upward "stickiness" in home prices is observable, we would expect that positive abnormal search volume for the keyword "real estate" would be associated with a stronger abnormal positive price increase compare to the negative price decrease following a low search volume. Similarly, we would expect that the predictability of the positive sentiment proxied by volume search for the keyword "real estate" would be stronger than the predictability of the keyword "rent" that is our proxy for a negative housing sentiment. To investigate whether the predictability of volume search is asymmetric, we run two separate regressions, one testing for sensitivity to positive sentiment and the other to negative sentiment. In the regression that tests for sensitivity to positive sentiment we define dummy variables that for each MSA take the value of 1 when abnormal search volume for the keywords "real estate" and "rent" are in the top and bottom quartile, respectively, and 0 otherwise. Conversely, for the regression that examines the sensitivity to negative sentiment the dummy variables for each MSA take the value of 1 when abnormal search volume for the keywords "real estate" and "rent"

are in the bottom and top quartile, respectively, and 0 otherwise. We interact these dummy variables with the abnormal search volume variable as per the following equation:

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} + \sum_{p=1}^{p=4} \varphi_p RE_{i,t-p} * DRE_i +$$

$$\sum_{p=1}^{p=4} \psi_p Rent_{i,t-p} * DRent_i + \theta Pop.Growth_t + \eta_i + d_t + \varepsilon_i \quad (5)$$

where DRE and $DRent$ are the dummy variables for the keywords "real estate" and "rent", respectively, as described above. Higher φ and ψ coefficients under the positive sentiment specification compared with their values under the negative sentiment specification would be consistent with the assumption that home prices are upward "sticky" and that abnormal search volume has more predictability power when home prices increase.

The results from estimating equation (5) are shown in Table 4. In the positive sentiment specification, we observe that the coefficients of "real estate" search volume as well as its interaction with the positive dummy are both positive and statistically significant. This implies that abnormal search volume has more ability to predict future home price changes when search volume for "real estate" is abnormally high rather than abnormally low. These results are confirmed in the negative sentiment specification by the negative coefficient of the interaction between "real estate" search and the negative dummy. Overall, the results presented in Table 4 are consistent with previous evidence from the literature that home prices are characterized by upward "stickiness". That is, homeowners are quick to increase their asking price when housing demand picks up, but are reluctant to reduce their reservation price when housing demand declines.

4.4 Search intensity in "hot" and "cold" markets

In housing markets that exhibit rapid price appreciation ("hot" market) homeowners and investors are more likely to display a return chasing behavior compared with "cold" housing markets.[x] Therefore, home price changes in "hot" markets are expected to be more sensitive to housing demand. To test this hypothesis we divide our sample to "hot" and "cold" markets based on a pre-sample time period (2001-2004). The cities in our sample are classified as "hot" or "cold" markets based on their above or below the median returns during the 2001-2004 time period, respectively. Equation (2) is employed on "hot" and "cold" markets separately. We also employ the following equation on the full sample:

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} + \sum_{p=1}^{p=4} \rho_p RE_{i,t-p} * Dhot_i +$$

$$\sum_{p=1}^{p=4} \omega_p Rent_{i,t-p} * Dhot_i + \theta Pop.Growth_t + \eta_i + d_t + \varepsilon_i \qquad (6)$$

where *Dhot* is a dummy indicator that is equal to 1 if the city is classified as "hot" and 0 otherwise.

The results of the "hot" versus "cold" markets analysis is presented in Table 5. Consistent with our hypothesis, "hot" markets (column 1) are more sensitive than "cold" markets (column 2) to positive buyer sentiment. The coefficient of the *RE* variable for the "hot" markets is statistically significant and roughly two and half times larger than the same coefficient for the "cold" markets (0.0183 versus 0.0070). We can reach a similar conclusion from the regression specification of equation (6) on the full sample (column 3). The positive and statistically coefficient of the interaction term between *RE* and *Dhot* suggests that home prices in "hot" markets are more sensitive to changes in housing demand. Moreover, the same specification shows that the coefficient of the *Rent*Dhot* factor is negative and statistically significant.

4.5 Search intensity, longer horizon returns and home price reversals

Thus far, our analysis has focused on search intensity as a proxy for buyer sentiment and the ability of this proxy to forecast abnormal price changes over a relatively short horizon (next quarter). However, a common theme that runs through the finance and asset valuation literature (e.g. Barber, Odean and Zhu, 2009; Brown and Cliff, 2005; Joseph, Wintoki and Zhang, 2011) is that while investor sentiment (or its proxy) tends to be positively correlated with asset returns in the short term, it also tends to be negatively correlated with asset returns over a longer horizon. In other words, asset prices (in our case, housing prices) may rise abnormally in a particular city following periods of high search intensity, but this rise should be followed by a relative reversal after several quarters.

To investigate this possibility, we sort the MSAs in our sample into quartiles based on the search intensity in the prior quartile and track the mean (and median) returns of the quartile portfolios over eighteen quarters following portfolio formation. The results of this analysis are shown in Figure 3 for the highest and lowest search quartiles.

Panel A of Figure 3 shows that in the eight quarters after portfolio formation, mean (median) house prices are 5.00% (4.66%) *higher* (after adjusting for the national average change in prices) in the cities with the *highest* search intensity. In stark contrast, over the same eight quarter period, mean (median) house prices are 3.45% (2.62%) *lower* in the cities with the *lowest* search intensity.

However, over a longer horizon as shown in Panel B of Figure 3, the picture changes significantly. From quarters nine through eighteen after portfolio formation, there is a sharp reversal in returns for both the high search and low search quartiles. In the ten-quarter period after the initial rise in house prices, mean (median) house prices are 1.63% (2.63%) *lower* in the

cities with the *highest* search intensity, while mean (median) house prices are 5.97% (6.08%) *higher* in the cities with the *lowest* search intensity.

This dramatic long run reversal is illustrated in Figure 4. We sort our cities in three different ways: into (1) quartiles, (2) terciles or (3) two groups based on their search intensity for the prior quarter. We then form a portfolio that is comprised of long position in MSAs associated with the highest abnormal search volume and a short position in MSAs with abnormally low search volume. Figure 4 shows that the returns from this portfolio rises for about nine quarters after portfolio formation and then begins a gradual reversal so that by the eighteenth quarter, the return on this "long-short" portfolio is essentially indistinguishable from zero. Figure 4 also shows that the bigger the difference in search intensity between the most searched cities and the least searched cities, the bigger the short-run return differentials and long-run reversal.

5. Granger causality tests and panel vector auto regressions

Most of our analysis in the paper has relied heavily on the cross-sectional variation in our data set. Therefore, our basic result thus far can be summarized as follows: in any given quarter, cities with higher (lower) levels of search intensity will experience higher (lower) growth in residential real estate prices. Despite the fact that our inference is drawn mostly from the cross-sectional properties of our data set, it is important to note that our data has a time-series element as well. While we have tried to account for potential reverse causality as best as we can (for example, by including up to four lags of both past returns and past search intensity), there is a small but non-trivial probability that our results are driven entirely by the possibility that past returns cause higher search intensity without any of the reverse.

In this section, we conduct formal Granger causality tests to further assess the direction of causality. A variable *x* is said to *Granger cause* another variable *y* if we can reject the hypothesis that the coefficients on the lags of variable *x* in the vector auto regression (VAR) equation of variable *y* are zero (Granger, 1969).

Our tests are based on a panel VAR of the form:

$$R_{i,t} = \alpha + \sum_{p=1}^{p=k} \beta_k R_{i,t-k} + \sum_{p=1}^{p=k} \gamma_k RE_{i,t-k} + \theta Pop.Growth_{i,t} + \eta_i + d_t + \varepsilon_i \quad (7)$$

$$RE_{i,t} = \alpha + \sum_{p=1}^{p=k} \beta_k R_{i,t-k} + \sum_{p=1}^{p=k} \gamma_k RE_{i,t-k} + \theta Pop.Growth_{i,t} + \eta_i + d_t + \varepsilon_{i,t} \quad (8)$$

where the variables are as described in section 4.1 and defined in equation (2). We conduct the tests for $k = 1$ to 4 lags. The VARs are estimated using the bias corrected fixed effects estimator described in section 4.1.

The results of our Granger causality tests are presented in Table 6. Columns (1) and (2) show the *F*-statistics and p-value results of the hypotheses tests that the lags of search intensity in the return VAR of equation (7) are zero. Columns (3) and (4) show the *F*-statistics and p-value results of the hypotheses tests that the lags of return in the search intensity VAR of equation (8) are zero. The results show that search intensity strongly Granger causes abnormal returns. We are able to reject (at a one percent level of significance) the null that lags of search intensity are not significant in predicting current abnormal returns at all lag horizons up to four lags. Not surprisingly, there is evidence of bi-directional causality: search intensity predicts returns and vice versa. However, the evidence of causality from past returns to search intensity appears weaker than the reverse; indeed if we used only two lags we would not have rejected the hypothesis that past returns do not Granger cause current search intensity.

6. Conclusion

Recent advances in technology present us with the opportunity to directly observe aggregated people's intentions in ways that were impossible just a few years ago. In this paper we use Google search intensity for housing related queries as a proxy for housing sentiment. Using our proxy for housing sentiment we explore the extent to which it can help predict future home price changes in cross section. Our general results reveal that abnormal change in housing prices for a particular city follows abnormal housing related search volume for that city. The magnitudes of the results are economically meaningful, most noticeable for cities with inelastic supply of available land that can be developed and robust to housing momentum as well as population growth.

To our knowledge, this is the first paper that directly relates aggregated internet search data to abnormal cross sectional home prices changes and we hope that more research on this topic will follow. Future research should examine how home builders, future traders and mortgage originators may use search data to better position themselves for housing price moves. Additionally, it will be interesting to see whether the ability of search intensity to predict future price movements dissipates as this topic gains the attention of practitioners.

Acknowledgements

We would like to thank Ko Wang (editor), three anonymous referees, Kissan Joseph, and participants at the 2012 American Real Estate Society (ARES) Annual Meeting, for many useful comments. We would also like to thank Dev vrat Khanna and Hatem Shoshaa for excellent research assistance.

References:

Barber, B. M., T. Odean and N. Zhu., Do Retail Trades Move Markets? *Review of Financial Studies*, 2009, 22, 151 – 186.

Baryla, E. A., L. V. Zumpano and H. W. Elder, An Investigation of Buyer Search in the Residential Real Estate Market Under Different Market Conditions. *Journal of Real Estate Research*, 2000, 20 (1/2), 75 – 91.

Batelle, J., 2005. Search. Penguin Group, New York, NY.

Beatty, S. E. and S. M. Smith, External Search Effort: An Investigation Across Several Product Categories. *Journal of Consumer Research*, 1987, 14, 83 – 95.

Beracha, E. and K. H. Johnson, Lessons from over 30 Years of Buy versus Rent Decisions: Is the American Dream Always Wise? *Real Estate Economics,* 2012, 40(2).

Beracha, E. and H. Skiba, Momentum in Residential Real Estate. *Journal of Real Estate Finance and Economics,* 2011, 43(3), 299 – 320.

Bokhari, S. and D. Geltner, Loss Aversion and Anchoring in Commercial Real Estate Pricing: Empirical Evidence and Price Index Implications. *Real Estate Economics*, 2011, 39(4), 635 – 670.

Brown, G. W. and M. T. Cliff, Investor Sentiment and Asset Valuation. *Journal of Business*, 2005, 78, 405 – 440.

Bruno, G.S.F, Approximating the Bias of the LSDV Estimator for Dynamic Unbalanced Panel Data Models. *Economics Letters*, 2005, 87, 361 – 366.

Bruno, G.S.F, Estimation and Inference in Dynamic Unbalanced Panel Data Models with a Small Number of Individuals. *The Stata Journal*, 2005, 5(4), 473 – 500.

Calhoun, C. A., OFHEO House Price Indices: HPI Technical Description. Office of Federal Housing Enterprise Oversight, 1996.

Case, K. E. and R. J. Shiller, The Efficiency of the Market for Single-Family Homes. *American Economic Review*, 1989, 79(1), 125 – 137.

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

Choi, H. and Varian, H., Predicting the Present with Google Trends, working Paper, 2009.

Clayton, J., Further Evidence on Real Estate Market Efficiency. *Journal of Real Estate Research*, 1998, 15 (1/2), 41 – 57.

Da, Z., J. Engelberg and P. Gao, In Search of Attention. *Journal of Finance*, 2011 Forthcoming.

Daniel, K., D. Hirshleifer and A. Subrahmanyam, Investor Psychology and Security Market Under and Overreactions. *Journal of Finance*, 1998, 53, 1839 – 1835.

Engelhardt, G., Nominal Loss Aversion, Housing Equity Constraints, and Household Mobility: Evidence from the United States, *Journal of Urban Economics*, 2003, 53, 171-195.

Fu, Y. and K. L. Ng, Market Efficiency and Return Statistics: Evidence from Real Estate and Stock Markets Using a Present Value Approach. *Real Estate Economics*, 2001, 29(2), 227–250.

Gau, G. W., Weak Form Tests of the Efficiency of Real Estate Investment Markets. *Financial Review*, 1984, 19(4), 301–320.

Genesove, D. and C. Mayer, Loss Aversion and Seller Behavior: Evidence from the Housing Market, *Quarterly Journal of Economics*, 2001, 116, 1233 – 1260.

Ginsberg, J., M. H. Mohebb, R. S. Patel, L. Brammer, M. S. Smolinsky and L. Brilliant, Detecting Influence Epidemics Using Search Engine Query Data. *Nature*, 2009, 457, 1012 – 1014.

Glaeser, E., J. Gyourko and A. Saiz, Housing Supply and Housing Bubbles, *Journal of Urban Economics*, 2008, 64, 198 – 217.

Gwin, C.R., International Comparisons of Real Estate E-nformation on the Internet. *Journal of Real Estate Research*, 2004, 26 (1), 1 – 23.

Harrison, J.M. and D. M. Kreps, Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations, *Quarterly Journal of Economics*, 1978, 92, 323 – 336.

Hohenstatt, R., M. Kasbauer and W. Schaffers, "Geco" and its Potential for Real Estate Research: Evidence from the U.S. Housing Market. *Journal of Real Estate Research*, 2011, 33 (4), 471 – 506.

Joseph, K., M. B. Wintoki and Z. Zhang, Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search. *International Journal of Forecasting*, 2011, 27, 1116 – 1127.

Kahneman, D., Attention and Effort (Prentice-Hall), Englewood Cliffs, NJ, 1973.

Kahneman, D. and A. Tversky, The Psychology of Preferences. *Scientific American*, 1982, 246, 167 – 173.

Kiviet J.F., On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics*, 1995, 68, 53 – 78.

Klein, L. R. and G. T. Ford, Consumer search for information in the digital age: an empirical study of purchase search for automobiles. *Journal of Interactive Marketing*, 2003, 17, 29 – 49.

Nerlove, M., Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time-Series of Cross-Sections. *Economic Studies Quarterly*, 1967, 18, 42 – 74.

Nickell, S., Biases in Dynamic Models with Fixed Effects. *Econometrica*, 1981, 49, 1417 – 1426.

Paciorek, A. and T. Sinai, Does Home Owning Smooth the Variability of Future Housing Consumption? *Journal of Urban Economics,* 2011, 71, 244 – 257.

Rangaswamy, A., C. L. Giles and S. Seres, A Strategic Perspective on Search Engines: Thought Candies for Practitioners and Researchers. *Journal of Interactive Marketing,* 2009, 23, 49 – 60.

Saiz, A., The Geographic Determinants of Housing Supply. *The Quarterly Journal of Economics*, 2010, 125, 1253 – 1296.

Scheinkman, J.A. and W. Xiong, Overconfidence and Speculative Bubbles. *Journal of Political Economy*, 2003, 111, 1183 – 1219.

Seiler, M., V. Seiler and M. Lane, Mental Accounting and False Reference Points in Real Estate Investment Decision-Making, *Journal of Behavioral Finance*, 2012, 13(1), forthcoming.

Seiler, M., V. Seiler, S. Traub and D. Harrison, D., Regret Aversion and False Reference Points in Residential Real Estate, *Journal of Real Estate Research*, 2008, 30(4), 461 – 474.

Shefrin, H. and M. Statman, The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *Journal of Finance*, 1985, 40(3), 777 – 790.

Stigler, G. J., The Economics of Information. *Journal of Political Economy*, 1961, 29, 213 – 225.

Thaler, R., Mental Accounting in and Consumer Choice, *Marketing Science*, 1985, 4(3), 199 – 214.

Tuna, C., New Ways to Read Economy. Wall Street Journal, April 8, 2010.

Wang, K., J. Erickson and S. Chan, Does the REIT Stock Market Resemble the General Stock Market? *Journal of Real Estate Research*, 1995, 10 (4), 445 – 460.

Figure 1: Google Insights screen shot of search terms "real estate Miami" and "rent Miami"

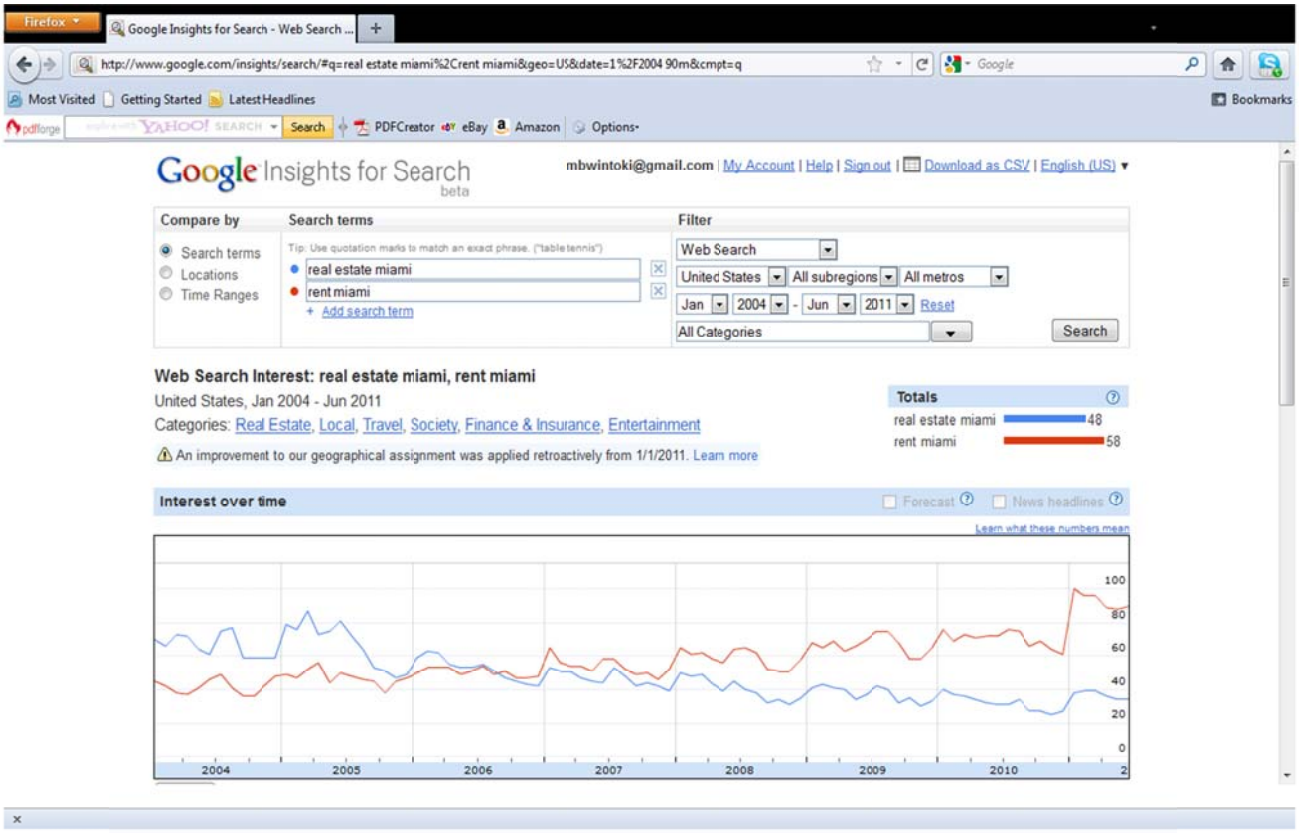


Figure 2A: Home Prices and Search Intensity for Boise, ID

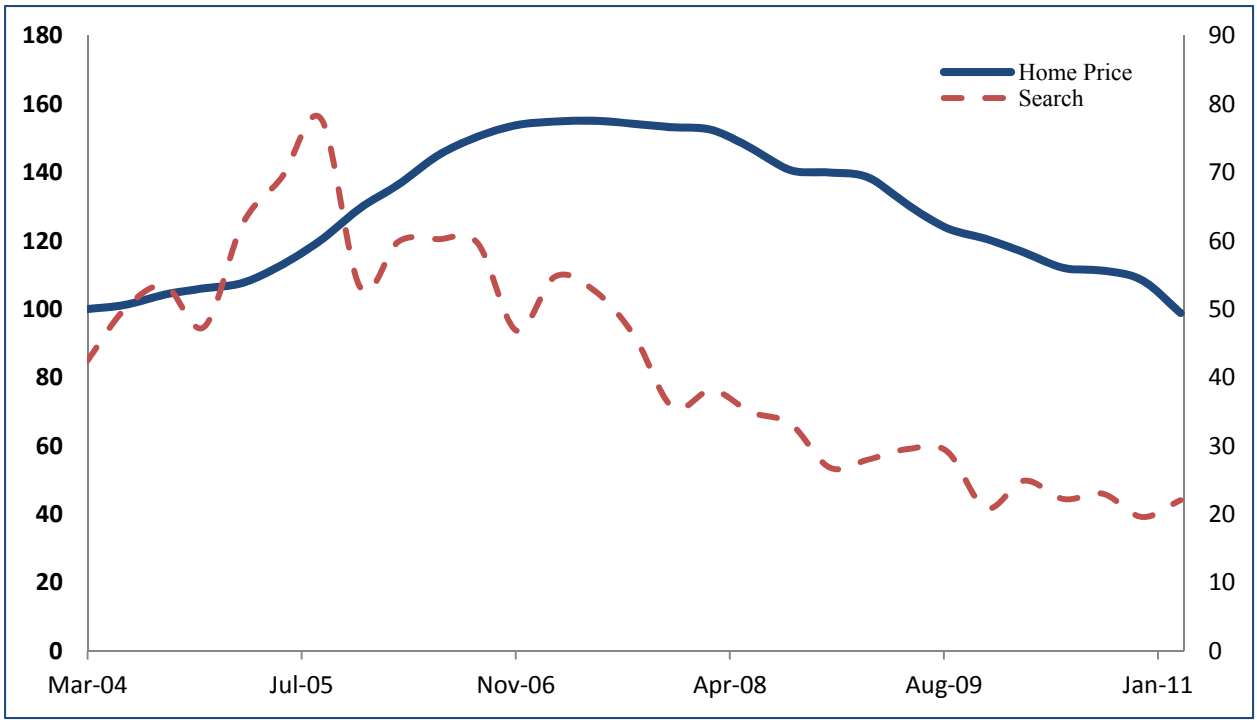
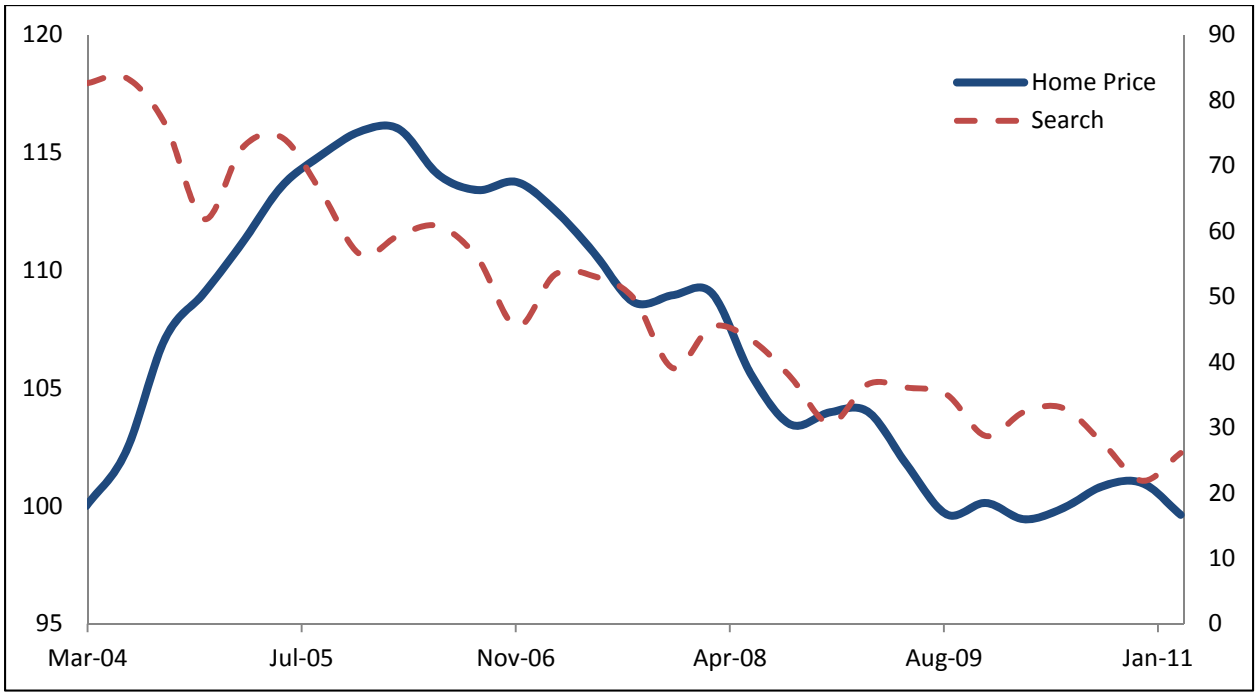


Figure 2B: Home Prices and Search Intensity for Boston, MA



Home Price (2004:1 = 100) on left axis; Search intensity on right axis

Figure 2C: Home Prices and Search Intensity for Des Moines, IA

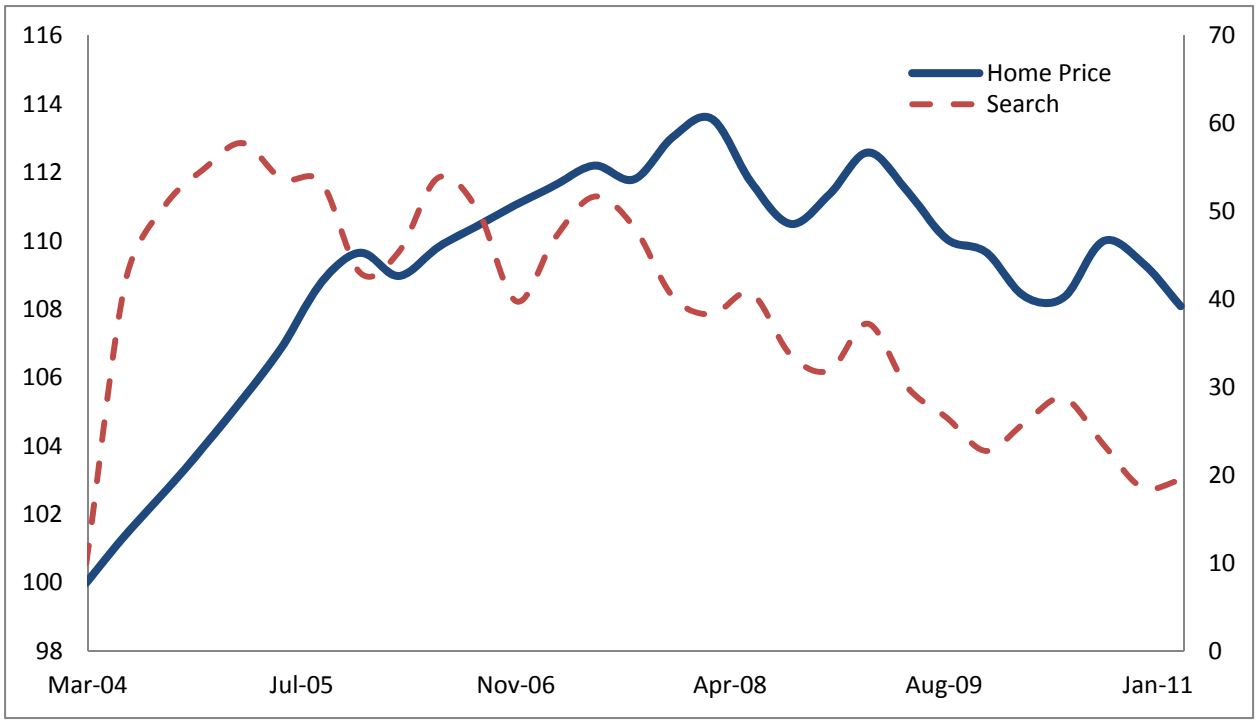
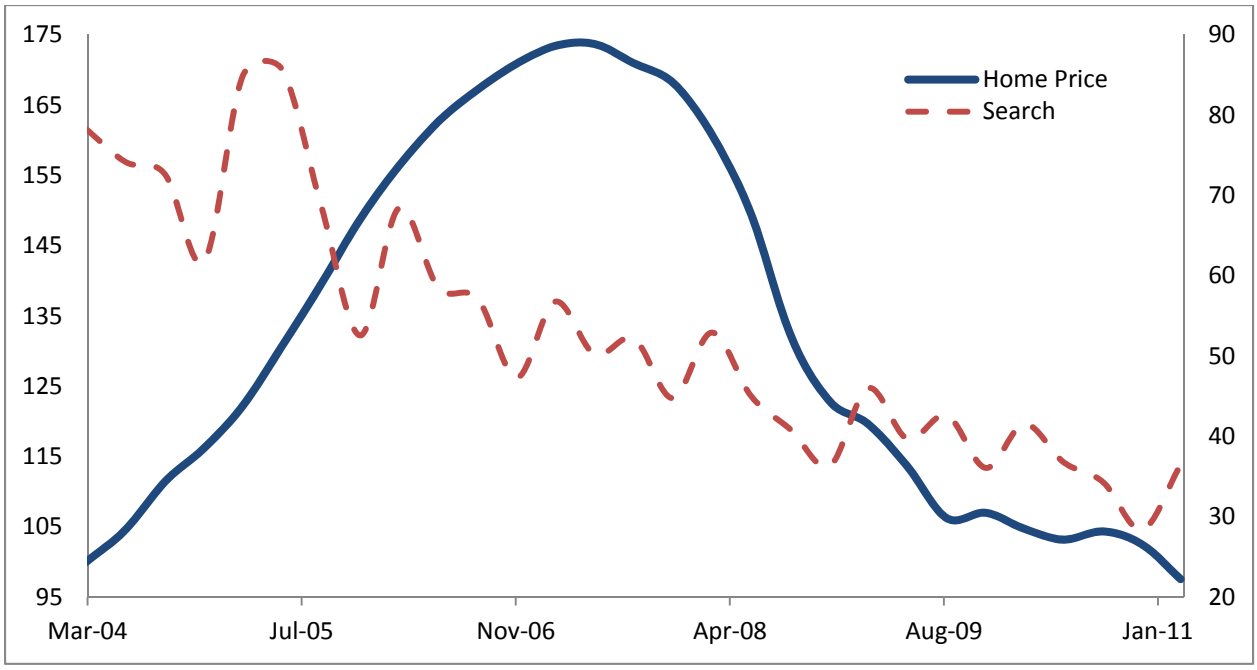


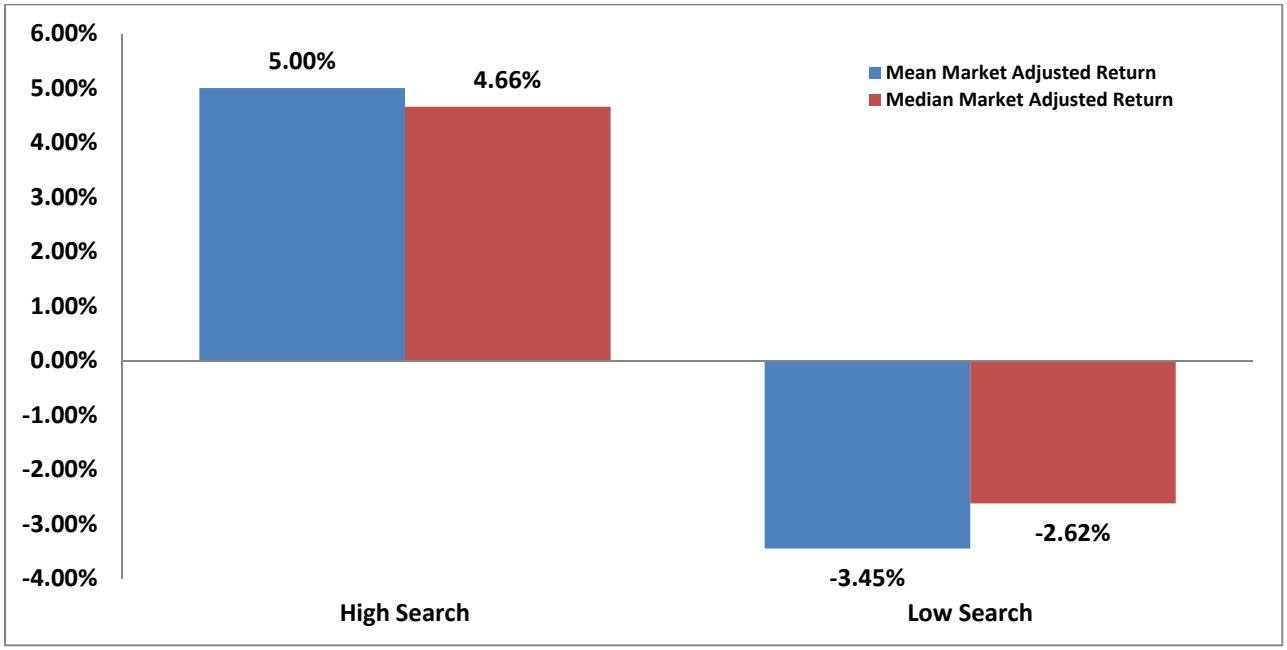
Figure 2D: Home Prices and Search Intensity for Miami, FL



Home Price (2004:1 = 100) on left axis; Search intensity on right axis

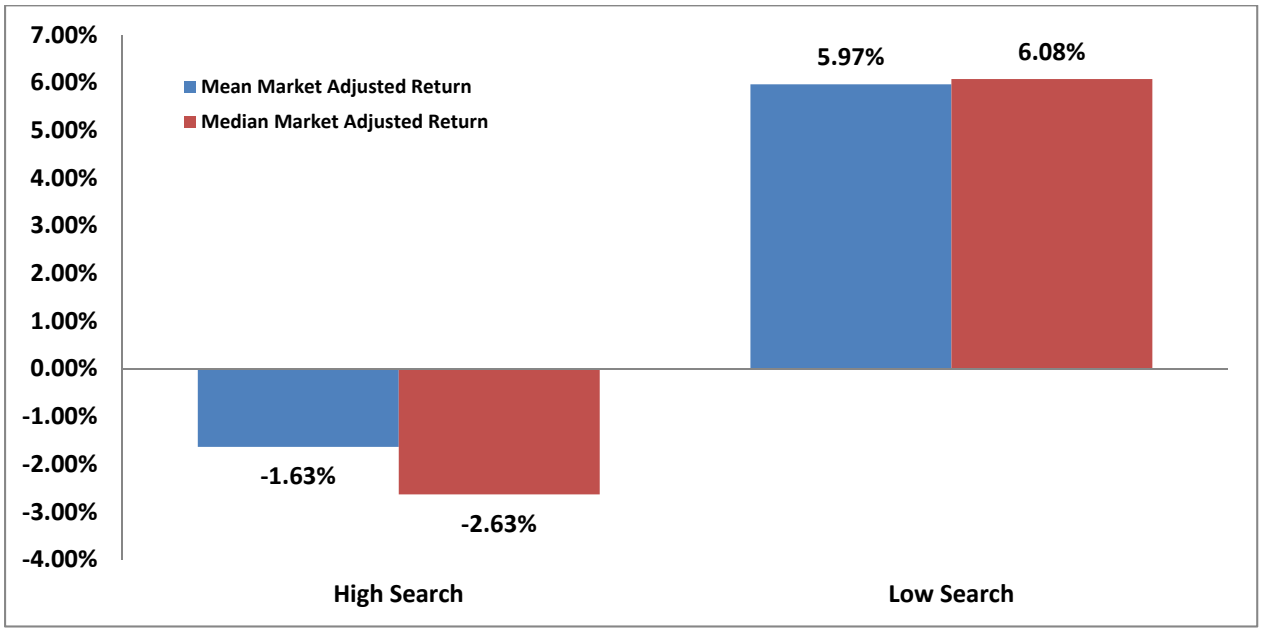
Figure 3: Long run housing returns to the most and least searched cities.

Panel A: Mean (Median) Returns from Quarter 1 to Quarter 8 following portfolio formation



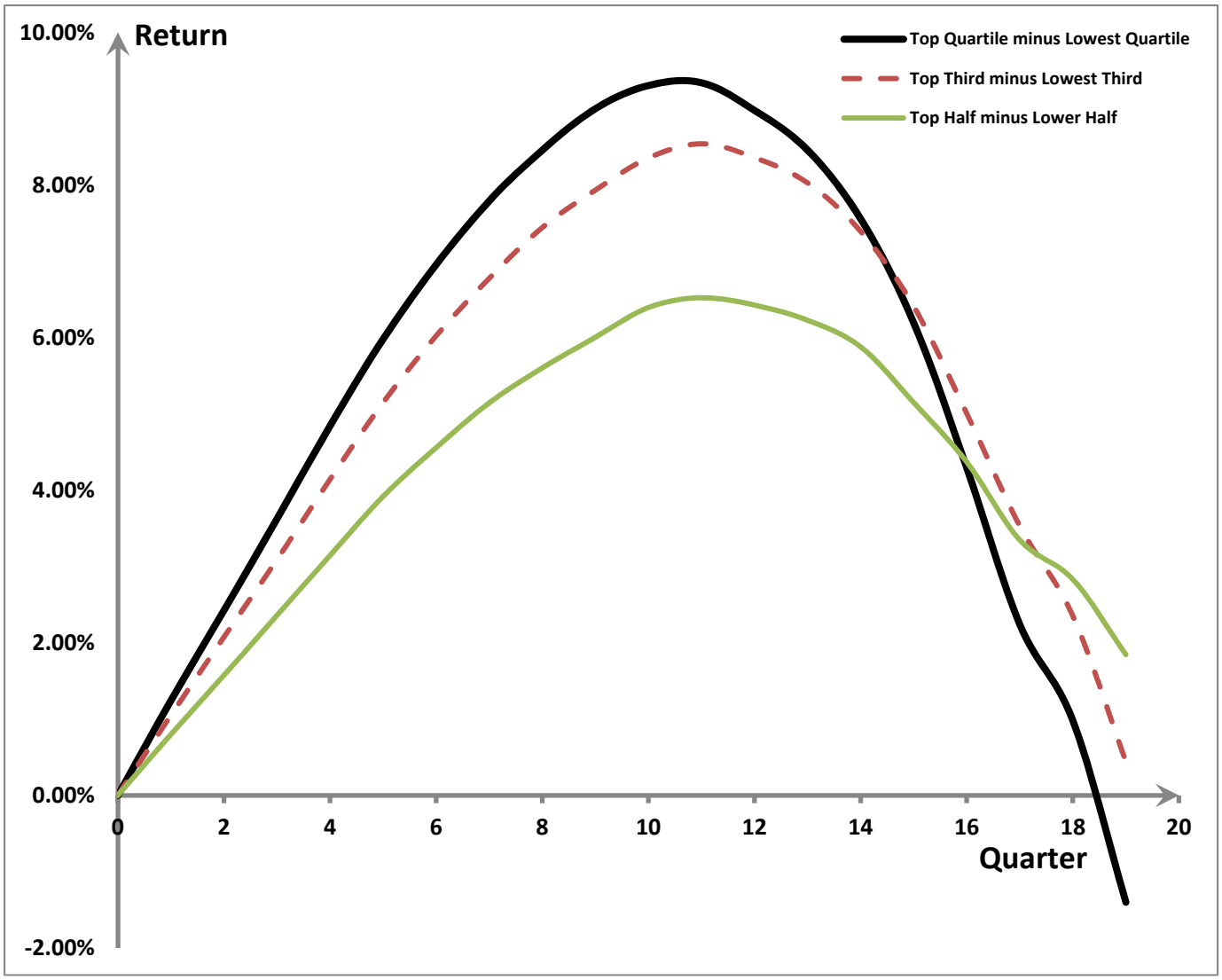
MSAs in our sample are sorted into quartiles based on the search intensity in the prior quarter. The mean and median market-adjusted returns of the quartile portfolios over the eight quarters following portfolio formation are recorded.

Panel B: Mean (Median) Returns from Quarter 9 to Quarter 18 following portfolio formation



MSAs in our sample are sorted into quartiles based on the search intensity in the prior quarter. The mean and median market-adjusted returns of the quartile portfolios over the quarters 9-18 following portfolio formation are recorded.

Figure 4: Cumulative returns for a "long-short" search intensity sorted portfolio.



The sample of MSAs is sorted in three different ways: into (1) quartiles; (2) terciles or (3) two groups based on their search intensity for the prior quarter. We then form a portfolio that is comprised of long position in MSAs associated with the highest abnormal search volume and a short position in MSAs with abnormally low search volume.

Table 1: Abnormal search volume and home price changes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$R_{i,t-1}$	0.2605***	0.3084***	0.2903***	0.2962***	0.2646***	0.2644***	0.2450***	0.2520***
	(19.92)	(21.40)	(20.35)	(20.60)	(19.81)	(20.75)	(19.17)	(19.75)
$RE_{i,t-1}$	0.0166***		0.0167***		0.0167***		0.0163***	
	(5.60)		(4.16)		(4.79)		(3.84)	
$Rent_{i,t-1}$		0.0034	-0.0016			0.0039	-0.0012	
		(1.06)	(-0.48)			(1.42)	(-0.45)	
$RE_{i,t-1} - Rent_{i,t-1}$				0.0057**				0.0048**
				(2.00)				(2.18)
$Pop.Growth$	0.0814***	0.0810**	0.0763**	0.0720**				
	(2.85)	(2.53)	(2.36)	(2.00)				
Number of cities	245	219	219	219	245	219	219	219
Total number of observations	5,145	4,581	4,581	4,581	5,145	4,581	4,581	4,581

The results reported in Table 1 span the 2004:1 to 2011:1 time period and generated as per the equations below with 4 lags. For brevity, only the magnitude and the statistical significance of the first lag are reported. The coefficient of the variable of interest, $RE_{i,t-p}$, is only significant for the first lag. The absolute values of the t-statistics are reported under the coefficients (*significant at 10% level, **significant at 5% level, ***significant at 1% level). As a robustness test we also examine the above eight regression specifications against a population weighted housing index that is based only on the cities included in our sample. These results are available upon request and practically indistinguishable from the results presented above.

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} + \theta Pop.Growth_i + \eta_i + d_t + \varepsilon_i \quad (2)$$

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \mu_p (RE_{i,t-p} - Rent_{i,t-p}) + \theta Pop.Growth_{i,t} + \eta_i + d_t + \varepsilon_{i,t} \quad (3)$$

Table 2: Abnormal search volume and home price changes – Population quadrants

	(Q1)	(Q2)	(Q3)	(Q4)
$R_{i,t-1}$	-0.0433	0.2144***	0.4548***	0.6395***
	(-1.32)	(6.32)	(14.91)	(23.46)
$RE_{i,t-1}$	0.0175**	0.0231***	0.0160*	0.0232**
	(2.20)	(2.90)	(1.80)	(1.96)
$Rent_{i,t-1}$	0.0076	0.0005	-0.0056	-0.0100
	(1.20)	(0.10)	(-0.87)	(-1.30)
$Pop.Growth$	0.0400	0.3180***	0.1496	0.0477*
	(0.42)	(2.69)	(1.61)	(1.72)
Number of cities	44	58	63	63
Total number of observations	912	1,158	1,230	1,281

The results reported in Table 2 span the 2004:1 to 2011:1 time period and generated as per the equation below with 4 lags. For brevity, only the magnitude and the statistical significance of the first lag are reported. The coefficient of the main variable of interest, $RE_{i,t-p}$, is only significant for the first lag. Q1 through Q4 are quartiles of our MSA sample in terms of population, where Q1 and Q4 represent the quartile of the MSAs with the lowest and highest population, respectively. MSAs are classified once a year into quartiles, so that the same MSA may appear in more than one quartile throughout the analysis. The absolute values of the t-statistics are reported under the coefficients (*significant at 10% level, **significant at 5% level, ***significant at 1% level).

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} + \theta Pop.Growth_i + \eta_i + d_t + \varepsilon_i \quad (2)$$

Table 3: Abnormal search volume and home price changes – Land supply elasticity

	(1)	(2)	(3)
	Full sample	High Land Supply Elasticity	Low Land Supply Elasticity
$R_{i,t-1}$	0.6258***	0.5088***	0.6559***
	(26.12)	(14.13)	(17.57)
$RE_{i,t-1}$	0.0396***	0.0024	0.0225*
	(2.76)	(0.21)	(1.71)
$Rent_{i,t-1}$	-0.0079	-0.0198	-0.0060
	(-0.67)	(-0.68)	(-0.63)
$RE_{i,t-1} * LE_i$	-0.0136*		
	(-1.85)		
$Rent_{i,t-1} * LE_i$	-0.0031		
	(-0.48)		
$Pop.Growth$	0.0220	0.0195	0.0151
	(0.69)	(0.14)	(0.44)
Number of cities	72	37	35
Observations	1,512	777	735

The results reported in Table 3 span the 2004:1 to 2011:1 time period and generated as per the equation below with 4 lags. For brevity, only the magnitude and the statistical significance of the first lag are reported. The coefficient of the main variables of interest: $RE_{i,t-p}$ and $RE_{i,t-p} * LE_i$ are only significant for the first lag. The absolute values of the t-statistics are reported under the coefficients (*significant at 10% level, **significant at 5% level, ***significant at 1% level).

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} +$$

$$\sum_{p=1}^{p=4} \sigma_p RE_{i,t-p} * LE_i + \sum_{p=1}^{p=4} \omega_p Rent_{i,t-p} * LE_i + \theta Pop.Growth_t + \eta_i + d_t + \varepsilon_i \qquad (4)$$

Table 4: Abnormal search volume and home price changes – Positive vs. negative sentiment

	(1)	(2)
	Positive sentiment	Negative sentiment
$R_{i,t-1}$	0.2890***	0.2808***
	(20.15)	(19.20)
$RE_{i,t-1}$	0.0129**	0.0277***
	(2.55)	(6.01)
$Rent_{i,t-1}$	-0.0017	0.0008
	(-0.48)	(0.22)
$RE_{i,t-1} * DRE_i$ (positive)	0.0101*	
	(1.78)	
$Rent_{i,t-1} * DRent_i$ (positive)	0.0001	
	(0.02)	
$RE_{i,t-1} * DRE_i$ (negative)		-0.0311***
		(-3.95)
$Rent_{i,t-1} * DRent_i$ (negative)		-0.0045
		(-0.61)
$Pop.Growth$	0.0724***	0.0697***
	(2.92)	(2.74)
Number of cities	219	219
Total number of observations	4,581	4,581

The results reported in Table 4 span the 2004:1 to 2011:1 time period and generated as per the equation below with 4 lags. For brevity, only the magnitude and the statistical significance of the first lag are reported. The coefficient of the variables of interest: $RE_{i,t-p}$ and $RE_{i,t-p} * DRE_i$ are only significant for the first lag. Under the positive sentiment specification *DRE* and *DRent* are set to 1 when the abnormal volume search for the terms "real estate" and "rent" are in the top and bottom quartile of MSA i, respectively, and to 0 otherwise. Conversely, under the negative sentiment specification *DRE* and *DRent* are set to 1 when the abnormal volume search for the terms "real estate" and "rent" are in the bottom and top quartile of MSA i, respectively, and to 0 otherwise. The absolute values of the t-statistics are reported under the coefficients (*significant at 10% level, **significant at 5% level, ***significant at 1% level).

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} + \sum_{p=1}^{p=4} \varphi_p RE_{i,t-p} * DRE_i +$$

$$\sum_{p=1}^{p=4} \psi_p Rent_{i,t-p} * DRent_i + \theta Pop.Growth_t + \eta_i + d_t + \varepsilon_i \quad (5)$$

Table 5: Abnormal search volume and home price changes in "hot" and "cold" markets

	(1)	(2)	(3)
	"Hot" markets	"Cold" markets	Full sample
$R_{i,t-1}$	0. 3672***	0. 0152	0. 2862***
	(16.78)	(0.72)	(19.91)
$RE_{i,t-1}$	0. 0183***	0.0070*	0.0111**
	(2.96)	(1.93)	(2.29)
$Rent_{i,t-1}$	-0. 0024	0.0032	0.0060
	(-0.56)	(1.08)	(1.35)
$RE_{i,t-1} * Dhot_i$			0. 0132**
			(2.03)
$Rent_{i,t-1} * Dhot_i$			-0. 0144***
			(-2.71)
$Pop.Growth$	0. 0569**	0.0350	0. 0759***
	(2.11)	(0.68)	(3.06)
Number of cities	112	107	219
Observations	2,334	2,247	4,581

The results reported in Table 5 span the 2004:1 to 2011:1 time period and generated as per the equation below with 4 lags. For brevity, only the magnitude and the statistical significance of the first lag are reported. The coefficient of the variables of interest: $RE_{i,t-p}$ and $RE_{i,t-p} * Dhot_i$ are only significant for the first lag. *Dhot* is set to 1 if housing performance in MSA i was above the sample median during the 2001-2004 time period and to 0 otherwise. The absolute values of the t-statistics are reported under the coefficients (*significant at 10% level, **significant at 5% level, ***significant at 1% level).

$$R_{i,t} = \alpha + \sum_{p=1}^{p=4} \beta_p R_{i,t-p} + \sum_{p=1}^{p=4} \gamma_p RE_{i,t-p} + \sum_{p=1}^{p=4} \delta_p Rent_{i,t-p} + \sum_{p=1}^{p=4} \rho_p RE_{i,t-p} * Dhot_i +$$

$$\sum_{p=1}^{p=4} \omega_p Rent_{i,t-p} * Dhot_i + \theta Pop.Growth_t + \eta_i + d_t + \varepsilon_i \qquad (6)$$

Table 6: Granger causality test results

	H_0: Search intensity does not cause abnormal returns		H_0:Abnormal returns do not cause search intensity	
	(1)	(2)	(3)	(4)
k	*F*-statistic	(*p*-value)	*F*-statistic	(*p*-value)
1	100.59	(0.00)***	4.22	(0.04)**
2	28.00	(0.00)***	0.78	(0.46)
3	17.91	(0.00)***	2.47	(0.06)*
4	12.23	(0.00)***	4.58	(0.01)***

The results reported in Table 6 span the 2004:1 to 2011:1 time period and shows the results of Granger causality tests applied to the panel vector auto regression (VAR) of equations (7) and (8). *k* denotes the lags used in the test. F-statistics and the corresponding p-values and levels of significance (*significant at 10% level, **significant at 5% level, ***significant at 1% level) are reported in the table. Columns (1) and (2) show the *F*-statistics and p-value results of the hypotheses tests that the lags of search intensity in the return VAR of equation (7) are zero. Columns (3) and (4) show the *F*-statistics and p-value results of the hypotheses tests that the lags of return in the search intensity VAR of equation (8) are zero.

$$R_{i,t} = \alpha + \sum_{p=1}^{p=k} \beta_k R_{i,t-k} + \sum_{p=1}^{p=k} \gamma_k RE_{i,t-k} + \theta Pop.Growth_{i,t} + \eta_i + d_t + \varepsilon_i \quad (7)$$

$$RE_{i,t} = \alpha + \sum_{p=1}^{p=k} \beta_k R_{i,t-k} + \sum_{p=1}^{p=k} \gamma_k RE_{i,t-k} + \theta Pop.Growth_{i,t} + \eta_i + d_t + \varepsilon_{i,t} \quad (8)$$

End noted

[i] According to Hitwise (http://www.hitwise.com/us/press-center/press-releases/experian-hitwise-reports-bing-powered-share-of-s/) Google accounts for 64.42% of U.S. searches as of March 2011.

[ii] The four risk factors commonly used in the finance literature are market return, firm size, firm book value, and return momentum.

[iii] A situation under which, the value of the home is lower than its outstanding mortgage balance.

[iv] The data is available at http:www.google.com/insights/search/.

[v] The top nine searches associated with the term "real estate Miami: are "miami florida", "florida real estate", "Miami beach", "Miami homes", "miami realty", "south beach Miami", "Miami condos", "miami condo", "Miami realtor".

[vi] The repeated sales housing price index is available on: http://www.fhfa.gov/. FHFA defines a repeated sale when the same physical address originates at least two mortgages and those mortgages are purchased by either Freddie Mac or Fannie Mae. The use of repeated sales of the same physical address controls for properties' characteristics, and reduces the effect of changes in construction quality over time on changes in housing prices. For more detail about the index construction, see Calhoun (1996).

[vii] In this version of the paper we include four quarterly lags. However, other specifications that are excluded from this paper illustrate that the coefficients of our variables of interest ($RE_{i,t-1}$ and $Rent_{i,t-1}$) are robust to different number of *R*, *RE* and *Rent* lags.

[viii] We thank an anonymous referee for bringing the inherent noisiness of our *Rent* factor to our attention.

[ix] We examine the top 10 search terms associated with "rent *i*" for each city in our sample. Search terms that are not real estate related (car rental, boat rental, etc…) account for less than 5% of all "rent *i*" related searches. Obtaining data on more specific search terms such as "housing for rent *i*" or "real estate for rent *i*" was unsuccessful because data on these search terms was unavailable for many of the cities in our sample.

[x] This behavior is not limited to retail investors or residential home prices; Wang, Erickson and Chan (1995) find that the level of financial analysts' coverage and stock turnover intensity are higher when the REIT stock market is "hot".