Search and Predictability of Prices in the Housing Market*

Stig V. Møller† Thomas Q. Pedersen‡ Christian M. Schütte§
Allan Timmermann¶

May 2022

Abstract

We develop a new housing search index (HSI) extracted from online search activity on a limited set of keywords related to the house buying process. We show that HSI has strong predictive power over subsequent changes in house prices, both in-sample and out-of-sample and after controlling for the effect of commonly used predictors, and relate our findings to models of search-induced frictions. Our results imply that search data can be used as an early indicator of where the market is going.

Keywords: Internet search, housing markets, housing demand, forecasting, frictions, inelastic housing supply.
JEL codes: C10, E17, G10, R3.

*We thank two anonymous reviewers, an Associate Editor, and the Department Editor, David Sraer, for insightful and constructive comments. In addition, we thank Anil Kumar, participants at the 2019 SoFiE Conference in Shanghai, the IAAE 2019 Annual Conference in Nicosia, the 2019 Nordic Econometric Meeting in Stockholm, the 2019 International Symposium on Forecasting in Thessaloniki, the 2019 CFE in London and seminar participants at Aarhus University for comments. We are grateful to Daryl Fairweather, Redfin Chief Economist, for kindly providing us with data on Redfin's Housing Demand Index. We acknowledge support from the Danish Finance Institute and from the Independent Research Fund Denmark (DFF 7024-00020B, DFF 7015-00017B).

†CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus and the Danish Finance Institute. E-mail: svm@econ.au.dk.

‡CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus and the Danish Finance Institute. E-mail: tqpedersen@econ.au.dk.

§CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus and the Danish Finance Institute. E-mail: christianms@econ.au.dk.

¶UCSD, CEPR, and CREATES, University of California, San Diego, La Jolla, CA 92093. E-mail: atimmermann@ucsd.edu.

1 Introduction

Fluctuations in house prices have a profound impact on household welfare, financial stability, and the broader economy. For example, Case et al. (2012) estimate that the decline in U.S. housing wealth during 2005-2009 implied a decline in consumption of about $350 billion per year. Further, in an analysis covering more than 60 countries, Reinhardt and Rogoff (2009) show that house price bubbles have historically been among the best predictors of banking crises across both advanced and emerging market economies. In response to the importance of variation in house prices for macroeconomic stability, the European Commission recently included house prices in its early warning system for macroeconomic imbalances (the "MIP Scoreboard"). Reliable and accurate predictions of house prices are evidently of great importance for policy makers as they can be used to predict future costs of living and revenues from real estate taxes or provide an early warning of an incipient price boom–or, potentially a weakening–in the housing market. Accurate forecasts are also valuable to households planning to buy or sell in the residential real estate market, particularly if available at the local market level.

The housing market is characterized by a highly heterogeneous and complex product, local segmentation, and a slow price discovery process caused by a variety of frictions. Buying a house is, therefore, a search intensive process involving a lengthy review of homes for sale and price comparisons across the inventory of homes listed for sale at a given point in time. Much of this search process is conducted online. A recent report by the National Association of Realtors (NAR, 2020) shows that home buyers use the internet as their main source of information about the housing market, with as many as 93% of home buyers using the internet to search for a home.

This paper develops and tests a set of hypotheses about the relation between online housing search volume and changes in house prices. Our first and main hypothesis is that search activity, which tracks peoples' intentions of buying a house and thereby proxies for housing demand, should have a positive relation with house prices. Given various frictions in the housing market, an increase in search activity is propagated into future periods, implying sluggish price adjustment in response to an increase in demand such that search activity should hold predictive power for future variation in house prices – an insight that follows directly from theoretical search-based models (e.g. Berkovec and Goodman, 1996, Genesove and Han, 2012, and Carrillo et al., 2015). Because the house search process tends to be lengthy, our second hypothesis is that internet search volume has predictive power at both short and long horizons, but also that its predictive power declines at longer horizons where frictions are less likely to be binding. Our third hypothesis is that the predictive power of housing search, being a proxy for housing demand, is particularly strong in housing markets with low supply elasticity. Since housing markets are inherently local and segmented, our fourth hypothesis is that local search activity contains important information about local house prices beyond what is captured in national search activity.

The intense and lengthy search process involved in buying a house coupled with large frictions in the housing market means that it is natural to expect internet search volume for housing to have predictive power for future house prices. Using Google Trends search data, we start out with the keyword "buying a house" and add related search terms supplied by Google, all of which are related to the search process of future home buyers. To capture common variation across search volume indices, we define the Housing Search Index (HSI) as the first principal component of the search volume indices. This provides us with a simple and intuitive measure of housing demand. We validate our search-based measure of demand by comparing it to data on home tours and writing offers.

We show that demand for housing as measured through online search activity predicts future house prices at both short and long horizons. At the one-month horizon, HSI explains more than 50% of the variation in national house price growth, while at the one-year horizon the explanatory power is close to 65%. The predictive power of HSI peaks at horizons around 3-8 months, which is consistent with the time buyers typically spend finding a home from the initial search process to closing the deal. Across horizons, HSI produces far more accurate forecasts of future house prices than standard housing market determinants – a result that holds both in-sample and out-of-sample. Overall, HSI tracks the housing market with relatively high accuracy. The index captures not only the turbulence surrounding the financial crisis and the more stable period the housing market has experienced in recent years, but also the unusual development in house prices following the outbreak of the Covid-19 pandemic.

Demand for housing is generally believed to be a function of key macroeconomic variables such as interest rates, employment and credit conditions. To better understand the mechanism behind housing search activity, we examine the relation between HSI and a range of variables typically used to explain dynamics in the housing market. We find that internet search for housing has a negative correlation with the level of the mortgage rate, indicating that households intensify search in times with low financing costs. Otherwise HSI has a relatively low correlation with key housing market determinants as well as with various risk premium proxies.

Google Trends provide data also on local online search volume. This is a key advantage relative to macroeconomic data since housing markets tend to be local in nature (Del Negro and Otrok, 2007, Gyourko et al., 2013, Glaeser et al., 2014, and Hernández-Murillo et al., 2017). In regressions across 77 Metropolitan Statistical Areas (MSAs), we show that local housing search is a strong predictor also for local house prices, generally explaining more than 40% of the one-month-ahead variation in MSA-level house prices. Furthermore, controlling for national search activity, we show that local housing search remains a significant predictor of local house prices, which is direct evidence that housing markets are influenced by local search dynamics.

We next exploit cross-sectional variation in local housing markets to corroborate our interpretation that HSI is a proxy for latent housing demand. We do so along two dimensions. First, our MSA-level regressions show a large dispersion in the economic effect on house prices from changes in search activity. Provided that HSI captures variation in housing demand, we would expect to see a larger economic effect in local housing markets with a more constrained housing supply. Using the supply inelasticity measure of Saiz (2010), we show that this is indeed the case. Second, theoretical search-based models imply that the time it takes for a house to be sold should fall in response to an increase in demand. We test this using HSI as proxy for demand and find supportive evidence of a negative relation between HSI and time-on-market.

The Covid-19 pandemic caused a massive shock to the U.S. economy and housing market and we would not necessarily expect the relation between search activity and house prices to remain robust during this period. To explore the impact of the pandemic on our results, we estimate an MSA-level panel model that includes HSI along with a measure of housing supply (for-sale inventory data from Zillow) and a stringency index for Covid-19 lockdown measures. We find that demand and supply effects along with Covid-19 restrictions combine to capture nearly two-thirds of the monthly variation in house prices across MSAs during the pandemic.

Other papers have studied the relation between online search and housing. Wu and Brynjolfsson (2015) find that search data are more effective for predicting house transactions than for predicting house prices and that online search has rather limited predictive power over house prices. This contrasts with our findings, but the reason for the difference is easy to comprehend: Wu and Brynjolfsson use two broad, predefined search categories (real estate listings and real estate agencies) containing several individual search terms, complicating the economic interpretation of their search activity measures. Conversely, we explicitly use terms that capture search activity from potential house buyers and therefore are more strongly related to housing demand and have a highly significant predictive power over variation in house prices across several horizons. Beracha and Wintoki (2013) use search volume for "real estate i", where "i" is the name of a city. They show that abnormal search volume for a city lead to abnormal changes in house prices for that city. We find that our suggested procedure has considerable stronger predictive power over future house prices compared to the procedure used by Beracha and Wintoki (2013).[1]

Our analysis is also related to the literature that exploits online search activity to measure peoples' attention and its impact on asset prices. For example, Da et al. (2011) construct a direct measure of investor attention through online search activity for individual stock tickers and show that an increase in attention predicts higher stock prices in the ensuing two weeks. At a more aggregate level, Da et al. (2015) use daily search activity to construct a Financial and Economic Attitudes Revealed by Search (FEARS) index using keywords such as recession, unemployment and bankruptcy. They find that the index predicts short-term return reversals as well as temporary increases in volatility.[2] Andrei and Hasler (2015) provide both a theoretical framework and empirical results which support attention as a key determinant of asset prices.

We contribute to this literature by showing that demand for housing as measured through online search activity is a strong predictor of house prices. The predictive ability of search activity for house prices follows naturally from the high search intensity involved in buying a house as well as the frictions present in the housing market. Consequently, search activity has a relatively large and long-lasting impact on future house prices – both in absolute terms and when compared to other asset classes.

Our paper is also directly related to the literature on predictability of house prices, including studies such as Rapach and Strauss (2009), Plazzi et al. (2010), Ghysels et al. (2013), Soo (2018),

[1] We focus on the link between search behavior and *future* house prices. Gargano et al. (2021) study the reverse relation, namely how *past* price growth explains differences in search behavior across prospective home buyers. They find that prospective home buyers experiencing higher growth in their postcode of residence search more broadly across locations and house characteristics.

[2] Joseph et al. (2011) also find that the more difficult stocks are to arbitrage, the stronger the link between search intensity (as measured by online ticker search) and future returns.

Cox and Ludvigson (2019), and Bork et al. (2020). This literature typically uses either economic variables such as interest rates, employment and credit conditions or sentiment-based variables as predictors. The underlying intuition here is that supply and especially demand are largely driven by these variables which, consequently, contain important information about future house prices. We extend this literature by proposing a more direct measure of demand and show that it strongly outperforms standard variables used to predict future house prices.

In addition to the higher predictive accuracy of our HSI measure, there are several other advantages of using online search data in forecasting house prices compared to data gathered from government agencies. Many macroeconomic variables are often announced with a substantial time delay, only available at a low frequency, and subject to substantial data revisions, complicating real-time forecasting. In contrast, Google search data are readily available at a high frequency without time delay and are not subject to data revisions.[3]

The rest of the paper is structured as follows. Section 2 explains how we build on the theoretical insights from search-based models as well as how we measure housing demand and construct the national and local search indices. This section also contains an analysis of how HSI relates to standard housing market determinants. Section 3 contains an empirical analysis of the predictive power of search activity in the housing market over future house prices. Section 4 explores variation in local housing markets and relates our findings to variation in local supply elasticities. Section 5 analyzes the housing market during the Covid-19 pandemic. Finally, Section 6 contains concluding remarks.

2 Search Activity in the Housing Market

Online search volume has been shown to track investor sentiment in stock and bond markets (Da et al., 2015). It is plausible to expect that search activity also contains valuable information for tracking and quantifying variation in the demand for housing – a highly complex and segmented market. Specifically, aggregate internet search volume for phrases such as "buying a house" is likely to reflect genuine interest in actually buying a house and should thereby provide a timely and observable signal that is correlated with the underlying (latent) variation in housing demand.

[3] Guo (2009) and Ghysels et al. (2017) show that asset return predictability from macroeconomic data tends to be considerably weaker when using real-time macroeconomic data as opposed to using revised macroeconomic data.

2.1 Search as a Leading Indicator for Housing Demand

We start by motivating our choice of housing search activity as a leading indicator for demand in the housing market, building on theoretical work from the search and matching literature. The idea behind these search models is that since no central clearing house exists, buyers and sellers look for each other until they are matched. Since search is a costly activity, agents will aim at optimizing the effort over time. Several models within this framework imply that positive (negative) demand shocks lead to subsequent positive (negative) house price changes, motivating why housing search as a proxy for demand should have predictive power over future house price changes.[4]

Piazzesi et al. (2020) point out that, although supply in the housing market can be proxied by the number of homes available for sale in a given market, demand (the number of potential buyers), remains unobserved. A similar observation is made by Han and Strange (2015) who argue that although we have measures for seller time-on-market, there is no parallel for buyer time-on-market as a proxy for buyer search. Since buyers are arguably more active than sellers, empirical research on buyer search intensity is essential for reaching a better understanding of housing markets. Our paper attempts to make up for this shortcoming, arguing that we can use internet search activity, segmented by local markets at the MSA level, as a proxy for the search behavior of home buyers across time.

Our study is related to Piazzesi et al. (2020) who document that search activity is positively correlated with house prices in the cross-section of U.S. cities. In contrast to their study, we characterize search intensity dynamics over time at an aggregate MSA level across the U.S., instead of focusing on cross-sectional search for individual houses at a single point in time. Our study confirms the positive relationship between search activity and prices but by analyzing the time series dimension, we can capture the effect of current search intensity on future price appreciation. In this respect, our study confirms the theoretical predictions of Berkovec and Goodman (1996), who present a model in which frictions in the search and matching process imply that current demand shocks impact not only current but also future house price changes. In their model, buyers and sellers have imperfect information about the underlying market conditions, implying that price expectations adjust gradually in response to a demand shock.[5]

[4] See Han and Strange (2015) for a detailed survey of the literature on housing search models.

[5] Krainer (2001) and Novy-Marx (2009) also analyze frictions in the search and matching process of home buyers and sellers.

Carrillo et al. (2015) develop a search and matching model in which measures of market tightness, defined as the ratio of buyers and sellers in the market, predict future house price changes. More buyers entering the market during times of increasing demand leads to market tightness which in turn is followed by an increase in the bargaining power of sellers and higher likelihood of a sale. Since buyers and sellers do not hold perfect information about market conditions (e.g., the size of demand shocks), an increase in market tightness today leads to an increase in house prices in the future.[6] Other search-based models can generate similar mechanisms of sluggish price adjustments. For instance, building on Wheaton (1990), Diaz and Jerez (2013) specify a search model that propagates the effect of aggregate shocks to future periods. A key element of their model is that search and matching frictions produce trading delays such that not all agents seeking to buy a new home can do so right away, implying that the effect of aggregate shocks is propagated to future periods. Genesove and Han (2012) develop a search and matching model in which lagged seller response, due to gradual adjustment of the seller's reservation price, results in sluggish price adjustments after a demand shock. In a similar vein, Head et al. (2014) show that time-consuming search and matching generates sluggish price adjustments in response to a shock.

Taken together, the theoretical insights from search-based models imply that a shift in current demand will lead to future price changes, which is the main hypothesis of this paper.

2.2 Construction of the Housing Search Index

To quantify internet search activity, we use Google Trends data from which we obtain a time series index on the volume of queries for a given search term in a given geographic area.[7] Google Trends provides a set of related queries for every main query. The list of related queries (or, equivalently, *related terms*) includes between 0 and 25 different terms, with the final number depending on the search volume of the main query, i.e. high volume series will usually have 25 related queries while lower volume series will feature fewer. Google does not disclose the methodology it uses to select related queries, but the resulting terms are usually intuitively related to the main query. From the perspective of quantifying housing demand, this feature is appealing for two reasons. First, each

[6] van Dijk and Francke (2018) create a proxy for tightness in the Dutch housing market which relates positively to changes in house prices.

[7] Google dominates the U.S. search engine market with a 63 percent market share as of October 2018 (Statista, 2018). Data on search volume are also available for other services owned by Google such as Image Search, News Search, Google Shopping and YouTube Search, but these account for far smaller volumes than general Google searches.

semantically related keyword can provide additional information about housing demand beyond that contained in the original query. Second, since related terms are likely to be correlated, this induces a natural factor structure which allows us to build an aggregate measure of housing demand.

Google Trends data are available from 2004 onwards. Our sample runs from 2004:1 to 2021:1 at the monthly frequency.[8] To obtain a simple and clean measure of housing demand, we initially use "buying a house" as our main search term and subsequently obtain a list of 22 related terms: "when buying a house", "buying a home", "buy a house", "mortgage", "buying a new house", "before buying a house", "how to buy a house", "real estate", "steps to buying a house", "buying a house calculator", "first time buying a house", buying a house process", "house buying process", "homes for sale", "building a house", "buying a house with bad credit", "cost of buying a house", "buying a house to rent", "mortgage calculator", "houses for sale", "buying a house tips", and "buying a foreclosure house". These search terms are all related to the home buying process and as such should proxy for housing demand. The three remaining related search terms are excluded either because they are unrelated to housing ("buying a car") or because the search volume is low. We define low volume series as those for which more than 10% of observations equal zero.[9]

Some of the related terms may be measured with more noise than others. To filter out the noise and more accurately estimate latent demand, we use a targeted PCA approach which ensures that only the most relevant search indices are included to compute the latent demand factor. Specifically, our implementation follows Bai and Ng (2008) as we use the elastic net estimator of Zou and Hastie (2005) to select the ten most relevant search indices and then apply principal component analysis to summarize the most important information from these indices into one common component. We interpret this principal component as a summary measure for housing search and refer to it as the Housing Search Index (HSI).[10]

Before extracting the first principal component, we transform the search indices as follows. Following Da et al. (2011, 2015) and Vozlyublennaia (2014), we first convert the series to their natural

[8] As noted by D'Amuri and Marcucci (2017), Google Trends are created based on a sample of queries that change according to the time and IP address used to download the data. To account for sampling error, we compute the index for all Google Trends queries using an average over 15 different days. The correlation across different samples is always above 0.99. Hence, the results are, for all practical purposes, robust to this issue.

[9] The two excluded terms are "help buying a house" and "buying a house cash".

[10] Our main goal is to produce a simple and easy-to-interpret index of housing search, which is why we use a simple targeted PCA approach. However, the predictive results that we report below are generally highly robust to using more advanced machine learning techniques. We refer to Section A.7 in the Online Appendix for further details.

logarithm.[11] To account for the possibility that the individual Google Trends series could follow different trends, we adopt a sequential testing strategy in the spirit of Ayat and Burridge (2000) and similar to Borup and Schütte (2022).[12] We further remove seasonality by regressing each series on monthly dummy variables and study the residuals from this regression.

2.3 Housing Search and Prices

Panel A in Figure 1 displays a time series of HSI along with the log growth rate in the seasonally adjusted monthly Federal Housing Finance Agency (FHFA) purchase-only house price index for the United States.[13] Housing search and growth in house prices move closely together. In particular, we note that HSI captures the negative growth rates in 2009-2010 that followed the collapse in the housing market, the subsequent recovery, as well as the more stable house price growth seen in recent years. HSI also captures the unusual development in house prices following the outbreak of the Covid-19 pandemic. As an initial response to Covid-19, house prices dropped slightly but subsequently experienced large positive growth rates – a development mirrored in the time-series movements of HSI.

To explore the dynamic relation between HSI and movements in house prices, Panel A of Figure 2 shows regression slope coefficients, associated t-statistics and R^2-values of monthly price changes from $t-1$ to t on lagging, contemporaneous and leading values of HSI:

$$p_t - p_{t-1} = \alpha_j + \beta_j HSI_{t+j} + \varepsilon_t, \ j = -12, ..., 12, \tag{1}$$

where p_t is the log of the FHFA house price index in month t. We find much larger coefficients

[11] There is no consensus in the literature as to whether Google Trends data are best characterized by stationarity, trend stationarity or a unit root since this can be very sensitive to the query in question. Vozlyublennaia (2011), Choi and Varian (2012), Bijl et al. (2016) and D'Amuri and Marcucci (2017) do not perform any differencing or detrending of the series, which suggests that the Google Trends data they use are stationary. Da et al. (2015) study the log-differences (growth rates) of their data.

[12] The idea is to successively test for stationarity, linear trend stationarity and quadratic trend stationarity using an augmented Dickey-Fuller (ADF) test. Specifically, the first test computes an ADF test with a constant term. If the null of non-stationarity is rejected, we stop and use the series without any transformation; conversely, if the null is maintained, we use an ADF test that includes both a constant and a linear time trend. If the null of this second test is rejected, we linearly detrend the series by using the residuals of a regression of the series on a constant and a time trend; otherwise we compute a final ADF test that includes a constant, a linear trend and a quadratic trend. If we reject the null of this test, we detrend the series but include a quadratic trend in the regression.

[13] It is well-known that house prices display strong seasonal variation with high prices during spring and summer and low prices during fall and winter. Section A.13 in the Online Appendix shows that a non-seasonally adjusted housing search index to a large extent captures the seasonal component in house prices.

and R^2-values using lags rather than leads of HSI, suggesting that movements in HSI precede movements in the FHFA house price index. The strongest statistical relation between HSI and changes in house prices occurs at lags of HSI ranging from one through four months. At these lags, the predictive power of HSI over monthly house price changes is more than 50%. Leads of HSI are also significantly related to house price changes, but increasing the lead length substantially reduces the magnitude of the slope coefficient, the degree of statistical significance, and R^2-values.

Table 1 shows results from tests of bi-directional Granger causality between HSI and house price changes. Regardless of lag length, we generally find that the Granger causality runs from HSI to house price changes and not the other way around, once two or more lags are included. Overall, the results indicate that HSI is a leading indicator of subsequent changes in house prices – a point we explore more in-depth in Section 3.

2.4 Housing Search and Transactions

If online search activity provides an accurate signal about peoples' intentions of buying a house, we should expect to find a positive relation between HSI and subsequent house sales. To explore this relation, Panel B of Figure 1 displays HSI along with monthly sales of existing housing units from the National Association of Realtors (NAR). The figure shows a strong positive relation between online search activity and house sales, which supports the conjecture that people only engage in a costly search process if they have true intentions of completing a transaction. The figure also shows that HSI tends to lead home sales, as we observe a substantial decrease in search activity prior to the large drop in house sales leading up to the financial crisis and likewise an increase in search activity prior to the increase in sales in 2009 and 2011-2012. Even in the unexpected event of the Covid-19 pandemic, we see how HSI leads house sales first with a small decrease in search activity as an initial response to the outbreak of the disease followed by a historically high degree of housing-related online search mirroring the development in transactions.

To evaluate the lead-lag relation between HSI and house sales, we undertake a similar regression analysis as that performed in equation (1):

$$sales_t = \alpha + \beta HSI_{t+j} + \varepsilon_t, \quad j = -12, ..., 12, \tag{2}$$

where $sales_t$ is the sales of existing single-family housing units from NAR in month t. Panel B of Figure 2 shows the slope coefficients, associated t-statistics and R^2-values as functions of j. The large values for $j < 0$ strongly suggest that search activity leads house sales. In contrast, we see no discernible relation between sales and future search activity, suggesting that increased sales activity does not prompt an increase in the volume of searches for buying a house. Consistent with this, the Granger causality tests in Table 1 imply that HSI helps to forecast home sales, while the reverse does not hold.[14]

Taken together, Panels A and B in Figure 2 suggest that online housing search volume leads both house prices and home sales but that the lead times are very different, being notably shorter (1-4 months) for house prices than for actual home sale transactions (10-12 months).

2.5 Demand Interpretation

It is important to validate our interpretation of HSI as a measure of latent demand. As mentioned in Section 2.1, although measures of housing supply are readily available, direct measures of demand are much harder to obtain. The best alternative measure of demand we could find comes from Redfin, which is one of the largest real estate brokers in the U.S. (https://www.redfin.com/). Redfin's Housing Demand Index (HDI) is described by the company as "the industry's first and only measure of housing activity prior to purchase". The index is based on the number of customers requesting home tours and writing offers in major metro areas of the U.S. Tours are weighted by averaging the number of tours per written offer.[15] This measure is arguably also closely related to the measure used by Piazzesi et al. (2020) who use user click data from the real estate website Trulia.com to measure search activity.

We believe Redfin's data is representative of the U.S. population of home buyers for two reasons. First, Redfin's website received more than 24 million unique visitors per month in 2020, and it is currently ranked as the 4th largest online brokerage in the U.S. by online market share. Second, in contrast to other competitors (such a Trulia.com) the company also operates offline using local brokers. This offline activity provides the company with a broader coverage of customers, making

[14] Home sales is highly persistent with an AR(1) coefficient of 0.97. As a robustness check, we also conducted the Granger causality tests using the first difference of home sales, which led to the same conclusion, namely that the Granger causality runs from HSI to home sales and not the other way around.

[15] More details on the methodology can be found here: https://www.redfin.com/news/redfin-housing-demand-index-methodology/

it more representative of the U.S. population as a whole.

HDI is constructed at a weekly frequency and starts in the first week of 2018. As a first validation we aggregate HDI to a monthly frequency by taking the average value per month and plot it together with HSI. As Panel A in Figure 3 illustrates, the two indices closely follow each other (correlation of 0.83) and both capture the large spike in demand that happened during the latter part of the Covid-19 period. To further validate our interpretation, we exploit the fact that Google Trends can be obtained at a weekly frequency and construct a weekly version of HSI using the same keywords and methodology as we use for the monthly version. Since the HDI data at the time of writing spans until the first week of April 2021, we expand the sample for the weekly HSI to end at this date, giving us a total of 170 observations.[16] The weekly HSI is plotted together with the Redfin's HDI in Panel B of Figure 3. The correlation between the two weekly indices is 0.93.

One natural concern from the above validation exercise is that the high correlation between HSI and Redfin's HDI makes HSI a superfluous measure of demand. However, HSI has advantages over HDI. First, Redfin's HDI is only available from the first week of 2018, whereas HSI starts in 2004, providing a longer historical sample to draw inference. Second, HSI can be constructed in real time for any MSA for which there is Google data, whereas HDI is only available (at the time of writing) at the national level and for 15 metropolitan areas. Finally, HSI tends to capture demand earlier than Redfin's HDI. In Table 1, we show the results from tests of bi-directional Granger causality between HSI and HDI. The tests show that lagged values of HSI have predictive power over HDI and not the other way around. These results suggest that HSI is able to capture latent demand earlier than HDI. The most likely explanation for this relation is that people often use Google to explore the housing market before actually committing to a specific realtor.

2.6 Housing Search and Other Housing Market Variables

Housing search activity is likely to be correlated with a variety of other economic variables. It is therefore important to address to what extent we can explain variation in housing search by means of macroeconomic fundamentals and other determinants of outcomes in the housing market. For example, does housing search increase in periods with low interest rates, high employment, good

[16] Results using a sample that ends in January 2021, as with the monthly series, are almost identical.

credit conditions, and high sentiment? Moreover, does housing search still predict movements in house prices after controlling for other economic variables?

To better understand the drivers behind housing search, we regress HSI on a set of commonly used housing market determinants. Motivated by studies such as Rapach and Strauss (2009), Plazzi et al. (2010), Ghysels et al. (2013), Bork and Møller (2018), Cox and Ludvigson (2019) and Bork et al. (2020), we include the following set of variables in our analysis:

- Employment (*payrolls*): The year-over-year log employment growth rate (total nonfarm payrolls).
- Inflation (*infl*): The log difference in the Consumer Price Index for all urban consumers (all items).
- Building permits (*permits*): The log difference in new private housing units authorized by building permits.
- Housing starts (*starts*): The log difference in new privately owned housing units.
- Term spread (*term*): The 10-year treasury constant maturity rate minus federal funds rate.
- Mortgage rate (*mort*): The level of the 30-year fixed mortgage rate.
- Price-rent ratio (*pr*): The log ratio of the house price to the rent of primary residences.
- Loans outstanding (*loans*): The log change in commercial and industrial loans outstanding.
- Sentiment (*sent*): Fraction of respondents who answer that now is a "good time" to buy a house from the University of Michigan's Survey of Consumers.[17]

In addition, we include broad economic activity indices and risk premium variables:

- The Chicago Fed National Activity Index (*cfnai*).
- The Aruoba, Diebold, and Scotti (2009) Business Conditions Index (*ads*).
- The price-dividend ratio (*pd*): The log ratio of the S&P500 index and twelve month trailing dividends.[18]

[17] All other above variables are from the Federal Reserve Bank of St. Louis (FRED) database.
[18] Data are obtained from Robert Shiller's website.

- Risk aversion (ra): Risk aversion index from Bekaert et al. (2021).
- Uncertainty (unc): Uncertainty index measured in annual volatility units from Bekaert et al. (2021).

Table 2 shows the results from the contemporaneous regression model

$$HSI_t = \alpha + x_t'\beta + \varepsilon_t, \tag{3}$$

where x_t contains the standard housing market determinants either individually in univariate regressions (left column) or combined in a multivariate regression (right column). In the univariate regressions, common house price predictors such as employment, inflation, the term spread, price-to-rent ratio, loans outstanding, and sentiment bear little-to-no relation to the volume of housing search. Building permits and housing starts are both significantly positively related to housing search volume. In addition, stock market based risk premium proxies such as the price-dividend ratio are significantly related to search activity. However, with R^2 values around 3-9%, these variables explain only a very small part of variation in HSI. The most striking relation is between the level of the mortgage rate and search activity with a strong indication that periods of low mortgage rates coincide with periods of high search activity.

Combining our full list of standard housing market determinants in a multivariate regression (right column), we notice that some of the results in the univariate regressions are reversed. Search activity is now strongly positively related to the price-rent ratio, indicating that search volume tends to increase in times with high valuation ratios. However, even with the full list of standard housing market determinants, we can only explain around 70% of the variation in HSI. With roughly 30% of the variation in HSI left unexplained, a relatively large component of time-series movements in the volume of housing search is, thus, uncorrelated with standard activity measures from the housing market.[19]

[19] Some of the information contained in these variables might represent noise that just happens to be correlated with HSI. We analyze this possibility in Section A.3 in the Online Appendix where we consider a placebo test that generates artifical times series by resampling from the panel of regressors. When regressing HSI on the 14 artifical time series, the median R^2 across bootstrap replications is above 20%.

2.7 Local Housing Search

Online search activity can be used to quantify a local component in housing demand. Specifically, Google Trends can be used to extract search activity that occurs within smaller geographical areas, allowing us to study the importance of housing search in the cross-section of local housing markets. This is an important feature because existing evidence suggests that local market factors help explain movements in house prices across the U.S. (e.g. Del Negro and Otrok, 2007, and Hernández-Murillo et al., 2017).

We later analyze whether the effect from housing search activity on house prices depends on local housing supply. To do this, we use Saiz's (2010) supply elasticity measures across Metropolitan Statistical Areas (MSAs). Saiz (2010) provides results for the 95 MSAs with a population over 500,000 in 2000. Google defines metropolitan areas slightly differently from the U.S. Office of Management and Budget (OMB) which leads us to exclude 18 MSAs from our analysis. For the remaining 77 MSAs there is a one-to-one mapping between the definitions of Google and OMB.

We define local housing search using the same keywords as for the aggregate U.S. housing market and exploit that Google Trends automatically includes geographical idiosyncrasies of home buyer search patterns in each MSA through the related terms. In this way, the search data will be heavily localized. While search activity for individuals residing in a given MSA counts in the overall search volume for that particular MSA, some individuals may also be interested in buying a home in one of the neighboring MSAs. To allow for such potential moves across MSA borders, we also include search activity in the state in which the MSA is located. Based on the local search activity, we construct local HSI measures using the same targeted PCA approach as for the national index.

To illustrate the differences across local housing markets, Figure 4 shows the local HSI along with the growth rate in the local Freddie Mac house price index for Miami and Wichita. Among the 77 MSAs included in our analysis, Miami and Wichita have the lowest and highest supply elasticity, respectively, cf. Saiz (2010). For Miami we see a very similar pattern in house prices as compared to the national market, although with a larger boom-bust cycle. We also observe a very strong relation between the local HSI and growth in house prices similar to that found for the national market. In contrast, house prices in Wichita did not experience a boom-bust cycle from 2004 to 2010 and monthly growth rates never stray far away from zero. We should therefore expect differences in the economic effect on local house prices from shocks to local $HSIs$, a point we explore in Section 4.

3 Search Volume and Predictability of House Prices

If online search activity tracks peoples' intentions of buying a house – and thus proxies for the demand for housing – we would expect increases in HSI to be associated with higher subsequent house prices. Given various frictions in the housing market, an increase in demand is propagated into future periods, which leads sluggish price adjustment in response to an increase in demand (Berkovec and Goodman, 1996, and Carrillo et al. 2015). Figures 1 and 2 support this conjecture by showing a strong positive relation between housing search and future growth in house prices.

To more formally explore the predictive power of housing search with respect to movements in house prices, we estimate predictive regressions

$$p_{t+h} - p_t = \alpha_h + \beta_h HSI_t + \theta_h' Z_t + \varepsilon_{t+h}, \tag{4}$$

where p_t is the log of the FHFA house price index, h is the forecast horizon, HSI is the predictive variable, and Z_t is a vector of control variables. We consider four different horizons, namely $h = 1, 3, 6$ and 12 months. To account for autocorrelation in house price growth (Case and Shiller, 1989) and overlaps in the dependent variable in (4) when $h > 1$, we compute bootstrap standard errors from a circular block bootstrap that resamples the data in blocks of consecutive observations, reproducing serial correlation and other dependencies in the data.[20,21]

Panel A in Table 3 reports the estimate of β_h, the corresponding bootstrap t-statistic in parenthesis, and the R^2 in square brackets. HSI is a strong predictor of future house prices with highly significantly positive slope estimates, consistent with future house prices rising when current search (demand) for housing is high. The predictive power of HSI is high when measured by the R^2, which ranges from 56% for $h = 1$ to around 70% for $h = 6$. The economic magnitude is also large, as a one standard deviation increase in HSI is associated with a 0.4% increase in expected house price growth at the one-month horizon rising to 4.4% at the one-year horizon. For example, starting from May 2020 and until the last sample point in January 2021, HSI stays more than two standard

[20] We resample the regressand and the regressor jointly in blocks with an average size of 24, which is close to the optimal block length according to the Politis and White (2001) automatic selection procedure. Section A.1 in the Online Appendix shows results for other choices of the average block length as well as results using Hodrick (1992) t-statistics aimed at circumventing the issue with overlapping data.

[21] We also examined the possibility that the predictive regressions suffer from small sample bias arising from cross-correlation in error terms as studied by Stambaugh (1999). We can rule this out as the innovations in HSI are only weakly correlated with those of the predictive regression.

deviations above its mean, which gives to rise to monthly house price growth predictions of more than 1% each month throughout this period. We see from these results that the effect of HSI on predicting house price changes is both statistically and economically significant.[22]

Panel B of Table 3 controls for all 14 variables described in Section 2.6. We see that HSI retains its statistical significance across all horizons, although its slope coefficient is somewhat reduced.[23] These results suggest that housing search activity carries important information about future house prices over and above that embedded in standard housing market predictors. To further verify this claim, we use the residuals from the multivariate regression in Table 2 to construct a version of the housing search index, $HSI^{\perp}$, that is orthogonal to the standard predictors. Panel C in Table 3 shows that the slope coefficients for $HSI^{\perp}$ remain positive and significant across horizons. This evidence suggests that HSI contains complementary information about future house prices that is not subsumed by standard housing market variables. In terms of predictive power, the R^2 generated by $HSI^{\perp}$ ranges from around 7% to 11% across horizons, which is still sizeable given that we have cleaned HSI from all information embedded in 14 control variables.[24] Furthermore, we analyze whether $HSI^{\perp}$ granger causes house price growth rates. As shown in Table 1, lagged values of $HSI^{\perp}$ predict future house price growth rates, while the reverse is not the case.

Finally, in Panel D of Table 3, we control for the AR(1) component in house price growth. A number of studies have documented that growth in house prices exhibit positive serial dependence, which can arise due to frictions and illiquidity and may also reflect the procedure used to construct the house price indices (Ghysels et al., 2013).[25] The results show that HSI stays strongly statistically significant when controlling for the AR(1) component. The coefficients on HSI are reduced across horizons, but the economic magnitude of the predictability remains substantial. The results suggest that HSI contains relevant information about future house price growth rates not already embedded in the AR(1) component.

[22] Across the four horizons, HSI delivers an R^2 that is at least 10 percentage points higher than the best performing individual search terms such as "buying a house", "when buying a house", "buying a home", "buying a new house", "building a house", and "cost of buying a house" which produce R^2 values around 20-50% across the four horizons. This shows the value added by extracting common information from a broad set of search terms.

[23] Section A.2 in the Online Appendix reports estimation results for the control variables, which are often insignificant.

[24] Some of the lost explanatory power may reflect that we control for a large number of variables and some of these variables may by chance explain part of the variance of HSI without being relevant variables for predicting house prices. In that sense, the R^2 values reported in Panel C of Table 3 may be viewed as a conservative measures of the additional predictive power gained from using HSI.

[25] The FHFA calulates their monthly repeat-sales house price index without the use of temporal aggregation. In contrast, the monthly Case-Shiller house price index is based on a three-month moving average window, implying that this index is substantially more autocorrelated than the FHFA index.

3.1 Comparison with Wu and Brynjolfsson (2015)

HSI focuses on the buying side of the housing market through the chosen keywords. Accordingly, we interpret the search index as a proxy for latent demand. Wu and Brynjolfsson (2015) also consider the use of online search activity to predict house prices and sales. Instead of using specific keywords, they consider predefined search categories supplied by Google Trends, namely "Real estate agencies" and "Real estate listings". Google classifies search queries into categories using an undisclosed natural language classification engine (Choi and Varian, 2012) and it is unclear how we should interpret these categories other than that they relate to the topic given by the name of the category. Wu and Brynjolfsson (2015) find that these two search categories hold some predictive power for future house prices, but also that prices are more difficult to predict than house sales.[26]

To facilitate a direct comparison with Wu and Brynjolfsson (2015), Table 4 explores the predictive power of the two search categories "Real estate agencies" and "Real estate listings" and compares these to HSI.[27] Panel A shows that the two predefined categories hold some predictive power for growth in house prices with R^2 values ranging from 9% for $h = 1$ to 18% for $h = 12$. Compared to HSI this degree of predictive power is, however, of limited magnitude, which is also evident from Panel B, where we use all three search-based predictors simultaneously. These results strongly suggest that a more carefully chosen set of keywords with a clear economic interpretation is important for the predictive power of online search compared to broad search categories.

Panel C in Figure 1 plots the two predefined search categories along with the log growth rate in the FHFA house price index. Compared to Panel A in the same figure, "Real estate agencies" and "Real estate listings" do not capture movements in house prices to the same extent as HSI. In particular, we notice that the predefined categories show an increase in search activity during the first part of the bust period and lag house prices in the second part of that period. They also do not spike after the outbreak of the Covid-19 pandemic. Regressing HSI on the two search categories "Real estate agencies" and "Real estate listings" produces an R^2 of only 4%, suggesting that these predefined search categories are only weakly correlated with HSI.

In summary, we confirm Wu and Brynjolfssons (2015) finding that the predefined search categories "Real estate agencies" and "Real estate listings" to some degree can predict future house prices,

[26] Dietzel (2016) uses a similar approach for real estate subcategories to analyze turning points in housing markets.

[27] We detrend and deseasonalize the predefined search categories similar to the other search indices as described in Section 2.2.

but also find that their predictive power is limited. A likely explanation of this is that these broad categories reflect both the buying and selling sides of the housing market. Our much stronger prediction results based on HSI suggest an additional explanation, namely that the predefined categories contain too much irrelevant information which reduces their predictive power.[28]

3.2 Predictability at Longer Horizons

Table 3 covers forecast horizons up to 12 months. Searching for a house is often a lengthy process so it is not surprising that HSI displays strong predictive power also over long horizons up to a year. However, we would also expect that its predictive power declines for very long horizons since home buyers have an incentive to limit the search period to avoid excessively large search costs. To visualize the predictive power over very long horizons, Figure 5 summarizes the slope coefficients, associated t-statistics and R^2 values for horizons up to five years ($h = 60$). The figure shows that HSI is a significant predictor of house price growth up to a horizon of roughly five years, but also that the explanatory power steadily declines after its peak at horizons around 3 to 8 months. Our finding that short-horizon effects are larger than long-horizon effects is consistent with the search models of Berkovec and Goodman (1996) and Genesove and Han (2012).[29]

Our 17-year sample from 2004-2021 means that we only have a limited number of independent observations at the longer horizons. Caution should therefore be exercised when interpreting these results, especially at the longest horizons. However, a decline in the predictive power at a horizon of roughly 8 months seems plausible given the time it typically takes to buy a home from the initial search process to closing the deal. NAR (2020) reports that the typical search time for a home is 10 weeks. Prior to searching for a home, buyers are likely to gather information about the house buying process itself. Once a buyer has found a house, the buyer and seller have to agree on a price, the house must be inspected, and the loan application must be approved, with the latter steps typically taking 40-50 days. As such, the predictive pattern of HSI is different from conventional

[28] We also compared our procedure with that of Beracha and Wintoki (2013), who analyze search activity for a particular MSA by using the search term "real estate i", where "i" is the given MSA. We find that our procedure has considerably stronger predictive power over future house prices than that of Beracha and Wintoki (2013). For example, in Miami, Toledo, and Houston, our local HSIs generate R^2s of 69%, 64%, and 51% at the one-month horizon compared to 20%, 9%, and 0% when using "real estate Miami", "real estate Toledo", and "real estate Houston", respectively. In general, the series of Beracha and Wintoki (2013) show weak co-movement with HSI.

[29] In the dynamic search model of Berkovec and Goodman (1996), a key mechanisms is lagged price responses to demand shocks. In their simulations, the price adjustment takes place within a few months following a change in demand.

predictors such as the price-rent ratio for which the predictive power builds up over long horizons.

3.3 Alternative Measures of House Price Changes

So far we have focused on the FHFA purchase-only house price index, which is one of the primary indices used in the literature. To illustrate that the predictive power of HSI is not only restricted to the FHFA house price index, we next consider other commonly used house price indices, namely the Case-Shiller national home price index, the Freddie-Mac house price index, and the Zillow home value index for single-family homes. The FHFA, Case-Shiller, and Freddie-Mac indices are similar in the sense that they are all constructed using a repeat sales methodology. In contrast, the Zillow index instead uses a valuation model to estimate prices for individual homes. All four indices differ in geographic coverage, price information source, and weighting scheme to form aggregate indices. These differences could be important when analyzing house price predictability, so Table 5 shows predictive results for each of the four house price measures. Despite the different methodologies used to measure house prices, our results show that the strong predictive power of HSI holds for all four house price indices implying that the choice of house price index is less important.

Another question is whether the predictive ability of HSI extends to commercial properties which, unlike residential properties, are purchased entirely from an investment perspective.[30] Similar to residential real estate, commercial real estate is characterized by various frictions that may induce sluggish price adjustments. HSI is designed to capture demand for residential properties, but may capture common variation in the residential and commercial real estate markets. To investigate this possibility, we use the CoStar composite value-weighted index of commercial properties across the U.S., which is constructed based on the repeat-sale methodology. From Panel E in Table 5, we see that HSI significantly predicts the growth rate in prices of commercial properties, although the extent of predictability is lower both in terms of statistical significance and predictive power.

3.4 Out-of-Sample Forecasting Tests

To be practically useful for policy makers and households, it is critical that HSI could have been used to improve forecasts of house prices in real time. Moreover, full-sample predictive regressions

[30] Plazzi et al. (2010) provide evidence of significant time-variation in expected returns on commercial properties.

such as those reported in Table 3 potentially overfit the data.

To address these issues, we next consider a set of out-of-sample forecasting experiments in which we recursively compute HSI and estimate the coefficients of the predictive model using only information available at the time of the forecast. We use the first three years of our sample (2004-2006) as our initial estimation period and reserve the remaining sample (2007-2021) for out-of-sample testing.[31]

Table 6 reports Campbell and Thompson (2008) out-of-sample R^2 values (R^2_{OoS}) and Diebold and Mariano (1995) t-statistics (t_{DM}) for comparing predictive accuracy against a given benchmark. In each case, R^2 values are computed relative to a "historical average" benchmark that assumes constant growth rates in house prices. The null hypothesis is $R^2_{OoS} \leq 0$, while the alternative hypothesis is $R^2_{OoS} > 0$.

We find that HSI is able to explain more than 50% of the out-of-sample variation in next month's growth in aggregate house prices. The predictive power increases with the forecast horizon and reaches its peak for $h = 3$ with $R^2_{OoS} = 65\%$, declining to $R^2_{OoS} = 54\%$ for $h = 12$. Diebold-Mariano tests strongly reject the null hypothesis that $R^2_{OoS} \leq 0$ at all forecast horizons.

Further, the table shows that forecasts based on HSI strongly outperform forecasts based on popular determinants of house prices across all horizons. In most cases, these variables generate negative R^2_{OoS} statistics. Exceptions include building permits and housing starts, but in both cases the R^2_{OoS} statistics are not significantly positive and never exceed 3%. For $h = 12$, sentiment generates the largest R^2_{OoS} statistic of 14% among the standard predictors, but again it is not statistically significant according to the Diebold-Mariano test. Stock market based risk premium measures such as the price-dividend ratio do not outperform the historical mean benchmark. The same goes for commonly used business cycle indicators such as $cfnai$.[32] These results make it less likely that the predictive ability of HSI stems from a typical risk compensation channel.

To assess if the strong predictive power of HSI is restricted to certain periods in time, we follow Welch and Goyal (2008) and plot the difference in the cumulative sum of squared forecast errors

[31] We use an expanding estimation window but obtain similar results with rolling windows. When generating the out-of-sample forecasts, we account for a two-month publication lag in house prices.

[32] These variables sometimes generate extremely negative R^2_{OoS} statistics, which typically arises from very substantial movements in the variables during the Covid-19 period, but with predictions in the opposite direction of the movements in the housing market.

(CSSFE) for $h = 1$ in Panel A of Figure 6. The benchmark is again constant growth rates in house prices. An upward sloping CSSFE implies that HSI delivers better forecasts than the benchmark and vice versa if the CSSFE is downward sloping. Figure 6 shows that HSI holds important information about future house prices irrespective of the market conditions, but also that online search activity is especially useful in turbulent times as witnessed under the financial crisis in 2007-2009 and the Covid-19 pandemic.

Finally, we compare the predictive ability of HSI to that of an AR(1) model, which captures the positive persistence in house price growth rates. As shown in Panel A of Table 6, HSI generates higher R^2_{OoS} values than the AR(1) model across all horizons. To more formally compare HSI with the AR(1) model, we use forecast encompassing tests (Chong and Hendry, 1986):

$$y_{t+h} = \varpi_h^{HSI} \hat{y}_{t+h}^{HSI} + \varpi_h^{ar1} \hat{y}_{t+h}^{ar1} + \varepsilon_{t+h}, \tag{5}$$

where $y_{t+h} = p_{t+h} - p_t$ is the realized h-month ahead log price growth rate and $\hat{y}_{t+h}$ is the forecasted value using either HSI or the AR(1) component. We implement the test by estimating:

$$e_{t+h}^{ar1} = \varpi_h^{HSI} \left(e_{t+h}^{ar1} - e_{t+h}^{HSI} \right) + u_{t+h}, \tag{6}$$

where $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$ is the forecast error. We test the null hypothesis that $\varpi_h^{HSI} = 0$, which would imply that the AR(1) forecast encompasses the forecast of HSI. We also estimate the reverse regression and test whether $\varpi_h^{ar1} = 0$. Panel B of Table 6 reports the results. The estimates of ϖ_h^{HSI} are strongly statistically significant across all horizons, implying that HSI contains relevant information beyond what is contained in the AR(1) forecast. Moreover, the weight on forecasts from the HSI model exceeds the weight on forecasts from the AR(1) model at all horizons.

In conclusion, our out-of-sample analysis confirms the strong in-sample predictive ability of HSI and shows that online search activity is a consistently strong predictor of future house prices in turbulent as well as in calmer periods. The analysis also emphasizes the strong predictive power of HSI compared both to standard house price determinants that generally have difficulties predicting future house prices out-of-sample as well as forecasts from an AR(1) model.[33]

[33] Section A.6 in the Online Appendix reports results from a bootstrap analysis, which further supports the strong out-of-sample predictive power of HSI over future movements on house prices.

4 Variation in Search across Local Housing Markets

National accounts data are often limited in geographic scope, and a key advantage of Google Trends data is that they have few geographical restrictions. This fact is particularly important for our analysis because housing markets are local in nature and we would not expect nationally aggregated data to capture all the complexities of local housing market dynamics.

To explore the predictive power of local versions of HSI, we estimate MSA-level regressions,

$$p_{it+h} - p_{it} = \alpha_i + \beta_i HSI_{it} + \varepsilon_{it+h}, \tag{7}$$

where p_{it} is the log of the Freddie Mac house price index and HSI_{it} is the housing search index, both for MSA i in month t. Figure 7 summarizes the results through a scatter plot of the estimated slope coefficients (β_i) versus R_i^2 values across the 77 MSAs introduced in Section 2.7. To ease comparisons across MSAs all search indices are standardized and the slope coefficients are multiplied by 1,200 to measure the annualized change in house prices after a one standard deviation change in search activity. For brevity, we only present results for $h = 1$, but the conclusion is robust across longer forecast horizons as we will verify in a panel setting in Section 4.1. The strong predictive power of HSI at the national level reappears in individual local housing markets with slope coefficients that are significantly positive for all but one MSA and with 54 MSAs generating R^2 values above 40%.

Across the 77 MSAs, the estimated slope coefficients range from 0.06 (Scranton) to almost 15 (Stockton) on an annualized basis. This implies a large dispersion in the economic effect on local house prices from shocks to demand as proxied by search activity. For example, a one standard deviation increase in the local HSI leads to an annualized 11.9% increase in expected house price growth in Miami the following month, while the corresponding response is only 0.1% in Wichita.

4.1 Local Variation in Supply Elasticities

The effect of changes in demand on prices should, in theory, be stronger in markets where the supply response is more inelastic compared to markets with a relatively flat supply curve where the supply response is more elastic. For example, in the search-model of Novy-Marx (2009), the amplification effect of a shock to demand is stronger in markets where agents are less responsive.

Following Saiz (2010), we therefore analyze whether the effect of HSI on house prices is stronger in MSAs with a more inelastic housing supply. We start by estimating predictive panel regressions, which allow us to analyze the average predictive relationship across all MSAs. In particular, we regress the h-month-ahead log house price growth in MSA i on the lagged housing search index in MSA i, constraining the slope coefficients to be identical across MSAs but allowing for individual MSA-specific fixed effects, i.e., imposing $\beta_i = \beta_j$ in (7). Following Thompson (2011), we compute standard errors that are robust to heteroskedasticity as well as correlation along both the time and MSA dimensions. Panel A in Table 7 shows the results. Local HSI significantly predicts local house price growth rates across all horizons. The predictive power of the local HSI as measured by the within-R^2 continues to be very large and is roughly 35% across all four horizons. Moreover, consistent with the national evidence, increased local housing search activity is associated with positive future growth rates in local house prices.

We next interact the local HSI with the degree of housing supply elasticity as computed by Saiz (2010). This allows us to analyze whether house prices in MSAs with a more inelastic housing supply react stronger to changes in housing demand as measured by search activity. To test this effect, we estimate

$$p_{it+h} - p_{it} = \alpha_i + (\beta + \beta_E \times Elasticity_i)\, HSI_{it} + \varepsilon_{it+h}, \tag{8}$$

where $Elasticity_i$ is the supply elasticity measure of Saiz (2010). Panel B in Table 7 shows the results. The results show that the relation between variation in local search and house prices is significantly stronger in MSAs with low supply elasticity compared to those with high supply elasticity. To visualize these results, Figure 7 shows the ten most supply-constrained MSAs in red and the ten least supply constrained MSAs in green. We see that there is some degree of clustering of the MSAs in accordance with the panel results in Table 7.

In Panel C of Table 7, we include local control variables. In general, access to fundamental variables at the MSA-level is quite limited, especially at the monthly frequency. We include local employment growth ($payrolls$), the local price-rent ratio (pr), and local realized volatility as a measure of uncertainty (unc).[34] We compute the realized volatility measure using a rolling window of three

[34] Data on employment and rent of primary residences are available from the Bureau of Labor Statistics. While employment data are available at the MSA-level, rental data are not available across all MSAs. We therefore use rental data at the regional level (West, North Central, Northeast, and South) and map the MSAs to the region in which the MSA is located.

months based on squared de-meaned returns [35] From the results in Panel C, we see that the inclusion of these control variables leads to only very small changes in slope coefficients on HSI. These results suggest that HSI contains predictive content beyond that contained in typical risk-compensation variables.

In conclusion, our results suggest that variation in local housing demand as proxied by our search index possesses strong predictive power over growth rates in local house prices. Moreover, changes in local housing demand have a larger economic impact on house prices in MSAs with a more constrained supply of housing.

4.2 National-Level versus MSA-Level Search

To analyze the extent to which housing markets are influenced by local search dynamics relative to national search activity, we next augment the panel regression model with the national-level HSI. As Panel A in Table 8 shows, local housing search stays statistically significant across all forecast horizons after controlling for national-level housing search. These results imply that housing markets are strongly influenced by local search dynamics, consistent with findings in the literature that housing markets are local in nature (Del Negro and Otrok, 2007, Gyourko et al., 2013, Glaeser et al., 2014, and Hernández-Murillo et al., 2017).

We also evaluate if the effect of local supply elasticity is affected by variation in the national-level housing search. Panel B of Table 8 shows that the effect of local-level search remains stronger in MSAs with low supply elasticity. In addition, Panel C of Table 8 shows that these results are largely unaffected by including control variables.

4.3 Local Variation in Time-on-Market

To shed further light on the economic channel that generates the predictive ability of HSI, we analyze the relation between HSI and the time it takes for a home to be sold. According to the search-based models of e.g. Stein (1995), Krainer (2001), Novy-Marx (2009), and Diaz and Jerez (2013), there is a negative correlation between price and time-on-market (TOM). Likewise, the search model of Genesove and Han (2012) in which sellers react to demand shocks with a delay

[35] This way of computing realized volatility follows from high-frequency measures of e.g. Andersen et al. (2001).

implies that TOM decreases following a positive shock to demand. Thus, given that HSI proxies for demand, we expect to see a negative relation between HSI and TOM.

To examine if this holds, we estimate panel regressions,

$$TOM_{it+h} = \alpha_i + \beta HSI_{it} + \varepsilon_{it+h}, \tag{9}$$

where TOM_{it} is the average time-on-market measured in days in MSA i in month $t+h$.[36]

Consistent with predictions from theoretical search models, Table 9 shows that there is a negative relation between TOM and HSI. The effect of HSI on TOM as measured by the magnitude of the predictive coefficient and the within-R^2 is strongest at the one and three month horizons. At these horizons, HSI is strongly statistically significant with two-way clustered t-statistics of -3.66 and -3.72, respectively. HSI remains statistically significant at the six month horizon but turns insignificant at the 12-month horizon. These results imply that an increase in search activity leads to a decrease in time-on-market with an effect that peaks at relatively short horizons. This evidence appears in line with the analysis Genesove and Han (2012) who find that short-run effects are stronger than long-run effects.

4.4 Out-of-Sample Forecasts

We next examine out-of-sample forecasts of house prices in each of the 77 MSAs with the first forecast made for 2007:1 and the last for 2021:1. We compute out-of-sample forecasts using both the MSA-level and national-level HSIs as predictive variables and use recursive estimation with an expanding window. Panel B in Figure 6 gives an overview of the out-of-sample R^2 values across MSAs. R^2_{OoS} values are systematically high across MSAs and forecast horizons. At the one-month horizon, the median R^2_{OoS} is 57% while the first and third quartile values are 49% and 62% across MSAs. The extent of predictability reaches its peak at the three-month horizon with a median R^2_{OoS} of 57%, which is only slightly reduced at longer horizons. For the annual horizon, the median R^2_{OoS} is 43%, while the first and third quartile values are 24% and 55%.

Panel C in Figure 6 visualizes the out-of-sample performance at the one-month horizon by plotting

[36] From the Zillow database, we have obtained time-on-market data for 72 out of 77 MSAs in our cross section over the period 2018:1 to 2021:1.

the median value of the cumulative squared forecast error for the no-predictability benchmark and that of the online search model. The out-of-sample performance of online search has been stable over time as HSI succeeds in consistently outperforming the no-predictability benchmark. Of special interest is the period around the outbreak of the Covid-19 pandemic which triggered an initial drop in house prices, quickly followed by rapidly rising house prices. Consistent with the findings for national-level house prices, Panel C illustrates that the MSA-level predictive power of online search actually strengthens during the pandemic.

The predictive power of online search is consistently strong over time, but gets even stronger during periods of turmoil such as the financial crisis in 2007-2009 and the Covid-19 pandemic. Theoretical models of search imply that the effect of shocks to demand is amplified and can lead to excess volatility (e.g. Novy-Marx, 2009 and Anenberg and Bayer, 2020). This mechanism of search models may explain why we find the largest gains in predictability relative to the historical mean benchmark during times of high price volatility.

Because HSI has the potential to act as an early indicator of where the market is going, it is important to examine how HSI performs under different market conditions and price paths. We therefore separately analyze the out-of-sample predictive power of HSI during periods of upturns and downturns in the housing market. Following Burnside et al. (2016), we define $\bar{p}_{it}$ as the centered moving average of the log house price in MSA i at time t:

$$\bar{p}_{it} = \frac{1}{2n+1}\sum\nolimits_{j=-n}^{n} p_{it+j}.$$

An upturn is then given as an interval of time for which $\Delta\bar{p}_{it} > 0$ for all t while a downturn is an interval of time for which $\Delta\bar{p}_{it} < 0$ for all t.[37] In Panel D of Figure 6, we show the median R^2_{OoS} across MSAs during periods of upturns and downturns. While HSI generates strong out-of-sample predictability in both upturns and downturns, R^2_{OoS} increases during periods of downturns. Thus, the predictive power of HSI is robust across different market conditions but is strengthened when the housing market is in a downturn.

We next analyze whether the predictability mainly is concentrated in MSAs with either high or low levels of housing market volatility. We sort the 77 MSAs in two groups based on local standard deviations of house price growth rates. Panel E of Figure 6 shows that out-of-sample predictability

[37] We set n to equal 5 months but obtain similar results with other reasonable choices of n.

is slightly higher in housing markets with high volatility but is generally strong across forecast horizons in both low and high volatility markets. Overall, there is strong and robust evidence that HSI is a useful out-of-sample predictor of house prices across MSAs.

4.5 Economic Significance

Our results show that there is substantial variation in search activity over time and across MSAs and that search activity precedes movements in future house prices – a finding that is in line with the theoretical search-and-matching modelling framework (e.g. Carillo et al., 2015).

To shed further light on the degree of time-variation in expected house price changes, we identify episodes of intense housing search activity based on a threshold of one standard deviation of the local HSI. Across MSAs, we identify 2,320 months with housing search activity more than one standard deviation above the (local) mean. Panel A in Figure 8 shows the median house price change following these episodes as well as the 1st and 3rd quartiles. From the figure, we see that the median growth rate in house prices following periods with high search is 0.7% at the one-month horizon, 1.9% at the three-month horizon, 3.6% at the six-month horizon and 6.7% at the annual horizon. These realized house price changes suggest that significant economic gains can potentially be achieved from being an early buyer in a market with increasing demand.

Panel A also illustrates that the potential savings from buying a house h months early in an increasing market varies strongly across MSAs. When search is one standard deviation above the mean, the 25th percentile growth rate is 0.4%, 1.2%, 2.5%, and 4.8% at the one-month, three-month, six-month, and one-year horizons, while for the 75th percentile, the price changes are 1.2%, 3.4%, 6.1%, and 10.1%, respectively.

Because HSI captures local changes in housing demand, we would also expect episodes with low search activity to coincide with subsequent downward pressure on house prices. We identify periods with low search activity as months in which the local housing search falls one standard deviation below its mean. Across MSAs, we find 2,802 events with low search activity. Panel B shows that episodes with low HSI are associated with a subsequent median decline in house prices of -0.3%, -0.8%, -1.5%, and -2.8% at the one-month, three-month, six-month, and one-year horizons. This suggests that it is risky to buy early during times when HSI is low. That is, there is a strong incentive to suspend or reduce search efforts because it is likely it will become possible to buy a house at a lower price, the longer the buyer waits.

Overall, our results suggest that the potential economic gains from exploiting predictability in house prices as identified by HSI can be quite large. However, frictions can generate trading delays, implying that not all agents seeking to buy a new home can do so right away. Furthermore, we cannot entirely rule out that search intensity comove with a time-varying risk-premium component such that episodes of intense search reflect a high required premium for buying a house at the relevant point in time. However, we offer two pieces of evidence that run counter to the time-varying risk premium interpretation.

First, Table 2 shows that proxies for time-varying risk premia such as the price-dividend ratio and the risk aversion and uncertainty indexes of Bekaert et al. (2021) only explain a very small portion of the variation in HSI. Moreover, Table 6 shows that, in contrast with HSI, these risk premium proxies produce very poor out-of-sample forecasts of changes in house prices.

Second, if variation in HSI reflects a time-varying risk premium, we would expect its predictive power over residential house prices to carry over to the REIT market. In contrast, if the predictive power of HSI arises from search frictions, it should hold little or no predictive power for publicly traded REITs whose prices are largely unaffected by search frictions. In Section A.12 in the Online Appendix, we show that HSI has very limited predictive ability over REIT returns.[38]

These findings suggest that the predictive power of HSI over future house prices does not arise from a risk compensation channel, but is more likely to reflect sluggish price adjustments in the residential real estate market due to frictions.

5 Search and House Prices During Covid-19

The Covid-19 pandemic triggered the sharpest reduction in economic activity recorded in modern times. Despite this massive contraction, in 2020 U.S. house prices experienced their largest gains

[38] In comparing the results for owner-occupied housing and the REIT market, we should however note that REITs income are ultimately driven by the demand for rental space and to some extent, rental properties are a substitute for owner-occupied housing. Consequently, if more households transition into homeownership, there will be less demand for rental properties, and REITs could experience a drop in value. This could potentially mute the degree of predictability by HSI for REIT returns even if there is a time-varying risk premium.

since 2005 (Mahertz, 2021).[39] Recessions are traditionally accompanied by stagnant or declining housing markets, so the increased house prices has led to wide speculation by experts and the media about its possible causes (Demsas, 2021, Friedman, 2021a, Passy, 2021). Some authors suggest that the primary cause of higher prices is supply constraints since the number of homes for sale and new houses built across metropolitan areas in the U.S. plummeted during 2020 (Badget and Bui, 2021, Friedman, 2021b). Others have pointed towards factors affecting demand (Romem, 2020), including falling mortgage rates, increased preference for more space and suburban housing resulting from a shift towards working from home and the increased adoption of new technologies that facilitate and accelerate the buying process due to social distancing norms.

From the perspective of theoretical models of search, a possible interpretation is that the Covid-19 pandemic caused a shock to demand due to a shift in housing preferences, which implied an increase in buyers entering the market and hence an increase in the bargaining power of sellers (e.g. Carillo et al., 2015). Due to sluggish price adjustments arising from frictions, the demand shock influences prices not only on initial impact but also in subsequent periods. In addition, the feedback effect described by Novy-Marx (2009) may have substantially magnified the initial shock, especially due to the tight supply constraints.

A combination of supply and demand shocks thus appear to have affected house prices during the pandemic. From a policy perspective it is important to quantify their respective effects since the optimal policy response depends heavily on their relative importance.[40] To address this point, we estimate the following panel regression

$$p_{it+1} - p_{it} = \alpha_i + \beta_D HSI_{it} + \beta_S S_{it} + \gamma' Z_{it} + \varepsilon_{it+1}, \tag{10}$$

where $p_{it+1} - p_{it}$ is the one-month change in the log house price index for MSA i in month $t+1$, HSI_{it} is our housing search index, S_{it} is a proxy for the housing supply given by the Zillow for-sale-inventory of houses (seasonally adjusted), and Z_{it} is a vector of controls including the Covid-19 stringency index of Hale et al. (2021) and the number of Covid-19 cases, all measured for MSA i in month t.

[39] At 11.3% in January 2021, trailing 12-month returns on the national FHFA index recorded their highest value in our sample. This compares with a pre-Covid-19 maximum of 10.1% for September 2005.

[40] A recent report from OECD (2020) on the housing market during Covid-19 notes that policy responses to curb housing demand can affect the long-run supply.

Table 10 shows results from estimating (10) over the period from February 2020 to January 2021, the end of our sample. While the short time span is a concern, the cross-sectional dimension of our data helps to achieve more powerful tests and reliable estimates.[41] For comparison with the full sample results (see Table 7, Panel A for $h = 1$), we first estimate (10) using only HSI as a proxy for demand (column 1). The slope coefficient during the Covid-19 period is slightly lower than for the full sample (0.33 vs. 0.42), while the R^2 is slightly higher (40.9% vs. 37.2%), demonstrating that HSI remained a robust predictor of growth in house prices during the pandemic. Next, we include supply (column 2) and finally also the lockdown stringency measure of Hale et al. (2021) (column 3) and the number of Covid-19 cases (column 4) as control variables. Across all specifications, HSI and the supply measures are highly statistically significant with the expected signs. For example, in the full specification, the coefficient estimates on HSI and supply measures are 0.13 and -0.16. A one standard deviation change in HSI is thus associated with a 1.6% predicted (annualized) change in house prices, while the corresponding impact from a supply change is 1.9%.

Interestingly, Covid-19 restrictions on their own were associated with large negative changes in house prices: the estimated slope coefficient on the stringency index of Hale et al. (2021) is -0.25 with a highly significant t-statistic. Controlling for Covid-19 restrictions is clearly important as their introduction leads to a reduction in the estimated slope coefficient on HSI from 0.33 to 0.13. These results suggest that tight supply constraints and increased demand for houses combined to lead to higher house prices during the pandemic.[42]

6 Concluding Remarks

In this paper, we show that online data on search for housing can be used to accurately quantify variation in the demand for housing both at the national (U.S.) and regional (metropolitan) level. Moreover, such data can be used to robustly predict changes in house prices, both in-sample and out-of-sample, at short and long-term horizons, and in periods with rapidly or slowly changing house prices.

Our housing search index produces significantly more accurate forecasts of house prices than con-

[41] The cross-sectional dimension consists of 72 MSAs for which we have data on both prices, supply and demand. The control variables in this regression are available only at the state level, so we map the state-level data to the individual MSAs for these variables.

[42] Section A.10 in the Online Appendix provides additional results on the Covid-19 period.

ventional measures of variation in housing demand such as employment, interest rates, sentiment, or proxies for risk. These variables only provide a partial account of housing demand and embed much less information about future house prices than our more direct measure obtained from search activity which reflects peoples' interest in buying a house regardless of whether the motive is based on fundamentals or is of a more speculative nature.

Our findings of strong predictability of future changes in house prices do not suggest arbitrage opportunities and also do not appear to be driven by time-varying risk premia. Instead, our results are more consistent with search-based models with frictions which imply that shocks affecting the housing market will only be reflected in future house prices through a gradual adjustment process.

References

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of American Statistical Assocication* 96, 42-55.

Andrei, D., Hasler, M. (2015). Investor attention and stock market volatility. *Review of Financial Studies* 28, 33-72.

Anenberg, E., Bayer, P. (2020). Endogenous sources of volatility in housing markets: the joint buyer-seller problem. *International Economic Review* 61, 1195-1228.

Ayat, L., Burridge, P. (2000). Unit root tests in the presence of uncertainty about the non-stochastic trend. *Journal of Econometrics* 95, 71-96.

Badget, E., Bui, Q. (2021). Where have all the houses gone? *The New York Times*, May 14.

Bai, J., Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146, 304-317.

Bekaert, G., Engstrom, E.C., Xu, N.R. (2021). The time variation in risk appetite and uncertainty. *Management Science*, forthcoming.

Beracha, E., Wintoki, M. (2013). Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research* 35, 283-312.

Berkovec, J.A., Goodman, J.L. (1996). Turnover as a measure of demand for existing homes. *Real Estate Economics* 24, 421-440.

Bijl, L., Kringhaug, G., Molnár, P., Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis* 45, 150-156.

Bork, L., Møller, S.V. (2018). Housing price forecastability: A factor analysis. *Real Estate Economics* 46, 582-611.

Bork, L., Møller, S.V., Pedersen, T.Q. (2020). A new index of housing sentiment. *Management Science* 66, 1563-1583.

Borup, D., Schütte, E. (2022). In search of a job: Forecasting employment growth using Google trends. *Journal of Business & Economic Statistics* 40, 186-200.

Burnside, C., Eichenbaum, M., Rebelo, S. (2016). Understanding booms and busts in housing markets. *Journal of Political Economy* 124, 1088-1147.

Campbell, J., Thompson, S. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 1509-1531.

Carrillo, P. E., de Wit, E. R., Larson, W. (2015). Can tightness in the housing market help predict subsequent home price appreciation? Evidence from the United States and the Netherlands. *Real Estate Economics* 43, 609-651.

Case, K. E., Shiller, R. J. (1989). The efficiency of the market for single family homes. *American Economic Review* 79, 125-137.

Case, K.E., Quigley, J.M., Shiller, R.J. (2012). Wealth effects revisited 1975-2012. Cowles Foundation Discussion Paper no. 1884. Yale University.

Choi, H., Varian, H. (2012). Predicting the present with Google Trends. *Economic Record* 88, 2-9.

Chong, Y., Hendry, D. (1986). Econometric evaluation of linear macroeconomic models. *Review of Economic Studies* 53, 671-690.

Cox, J., Ludvigson, S.C. (2019). Drivers of the great housing boom-bust: Credit conditions, beliefs, or both? *Real Estate Economics* 2019, 1-33.

Da, Z., Engelberg, J., Gao, P. (2011). In search of attention. *Journal of Finance* 66, 1461-1499.

Da, Z., Engelberg, J., Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. *Review of Financial Studies* 28, 1-32.

D'Amuri, F., Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting* 33, 801-816.

Del Negro, M., Otrok, C. (2007). 99 luftballons: Monetary policy and the house price boom across US states. *Journal of Monetary Economics* 54, 1962-1985.

Demsas, J. (2021). Covid-19 caused a recession. So why did the housing market boom? *Vox*, Feb 5.

Diaz, A., Jerez, B. (2013). House prices, sales, and time on the market: A search-theoretic framework. *International Economic Review* 54, 837-872.

van Dijk, D.W., Francke, M.K. (2018). Internet search behavior, liquidity and prices in the housing market. *Real Estate Economics* 46, 368-403.

Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253-263.

Dietzel, M. (2016). Sentiment-based predictions of housing market turning points with google trends. *International Journal of Housing Markets* 9, 108-136.

Friedman, N. (2021a). The pandemic ignited a housing boom – but it is different from the last one. *The Wall Street Journal*, March 15.

Friedman, N. (2021b). U.S. existing-home prices climb on tight supply. *The Wall Street Journal*, April 23.

Gargano, A., Giacoletti, M., Jarnecic, E. (2021). Local experiences, attention and spillovers in the housing market. *Journal of Finance*, forthcoming.

Genesove, D., Han, L. (2012). Search and matching in the housing market. *Journal of Urban Economics* 72, 31-45.

Ghysels, E., Horan, C., Moench, E. (2017). Forecasting through the rearview mirror: Data revisions and bond return predictability. *Review of Financial Studies* 31, 678-714.

Ghysels, E., Plazzi, A., Valkanov, R., Torous, W. (2013). Forecasting real estate prices. *Handbook of Economic Forecasting*, G. Elliot and A. Timmermann (Eds.), 509-580.

Glaeser, E.L., Gyourko, J., Morales, E., Nathanson, C.G. (2014). Housing dynamics: An urban approach. *Journal of Urban Economics* 81, 45-56.

Guo, H. (2009). Data revisions and out-of-sample stock return predictability. *Economic Inquiry* 47, 81-97.

Gyourko, J., Mayer, C., Sinai, T. (2013). Superstar cities. *American Economic Journal: Economic Policy* 5, 167-199.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B. Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour* 5, 529-538.

Han, L., Strange, W. (2015). The microstructure of housing markets: search, bargaining, and brokerage. *Handbook of Regional and Urban Economics*, Edited by Duranton, G., Henderson J.V., Strange, W.C., 5, 813-886.

Head, A., Lloyd-Ellis, H., Sun, H. (2014). Search, liquidity, and the dynamics of house prices and construction. *American Economic Review* 104, 1172-1210.

Hernández-Murillo, R., Owyang, M.T., Rubio, M. (2017). Clustered housing cycles. *Regional Science and Urban Economics* 66, 185-197.

Hodrick, R. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies* 5, 357-386.

Huang, D., Jiang, D., Tu, J., Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies* 28, 791-837.

Joseph, K., Wintoki, M.B., Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting* 27, 1116-1127.

Krainer, J. (2001). A theory of liquidity in residential real estate markets. *Journal of Urban Economics* 49, 32-53.

Mahertz, T. (2021). The U.S. housing market gained more value in 2020 than in any year since 2005. *Zillow*, Jan 26.

National Association of Realtors (NAR) (2020). Home Buyers and Sellers Generational Trends Report. Technical Report, National Association of Realtors.

Newey, W., West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.

Novy-Marx, R. (2009). Hot and cold markets. *Real Estate Economics* 37, 1-22.

OECD (2020). Housing amid covid-19: Policy responses and challenges. *OECD contributions*.

Passy, J. (2021). Is the U.S. housing market heading for a crash? Here is what the experts say. *MarketWatch*, May 3.

Piazzesi, M., Schneider, M., Stroebel, J. (2020). Segmented housing search. *American Economic Review* 110, 720-759.

Plazzi, A., Torous, W., Valkanov, R. (2010). Expected returns and expected growth in rents of commercial real estate. *Review of Financial Studies* 23, 3469-3519.

Politis, D.N., White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23, 53-70.

Rapach, D., Strauss, J. (2009). Differences in housing price forecastability across US states. *International Journal of Forecasting* 25, 351-372.

Reinhardt, C.M., Rogoff, K.S. (2009). *This Time is Different: Eight Countries of Financial Folly.* Princeton University Press.

Romem, I. (2020). Explaining the frenzy in the housing market. *The New York Times*, Dec, 3.

Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Jounal of Economics* 125, 1253-1296.

Soo, C. K. (2018). Quantifying animal spirits: News media and sentiment in the housing market. *Review of Financial Studies* 31, 3689-3719.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics* 54, 375-421.

Statista (2018). Market share of search engines in the United States 2008-2018, available from `www.statista.com`.

Stein, J.C. (1995). Prices and trading volume in the housing market: A model with down-payment effects. *Quarterly Journal of Economics* 110, 379-406.

Thompson, S.B. (2011). Simple formulas for standard errors that cluster by both firrm and time. *Journal of Financial Economics* 99, 1-10.

Vozlyublennaia, N. (2014). Investor attention, index performance and return predictability. *Journal of Banking and Finance* 41, 17-35.

Welch, I., Goyal, A. (2008). A comprehensive look at the performance of equity premium prediction. *Review of Financial Studies* 21, 1455-1508.

Wheaton, W.C. (1990). Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy* 98, 1270-1292.

Wu, L., Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In: A. Goldfarb, S. M. Greenstein, and C. E. Tucker (Eds.), *Economic analysis of the digital economy*, Chicago: University of Chicago Press.

Zou, H., Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of Royal Statistical Society*, Series B, 67, 301-320.

Table 1. Granger Causality Tests. The table reports results from standard Granger causality tests:

$$HSI_t = \delta + \sum_{i=1}^{p} \theta_i HSI_{t-i} + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t$$

$$y_t = \delta + \sum_{i=1}^{p} \theta_i y_{t-i} + \sum_{i=1}^{p} \gamma_i HSI_{t-i} + \varepsilon_t$$

where HSI is the housing search index and y is either house price changes, home sales, or Redfin's housing demand index (HDI). $HSI^{\perp}$ is the part of HSI that is orthogonal to the other predictive variables. The table shows p-values from the joint test that $\gamma_1 = \gamma_2 = ... = \gamma_p = 0$. We use the Newey and West (1987) estimator with automatic lag selection. For house prices and home sales, the sample period is 2004:1 to 2021:1 with monthly observations, while the sample frequency is weekly for HDI and the sample runs from the first week of 2018 until the first week of April 2021.

Null hypothesis	$p = 1$	$p = 2$	$p = 3$	$p = 4$
HSI does not Granger cause house price changes	0.000	0.000	0.000	0.000
House price changes do not Granger cause HSI	0.022	0.209	0.719	0.669
HSI does not Granger cause home sales	0.002	0.005	0.003	0.004
Home sales do not Granger cause HSI	0.916	0.411	0.226	0.204
HSI does not Granger cause HDI	0.000	0.000	0.000	0.000
HDI does not Granger cause HSI	0.616	0.667	0.661	0.144
$HSI^{\perp}$ does not Granger cause house price changes	0.002	0.016	0.052	0.066
House price changes do not Granger cause $HSI^{\perp}$	0.267	0.485	0.285	0.449

Table 2. The Relation Between Housing Search and Alternative Variables. The table reports results from regressions, $HSI_t = \alpha + x_t'\beta + \varepsilon_t$, where HSI_t is the housing search index and x_t contains standard house price determinants. We show results from univariate regressions using one variable at a time as well as from a multivariate regression. For each regression, the table reports estimates of β, corresponding t-statistics in parenthesis, and the R^2 in square brackets. We compute standard errors using the Newey and West (1987) estimator with automatic lag selection. All variables are standardized to facilitate comparison of the β estimates. The sample period is 2004:1 to 2021:1.

	Univariate	Multivariate
payrolls	−0.22	−0.02
	(−1.17)	(−0.28)
	[4.86]	
infl	0.03	0.01
	(0.26)	(0.22)
	[0.10]	
permits	0.30	0.05
	(3.24)	(1.34)
	[9.13]	
starts	0.16	−0.01
	(2.40)	(−0.47)
	[2.63]	
term	−0.04	0.51
	(−0.29)	(5.43)
	[0.19]	
mort	−0.57	−1.18
	(−4.57)	(−12.57)
	[32.05]	
pr	−0.00	1.09
	(−0.01)	(6.93)
	[0.00]	
loans	−0.10	0.15
	(−0.66)	(2.53)
	[1.10]	
sent	0.17	0.10
	(1.20)	(1.25)
	[2.92]	
cfnai	0.18	0.06
	(1.01)	(1.09)
	[3.27]	
ads	0.17	0.12
	(1.06)	(2.53)
	[2.89]	
pd	0.28	0.04
	(2.66)	(0.26)
	[7.87]	
ra	−0.20	−0.11
	(−2.21)	(−1.43)
	[4.20]	
unc	−0.20	0.10
	(−1.73)	(0.61)
	[3.89]	[70.99]

Table 3. Predicting House Prices With Housing Search and Alternative Variables. The table reports results from predictive regressions, $p_{t+h} - p_t = \alpha_h + \beta_h HSI_t + \theta_h' Z_t + \varepsilon_{t+h}$, where p is the log of the FHFA house price index, HSI is the housing search index, Z is a vector of control variables, and h is the forecasting horizon in months. Panel A reports results using HSI on its own (i.e. $\theta_h = 0$). Panel B controls for the 14 predictive variables defined in Section 2.6. Panel C uses the part of HSI that is orthogonal to the other predictive variables ($HSI^{\perp}$). Panel D controls for an AR(1) component. For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the R^2 in square brackets. We compute standard errors using a circular block bootstrap. All predictive variables are standardized and slope coefficients are multiplied by 100 to facilitate comparison across variables. The sample period is 2004:1 to 2021:1.

	$h=1$	$h=3$	$h=6$	$h=12$
Panel A: HSI alone				
HSI	0.43	1.26	2.44	4.37
	(6.39)	(6.74)	(6.19)	(5.35)
	[56.91]	[67.37]	[70.41]	[64.42]
Panel B: HSI and control variables				
HSI	0.27	0.77	1.56	2.31
	(4.66)	(3.84)	(3.36)	(3.27)
	[69.47]	[78.73]	[80.45]	[80.77]
Panel C: Orthogonalized HSI				
$HSI^{\perp}$	0.15	0.47	0.96	1.55
	(3.02)	(3.01)	(2.90)	(2.89)
	[7.12]	[9.39]	[10.93]	[8.15]
Panel D: Controlling for AR(1) component				
HSI	0.27	0.84	1.69	3.01
	(5.74)	(6.21)	(6.41)	(5.18)
	[64.35]	[74.71]	[77.88]	[70.65]

Table 4. Predicting House Prices With Alternative Search Indices. The table reports results from predictive regressions, $p_{t+h} - p_t = \alpha + \beta' x_t + \varepsilon_{t+h}$, where p_t is the log of the FHFA house price index, x_t is a vector of predictive variables, and h is the forecast horizon in months. For each regression, the table reports slope estimates, the corresponding t-statistics in parenthesis, and the R^2 in square brackets. We compute standard errors using a circular block bootstrap. All predictive variables are standardized to facilitate comparison of the β estimates and the log price change is multiplied by 100. Panel A shows results for the predefined search categories used by Wu and Brynjolfsson (2015), while Panel B includes HSI jointly with the predefined search categories. The sample period is 2004:1 to 2021:1.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: Predefined search categories				
Real estate agencies	0.25	0.79	1.75	3.60
	(1.86)	(2.11)	(2.71)	(3.23)
Real estate listings	−0.14	−0.44	−1.07	−2.35
	(−1.31)	(−1.34)	(−1.62)	(−1.88)
	[8.97]	[11.74]	[15.37]	[18.21]
Panel B: HSI joint with predefined search categories				
HSI	0.41	1.21	2.33	4.15
	(6.40)	(6.43)	(5.77)	(4.60)
Real estate agencies	0.10	0.24	0.46	0.92
	(1.61)	(1.43)	(1.46)	(1.16)
Real estate listings	−0.02	−0.06	−0.25	−0.87
	(−0.33)	(−0.41)	(−0.98)	(−1.96)
	[59.16]	[69.02]	[71.38]	[65.49]

Table 5. Alternative House Price Indices. The table reports results from predictive regressions, $p_{t+h} - p_t = \alpha + \beta HSI_t + \varepsilon_{t+h}$, where p_t is the log house price measured using either the FHFA index (Panel A), the Case-Shiller index (Panel B), the Freddie-Mac index (Panel C), the Zillow index (Panel D), or the CoStar commercial property index (Panel E). For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the R^2 in square brackets. We compute standard errors using a circular block bootstrap. The sample period is 2004:1 to 2021:1.

$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: FHFA			
0.43	1.26	2.44	4.37
(6.39)	(6.74)	(6.19)	(5.35)
[56.91]	[67.37]	[70.41]	[64.42]
Panel B: Case-Shiller			
0.46	1.40	2.73	5.12
(4.75)	(5.10)	(5.16)	(4.86)
[53.71]	[59.36]	[62.18]	[62.90]
Panel C: Freddie-Mac			
0.49	1.49	2.88	5.26
(4.89)	(5.29)	(5.26)	(4.92)
[65.38]	[68.34]	[67.92]	[63.74]
Panel D: Zillow			
0.39	1.22	2.51	5.02
(4.39)	(4.74)	(5.20)	(5.71)
[55.96]	[60.97]	[66.38]	[71.19]
Panel E: CoStar			
0.33	1.13	2.58	5.70
(1.90)	(2.18)	(2.31)	(2.16)
[5.83]	[9.95]	[18.68]	[31.06]

Table 6. Out-of-Sample Tests. Panel A reports the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OoS}) and in parenthesis the p-value from the Diebold and Mariano (1995) t-statistic, computed using the Newey and West (1987) estimator with h lags, where h is the forecast horizon in months. The null hypothesis is that the R^2_{OoS} is equal to zero or negative and the alternative hypothesis is that it is positive. Panel B reports coefficient estimates from forecast encompassing tests for whether the weights in (5) are equal to zero with p-values shown in parentheses.

	$h=1$	$h=3$	$h=6$	$h=12$
Panel A: R^2_{OoS} statistics				
HSI	51.37	64.54	63.91	54.25
	(0.00)	(0.00)	(0.00)	(0.02)
payrolls	−44.66	−94.68	−124.68	−61.46
	(0.93)	(0.96)	(0.95)	(0.91)
infl	−4.22	−1.52	−1.15	−0.81
	(0.87)	(0.77)	(0.77)	(0.86)
permits	2.91	2.58	0.29	2.36
	(0.39)	(0.40)	(0.48)	(0.33)
starts	1.02	1.47	1.65	1.14
	(0.33)	(0.21)	(0.16)	(0.22)
term	−5.84	−10.44	−14.92	−21.92
	(0.80)	(0.81)	(0.78)	(0.74)
mort	2.35	−5.20	−8.93	−33.21
	(0.43)	(0.59)	(0.64)	(0.78)
pr	−12.97	−26.63	−42.68	−70.34
	(0.93)	(0.98)	(0.96)	(0.92)
loans	−7.53	−13.91	−23.66	−49.02
	(0.79)	(0.90)	(0.88)	(0.87)
sent	−0.49	−2.47	−0.17	13.79
	(0.53)	(0.59)	(0.50)	(0.26)
$cfnai$	−65.29	−109.85	−111.84	−7.45
	(0.84)	(0.86)	(0.88)	(0.67)
ads	−91.49	−141.42	−139.03	−4.91
	(0.86)	(0.88)	(0.89)	(0.65)
pd	−32.02	−96.27	−179.35	−288.30
	(0.77)	(0.86)	(0.87)	(0.87)
ra	−189.95	−392.86	−511.52	−274.89
	(0.88)	(0.90)	(0.93)	(0.93)
unc	−23.38	−69.09	−120.67	−253.75
	(0.76)	(0.83)	(0.85)	(0.85)
$ar1$	44.13	55.54	54.95	31.15
	(0.00)	(0.00)	(0.00)	(0.14)
Panel B: Encompassing tests				
HSI	0.59	0.60	0.60	0.79
	(0.00)	(0.00)	(0.00)	(0.00)
$ar1$	0.41	0.40	0.40	0.21
	(0.00)	(0.00)	(0.00)	(0.25)

Table 7. Predicting Local House Prices With Local Housing Search: Evidence From Panel Regressions. The table reports results from fixed effects panel regressions of the form, $p_{it+h} - p_{it} = \alpha_i + \beta HSI_{it} + \beta_E HSI_{it} \times Elasticity_i + \phi' Z_{it} + \varepsilon_{it+h}$, where p_{it} is the log of the Freddie-Mac house price index in MSA i, HSI_{it} is the housing search index in MSA i, $Elasticity_i$ is supply elasticity in MSA i, Z_{it} contains control variables, and h is the forecast horizon in months. The control variables are local employment growth, the local price-rent ratio, and local realized volatility. In Panels A and B, we set $\beta_E = \phi = 0$ and $\phi = 0$, respectively. For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the within R^2 in square brackets. We compute standard errors using Thompson (2011) two-way clustered robust-statistics with h lags. HSI is standardized to facilitate interpretation of the β estimates. The sample period is 2004:1 to 2021:1.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: Local HSI				
HSI	0.42	1.22	2.32	4.19
	(9.43)	(7.59)	(6.26)	(5.58)
	[37.20]	[37.89]	[36.89]	[33.81]
Panel B: Effect of Supply Elasticity				
HSI	0.65	1.92	3.67	6.58
	(7.49)	(6.32)	(5.23)	(4.49)
$HSI \times Elasticity$	-0.12	-0.35	-0.67	-1.20
	(-4.02)	(-3.65)	(-3.22)	(-2.79)
	[40.06]	[40.93]	[39.99]	[36.57]
Panel C: Adding Control Variables				
HSI	0.61	1.81	3.41	5.77
	(7.98)	(6.94)	(5.70)	(4.19)
$HSI \times Elasticity$	-0.11	-0.34	-0.67	-1.18
	(-4.03)	(-3.73)	(-3.30)	(-2.75)
	[43.40]	[44.05]	[44.15]	[45.50]

Table 8. Predicting Local House Prices With National and Local Housing Search: Evidence From Panel Regressions. The table reports results from fixed effects panel regressions of the form, $p_{it+h} - p_{it} = \alpha_i + \beta_{US} HSI_{USt} + \beta HSI_{it} + \beta_E HSI_{it} \times Elasticity_i + \phi' Z_{it} + \varepsilon_{it+h}$, where p_{it} is the log of the Freddie-Mac house price index in MSA i, HSI_{USt} is the national-level housing search index, HSI_{it} is the housing search index in MSA i, $Elasticity_i$ is supply elasticity in MSA i, Z_{it} contains control variables, and h is the forecast horizon in months. The control variables are local employment growth, the local price-rent ratio, and local realized volatility. In Panels A and B, we set $\beta_E = \phi = 0$ and $\phi = 0$, respectively. For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the within R^2 in square brackets. We compute standard errors using Thompson (2011) two-way clustered robust-statistics with h lags. HSI is standardized to facilitate interpretation of the β estimates. The sample period is 2004:1 to 2021:1.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: U.S. vs. Local HSI				
U.S. HSI	0.28	0.86	1.66	3.01
	(10.35)	(9.78)	(8.23)	(5.56)
Local HSI	0.25	0.74	1.41	2.58
	(7.95)	(6.90)	(6.30)	(6.86)
	[48.52]	[50.36]	[49.90]	[46.27]
Panel B: Effect of Supply Elasticity				
U.S. HSI	0.28	0.85	1.64	2.97
	(10.16)	(9.64)	(8.20)	(5.57)
Local HSI	0.47	1.40	2.70	4.85
	(6.28)	(5.41)	(4.80)	(4.62)
Local $HSI \times Elasticity$	−0.11	−0.33	−0.64	−1.12
	(−4.07)	(−3.60)	(−3.18)	(−2.78)
	[51.10]	[53.11]	[52.69]	[48.71]
Panel C: Adding Control Variables				
U.S. HSI	0.30	0.88	1.63	2.61
	(10.51)	(9.27)	(7.89)	(5.24)
Local HSI	0.43	1.32	2.59	4.62
	(6.52)	(5.61)	(4.87)	(4.04)
Local $HSI \times Elasticity$	−0.11	−0.32	−0.63	−1.12
	(−4.27)	(−3.75)	(−3.23)	(−2.71)
	[54.27]	[54.80]	[53.93]	[52.82]

Table 9. Predicting Time-on-Market With Local Housing Search. The table reports results from fixed effects panel regressions of the form, $TOM_{it+h} = \alpha_i + \beta HSI_{it} + \varepsilon_{it+h}$, where TOM_{it} is the time-on-market measured in days in MSA i, HSI_{it} is the housing search index in MSA i, and h is the forecast horizon in months. For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the within R^2 in square brackets. We compute standard errors using Thompson (2011) two-way clustered robust-statistics with one lag. HSI is standardized to facilitate interpretation of the β estimates. The sample period is 2018:1 to 2021:1.

	$h=1$	$h=3$	$h=6$	$h=12$
HSI	-3.49	-3.57	-2.82	-0.72
	(-3.66)	(-3.72)	(-2.70)	(-0.82)
	[19.96]	[21.05]	[13.36]	[0.84]

Table 10. The Housing Market During the Covid-19 Pandemic. The table reports results from fixed effects panel regressions of the form, $p_{it+1} - p_{it} = \alpha_i + \beta_D HSI_{it} + \beta_S S_{it} + \gamma' Z_{it} + \varepsilon_{it+1}$, where p_{it} is the log of the Freddie-Mac house price index in MSA i in month t, HSI_{it} is the housing search index, S_{it} is housing supply as measured by the for-sale-inventory, and Z_{it} is a vector of control variables, including the monthly change in Covid-19 restrictions and Covid-19 cases. For each regression, the table reports slope estimates, the corresponding t-statistic in parenthesis, and the within R^2 in square brackets. We compute standard errors using Thompson (2011) two-way clustered robust-statistics with one lag. To facilitate interpretation of the estimates, we standardize all regressors. The sample period is 2020:2 to 2021:1.

HSI	0.33	0.23	0.13	0.13
	(9.38)	(5.99)	(4.35)	(4.31)
Supply		−0.16	−0.11	−0.16
		(−2.05)	(−1.92)	(−2.22)
Covid-19 restrictions			−0.25	−0.24
			(−6.94)	(−7.14)
Covid-19 cases				−0.07
				(−1.71)
R^2_{within}	[40.92]	[46.79]	[62.43]	[63.24]

Figure 1. Housing Search Index. Panel A shows the housing search index (HSI) along with the log growth rate in the seasonally adjusted Federal Housing Finance Agency (FHFA) purchase-only house price index. Panel B shows HSI along with the monthly sales of existing single-family housing units from the National Association of Realtors. Panel C shows search volume for the predefined search categories "Real estate agencies" and "Real estate listings" along with the log growth rate in the FHFA house price index. The sample period is 2004:1 to 2021:1.

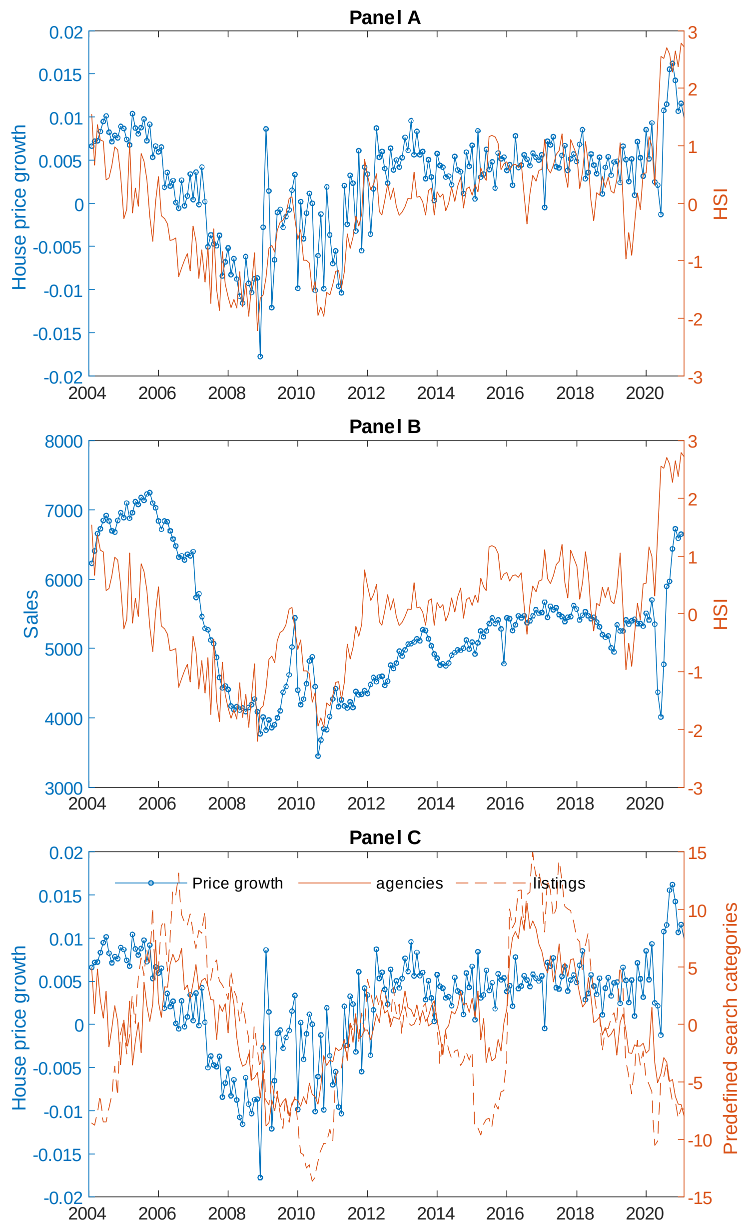

Figure 2. Lead-Lag Relations. Panel A shows regression slope coefficients, associated t-statistics and R^2 values of from regressing monthly price changes from $t-1$ to t on HSI_{t+j} for $j \in \{-12, 12\}$. Panel B shows the results from regressing monthly house sales at time t on HSI_{t+j} for $j \in \{-12, 12\}$. Standard errors are calculated using the Newey and West (1987) procedure with 12 lags.

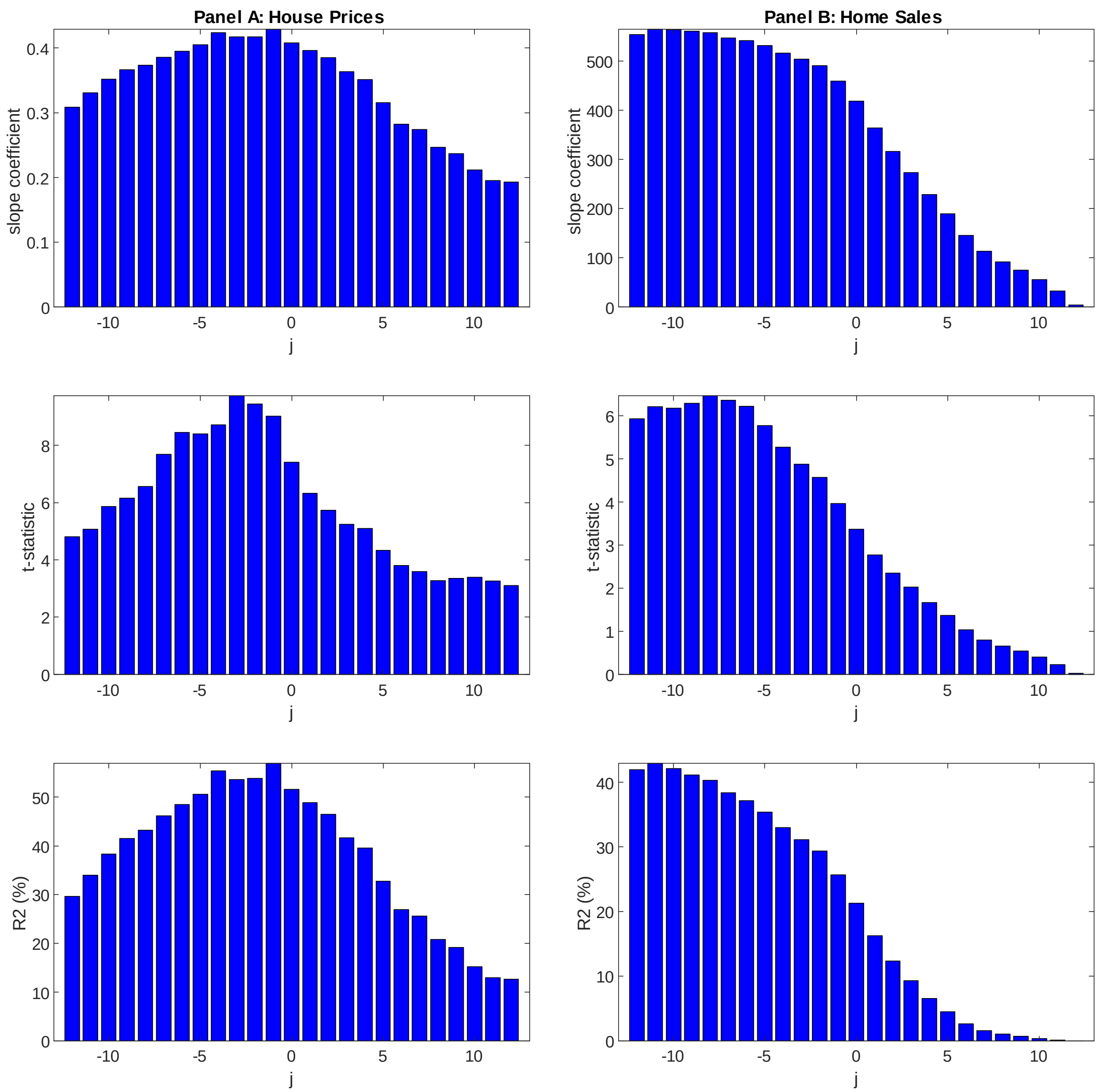

Figure 3. Housing Demand. The figure shows the Housing Search Index (HSI) along with the Housing Demand Index (HDI) contructed by Redfin. In Panel A, the sample frequency is monthly and covers the period 2018:1 to 2021:1. In Panel B, the sample frequency is weekly and the sample period runs from the first week of 2018 until the first week of April 2021.

Figure 4. Local Housing Search. Panel A and B show the local HSI and log growth rate in house prices in Miami (FL) and Wichita (KS), respectively. The sample period is 2004:1-2021:1.

Figure 5. Long-Horizon Predictability. The figure shows estimated slope coefficients, associated t-statistics and R^2 values from the regression, $p_{t+h} - p_t = \alpha + \beta HSI_t + \varepsilon_{t+h}$, as a function of h. We compute standard errors using a circular block bootstrap. The sample period is 2004:1-2021:1.

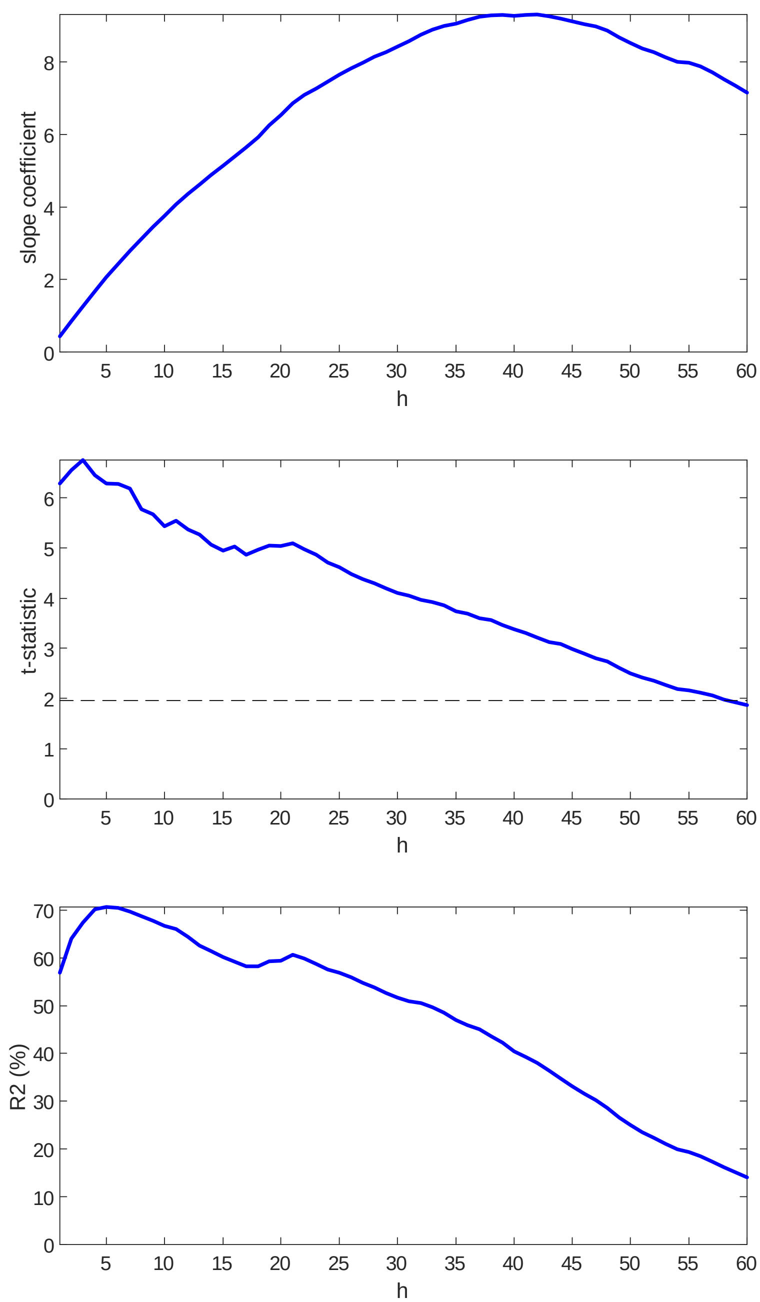

Figure 6. Out-of-Sample Forecast Errors. Panel A shows the cumulative sum of squared forecast errors of the no-predictability benchmark minus the cumulative sum of squared forecast errors of a model based on the national-level HSI. The forecast horizon is $h = 1$ month and the out-of-sample period runs from 2007:1 to 2021:1. Panel B shows the 1st quartile, median and 3rd quartile out-of-sample R^2 values at horizons of $h = 1, 3, 6,$ and 12 months across MSAs. Panel C shows the median of the cumulative sum of squared forecast errors of the no-predictability benchmark minus the cumulative sum of squared forecast errors of HSI across MSAs. Panel D plots the median out-of-sample R^2 at horizons of $h = 1, 3, 6,$ and 12 months across MSAs in downturns and upturns, whereas Panel E shows results for MSAs with high and low volatility.

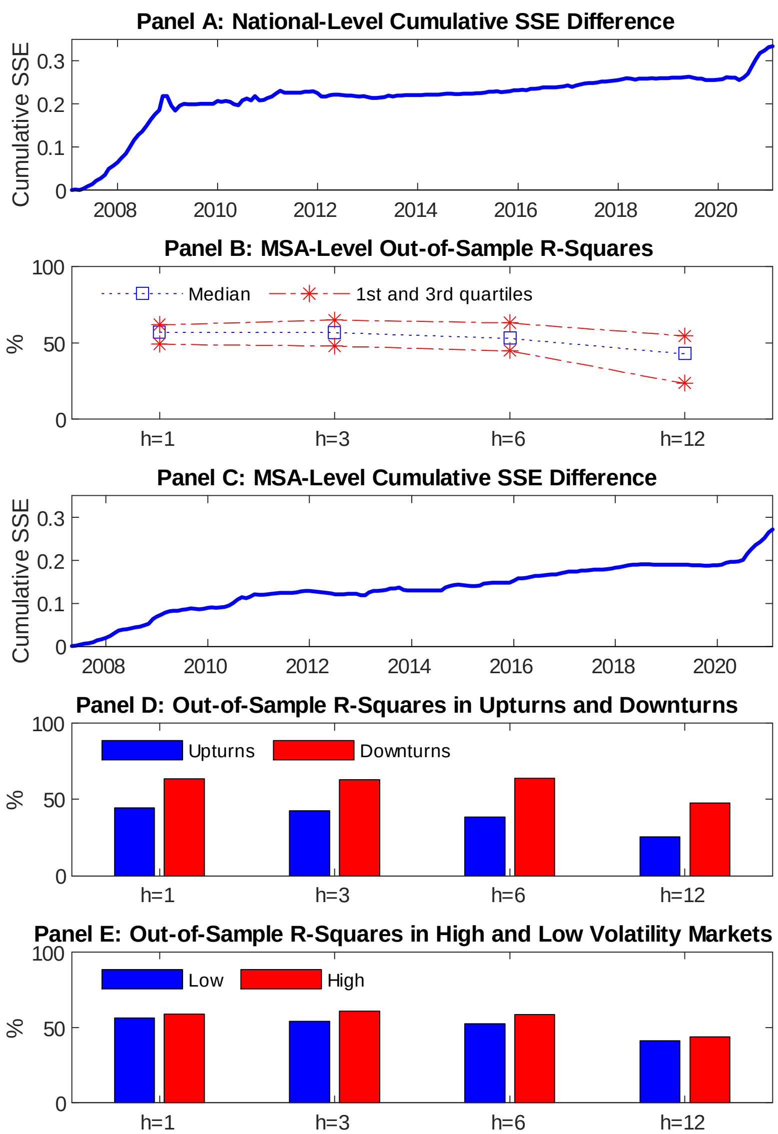

Figure 7. Local Predictive Power of $\boldsymbol{HSI}$. We run MSA-by-MSA forecasting regressions, $p_{it+1} - p_{it} = \alpha_i + \beta_i HSI_{it} + \varepsilon_{it+1}$, and plot the estimate of the predictive coefficient β_i against the R_i^2 across MSAs. HSI is standardized and all β_i estimates are annualized. The ten most supply-constrained MSAs, cf. Saiz (2010), are shown in red and the ten least supply-constrained MSAs are shown in green.

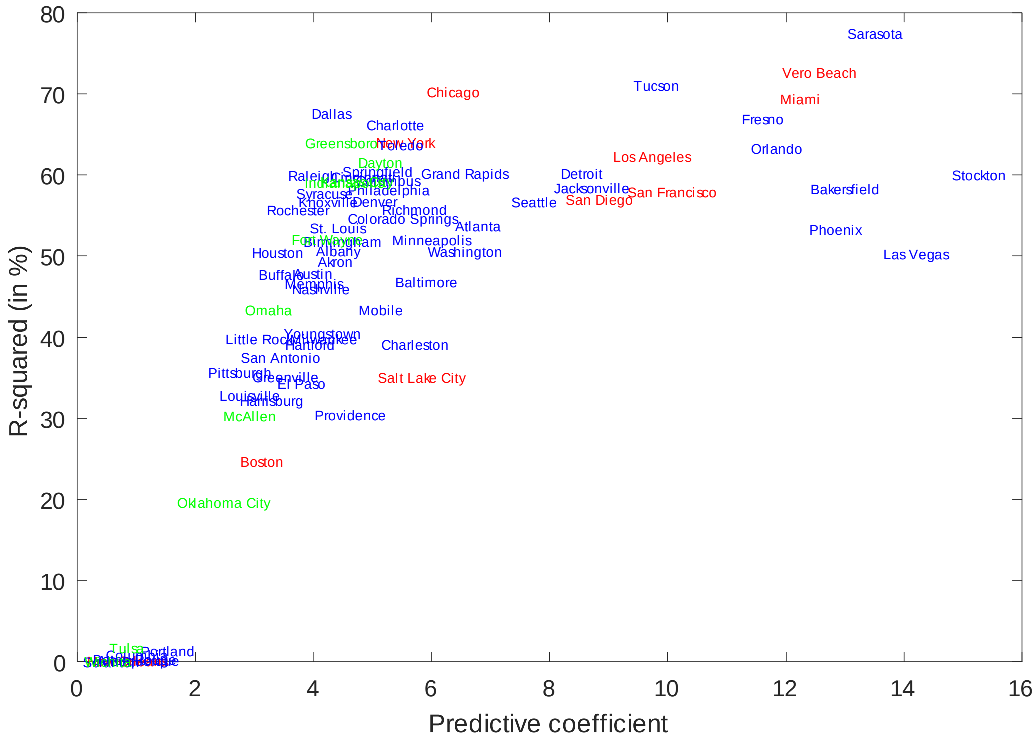

Figure 8. Economic Value of house price forecasts. Panel A shows 1st quartile, median and 3rd quartile growth rates in house prices after episodes where HSI is one standard deviation above the local mean. Panel B shows the results following events where HSI is one standard deviation below the local mean. Forecast horizons range from one-month ahead ($h = 1$) to one-year ahead ($h = 12$).