DISCUSSION PAPER SERIES

DP15875

Search and Predictability of Prices in the Housing Market

Stig Møller, Thomas Pedersen, Erik Christian Montes Schütte and Allan Timmermann

FINANCIAL ECONOMICS

CEPR

ISSN 0265-8003

Search and Predictability of Prices in the Housing Market

Stig Møller, Thomas Pedersen, Erik Christian Montes Schütte and Allan Timmermann

Discussion Paper DP15875
Published 03 March 2021
Submitted 02 March 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Search and Predictability of Prices in the Housing Market

Abstract

We develop a new housing seach 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 for subsequent changes in house prices, both in-sample and out-of-sample, and after controlling for the effect of commonly used predictors. Compared to the stock market, online search has much stronger predictive power over house prices and its effect also lasts longer. Variation in housing search is a particularly strong predictor of subsequent price changes in markets with inelastic housing supply and high speculation.

JEL Classification: C10, E17, G10, R3

Keywords: Internet search, housing markets, Housing Demand, Forecasting, inelastic housing supply

Stig Møller - svm@econ.au.dk
CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus and the Danish Finance Institute

Thomas Pedersen - tqpedersen@econ.au.dk
CREATES, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus and the Danish Finance Institute

Erik Christian Montes Schütte - christianms@econ.au.dk
Aarhus University, CREATES, Danish Finance Institute

Allan Timmermann - atimmermann@ucsd.edu
UCSD and CEPR

Acknowledgements
We thank Torben G. Andersen, 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 acknowledge support from the Danish Finance Institute and from the Danish Council of Independent Research and from the Independent Research Fund Denmark (7024-00020B).

Search and Predictability of Prices in the Housing Market*

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

February 2021

Abstract

We develop a new housing seach 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 for subsequent changes in house prices, both in-sample and out-of-sample, and after controlling for the effect of commonly used predictors. Compared to the stock market, online search has much stronger predictive power over house prices and its effect also lasts longer. Variation in housing search is a particularly strong predictor of subsequent price changes in markets with inelastic housing supply and high speculation.

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

*We thank Torben G. Andersen, 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 acknowledge support from the Danish Finance Institute and from the Danish Council of Independent Research and from the Independent Research Fund Denmark (7024-00020B).

†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. E-mail: christianms@econ.au.dk.

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

1 Introduction

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 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, 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-term horizons, but also that its predictive power declines at longer horizons where the supply of homes is more likely to shift, thus reducing the benefits from search. 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 as well as in markets with high degrees of speculation. 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. Our fifth and final hypothesis is that housing search intensifies when buyers expect home prices to appreciate and, conversely, is reduced when home prices are expected to depreciate.

The intense and lengthy search process involved in buying a house coupled with the 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 intuitively 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, which provides us with a simple and clean measure of housing demand.

We show that demand for housing as measured through online search activity predicts future house prices at both short and long-term horizons. At the one-month horizon, the HSI explains more than 50% of the variation in national house price growth, while at the one-year horizon the explanatory power exceeds 65%. The predictive power of HSI peaks at horizons around 5 to 10 months, which is consistent with the time buyers typically spend finding a home from the initial search process to closing the deal. Across horizons, the 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.

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 the search index and a range of variables typically used to explain dynamics in the housing market. We find that internet search for housing is positively related to employment and buyer sentiment as measured by University of Michigan's Survey of Consumers. In contrast, housing search appears unrelated to interest rates and credit conditions. Based on typically used housing market determinants, we are only able to explain one-third of the variation in the HSI, which suggests that other (unobserved) factors play an important role in households' decision to search for housing.

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. For example, at the annual horizon, the impact on house price growth from a one standard deviation change in housing search is on average 5.4% in low supply-elasticity MSAs versus 3.3% in MSAs with a high supply elasticity. Second, since speculation represents a source of housing demand (Gao et al., 2020), HSI should contain more predictive ability in markets where there are more housing speculation. Using the fraction of non-owner-occupied home purchases as proxy for housing speculation as in Gao et al. (2020), we find that HSI has significantly more predictive ability in MSAs that are more prone to housing speculation. This finding relates to Piazzesi and Schneider (2009), who show that a small number of optimistic investors can have a large price impact.

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 for house prices. This contrasts with our findings, but the reason for the difference is easy to comprehend. Importantly, Wu and Brynjolfsson use two broad and 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 bears a much closer relation to housing demand and consequently has strong and highly significant predictive power over variation in house prices across several horizons. In another related paper, Beracha and Wintoki (2013) use search volume for "real estate i", where "i" is the name of a city and 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 more predictive content for future house prices compared to the procedure used by Beracha and Wintoki (2013).

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 show that the index predicts short-term return reversals as well as temporary increases in volatility.[1] 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

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

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), 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 better predictive power 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.[2]

The rest of the paper is structured as follows. Section 2 describes 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 the search index relates to standard housing market determinants. Section 3 contains an empirical analysis of the predictive power of search activity in the housing market with regards to future house prices. Section 4 explores variation in local housing markets and relates our findings to variation in local supply elasticities

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

and speculative demand. Finally, Section 5 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.

2.1 Search as a Leading Indicator for Housing Demand

We start by briefly motivating our choice of housing search activity as a leading indicator for demand in the housing market by building on the theoretical backbone of the search and matching literature. The idea behind these 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, searchers 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, which helps motivate that housing search as a proxy for demand should contain predictive power for future house price changes.[3]

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) about buyer-side search intensity since they argue that

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

although we have measures for seller-time-on-market, there is no parallel for buyers' time-on-market as proxy for buyer search. Since buyers are arguably more active than sellers are, 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), but 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 as their study does. Piazzesi et al. document that search activity is positively correlated with house prices in the cross-section of U.S. cities. 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.[4]

Carrillo et al. (2015) develop a search and matching model in which measures of market tightness, which is 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 in predicted sale probabilities. 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

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

prices in the future.[5] 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 adjustment 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 demand today will lead to price changes in future periods.

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.[6] 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

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

[6] Other search engines exist. However, 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.

reasons. First, each 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 period runs from 2004:1 to 2019:9 and uses the monthly frequency.[7] 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 directly 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.[8]

Our aggregate measure of housing demand is constructed as the first principal component of the search volume indices for the 23 keywords that are all intuitively related to the search process of future home buyers. This principal component accounts for more than 25% of the total variance of the underlying search volume indices and the individual keywords generally have high positive loadings on the first principal component. We therefore interpret this principal component as a summary measure for

[7] 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.

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

housing search and refer to it as the Housing Search Index (HSI).[9]

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 logarithm.[10] 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 (2020).[11] 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 the HSI along with the log growth rate in the Federal Housing Finance Agency (FHFA) House Price Index. Housing search and growth in house prices move closely together. In particular, we note that the 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.

[9] As a robustness check, we have analyzed the effect of including the second and third principal components and there are no predictive gains of including these.

[10] 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. Yu et al. (2019) use an ADF test on three Google Trends queries: "oil inventory", "oil consumption" and "oil price" and find evidence of stationarity at the 5% level (10% level) in "oil inventory" ("oil consumption"), but these authors are not able to reject the null of a unit root for "oil price". Da et al. (2015) study the log-differences (growth rates) of their data.

[11] 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 by a similar methodology as before but include a quadratic trend in the regression; otherwise we take first differences.

To explore the dynamic relation between the 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 the HSI:

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

where p_t is the log of the FHFA house price index in month t and j ranges from $j = -12$ to $j = 12$. We find much larger coefficients and R^2-values using lags rather than leads of the HSI, suggesting that movements in the HSI precede movements in the FHFA house price index. The strongest statistical relation between the HSI and changes in house prices occurs at lags of the HSI ranging from one through four months. At these lags, the predictive power of the HSI over monthly house price changes exceeds 50%. Leads of the 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 the R^2-values.

Table 1 shows results from tests of bi-directional Granger causality between the HSI and house price changes. Regardless of lag length, we generally find that the Granger causality runs from the HSI to house price changes and not the other way around, once two or more lags are included. Overall, the results indicate that the 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 single-family 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.

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

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

where $sales_t$ is the sales of existing single-family housing units from NAR in month t, and j ranges from $j = -12$ to $j = 12$. Panel B of Figure 2 shows the slope coefficients, associated t-statistics and R^2-values as functions of j. Their large values for $j < 0$ strongly suggest that search activity leads house sales. The slope coefficients imply that a one standard deviation increase in the HSI is associated with an increase in house sales of almost 600,000 units one year ahead. 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 the HSI is useful in forecasting home sales, while the reverse is not the case.[12]

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 (2-3 months) for house prices than for actual home sale transactions (12 months).

[12] 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 the HSI to home sales and not the other way around.

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

To better understand the drivers behind housing search, we regress the HSI on a set of commonly used housing market determinants, focusing on the aggregate housing market to ensure the longest available data series. 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 (*employ*): The 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 change in the 30-year fixed mortgage rate.

- Price-rent ratio (pr): The log ratio of the house price to the rent of primary residence.

- 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.[13]

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 inflation, the term spread, mortgage rate, price-to-rent ratio, and loans outstanding bear little-to-no relation to the volume of housing search. That is, we cannot explain movements in housing demand as measured through online search activity by means of changes in interest rates and credit conditions – at least not during our sample period from 2004 to 2019. Building permits and housing starts are both significantly positively related to housing search volume. However, with R^2 values around 3-7%, they explain only a very small part of variation in the HSI. In contrast, housing search volume is strongly positively correlated with employment and sentiment as reflected in R^2-values around 15-18%. Housing search volume thus tends to increase in times with high employment and general optimism about conditions for buying a house.

Combining our full list of standard housing market determinants in a multivariate regression (right column), we can explain around 35% of the variation in the HSI. With almost two-thirds of the variation in the HSI left unexplained, a large component

[13] All other variables are from the Federal Reserve Bank of St. Louis (FRED) database.

of time-series movements in the volume of housing search is, thus, uncorrelated with standard activity measures from the housing market.

2.6 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 first analyze whether the effect from housing search activity on house prices depends on the 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.

Search volume and the number of related terms vary across geographical regions and so the number of predictor series for each MSA also varies. To handle this issue, we follow Bai and Ng (2008) and use a targeted PCA approach which ensures that only the most relevant search indices are included to compute local demand factors.[14]

To illustrate the differences across local housing markets, Figure 3 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 notable boom-bust cycle from 2004 to 2010 and monthly growth rates never stray far away from zero. Although the link between HSI and growth in house prices is less clear for Wichita, we do find a significant relation between the two. However, from Figure 3 we should expect differences in the economic effect on local house prices from shocks to local $HSIs$. We further explore this point 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 the 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 (e.g. Berkovec and Goodman, 1996,

[14] Specifically, for each MSA we use the elastic net estimator of Zou and Hastie (2005) to select the ten most relevant search indices, then apply principal component analysis to summarize the most important information from these ten indices into one common component which constitutes the local housing search index.

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 house price movements, we estimate predictive regressions

$$p_{t+h} - p_t = \alpha_h + \beta_h x_t + \varepsilon_{t+h}, \tag{4}$$

where p_t is the log of the FHFA house price index, h is the forecast horizon, and x_t is a predictive variable which is either our housing search indicator, HSI, or a variable taken from the list of housing market determinants described earlier. We consider four different horizons, namely $h = 1, 3, 6$ and 12 months.

Table 3 reports the estimate of β_h, the corresponding t-statistic in parenthesis, and the R^2 in square brackets.[15] The housing search index is seen to be a very strong predictor of future house prices with significantly positive slope estimates, consistent with future house prices rising when current search (demand) for housing is high. Moreover, the estimated slope coefficients increase with the horizon, indicating that the relation strengthens at longer horizons. In fact, the predictive power of the HSI is substantial, with R^2-values ranging from 52% at the one-month horizon to around 70% for $h = 6$ and 12.[16] The economic magnitude is also large, as a one standard deviation increase in the HSI is associated with a 4.55% increase in expected house price growth at the one-year horizon.

Among the standard predictors, employment and sentiment stand out for their statistical significance although they clearly explain less of the variation in house prices than housing search activity. Employment generates an R^2 around 20-25% across the

[15] t-statistics are computed using the Newey and West (1987) procedure with h lags.

[16] Estimating the predictive regression on the individual search terms, "buying a house" delivers the highest R^2 ranging from 37% for $h = 1$ to 57% for $h = 12$. Across the four horizons, HSI delivers an R^2 that is roughly 12-20 percentage points higher than this best performing individual search term. Although "buying a house" is a strong predictor in itself, these results emphasize the value added by also including related search terms.

four forecast horizons, while the R^2 associated with sentiment ranges between 11% and 30%. The remaining predictors yield R^2 values between 0 and 7%. None of the standard predictors thus come close to matching the predictive power of the HSI over future house prices.

These results suggest that housing search activity carries important information about future house prices over and above the information embedded in standard housing market predictors. To further verify this claim, we use the residuals from the multivariate regression in Table 3 to construct a version of the housing search index that is orthogonal to the standard predictors, which we denote by $HSI^{\perp}$. The final row in Table 3 shows that the slope coefficients for $HSI^{\perp}$ remain positive and highly significant with R^2-values between 23-36%. This corroborates our finding that the HSI contains important information about future house prices that is not subsumed by standard housing market variables.

3.1 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 the 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 4 summarizes the slope coefficients, associated t-statistics and R^2 values for horizons up to five years ($h = 60$). The figure shows that the HSI is a significant predictor of house price growth up to a horizon of roughly four years, but also that the explanatory power steadily declines after its peak at horizons around 5 to 10 months.

Our 16-year sample from 2004-2019 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 10 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. Among the keywords underlying HSI, this initial step is captured by "steps to buying a house" and "house buying process". 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.

3.2 Adjustments for Serial Correlation

A number of studies have documented that growth in house prices exhibit positive serial dependence (e.g., Case and Shiller, 1989). Serial correlation can arise due to frictions and illiquidity and may also reflect the procedure used to construct the house price indices (Ghysels et al., 2013).[17]

To verify that HSI's predictive ability of future house prices is not just driven by autocorrelation in house price changes, we consider two simple unsmoothing techniques. First, following Getmansky et al. (2004), we unsmooth the house price changes using an MA(3) filter.[18] In the case of illiquidity-driven smoothing, we would expect the effects to be relatively short-lived such that a three-month lag length specification should suffice. Panel A of Table 4 shows that the HSI has strong predictive power for the MA(3)-filtered house price changes and is highly statistically significant across

[17]Although the FHFA index is a repeat-sales index, the FHFA calulates their monthly house price index without the use of temporal aggregation, which would have been a direct source of autocorrelation. 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.

[18]That is, we estimate the equation, $p_t - p_{t-1} = \alpha + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3}$, and then use the residuals $\hat{\varepsilon}_t$ as the dependent variable.

all horizons. We observe an R^2 of more than 20% at the one-month horizon, rising to more than 50% at the one-quarter horizon and approximately 65% at the six-month and one-year horizons.

Next, we follow Fischer et al. (1994) and Miller et al. (1994), among others, and use an AR(1) model to unsmooth the house price changes.[19] In contrast to the MA(3) model, the AR(1) model allows for an infinite number of smoothing lags with exponentially decaying weights. Panel B of Table 4 shows that the HSI also has strong predictive power for the AR(1)-filtered house price changes. We record an R^2 of almost 10% at the one-month horizon, which increases to 39% at the one-quarter horizon, and 57% at the six-month and one-year horizons. Furthermore, the HSI is highly statistically significant across all horizons. These results confirm that the HSI retains its strong predictive ability when using unsmoothed house price changes.

3.3 Out-of-Sample Tests

In-sample predictive regressions such as those reported in Table 3 can be criticized for overfitting – particularly in the multivariate case – and also could not have been utilized in real time to generate forecasts of house prices, in part because the HSI uses full-sample information. To address these issues, we next consider a set of out-of-sample forecasting experiments in which we recursively compute the 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 (2007-2019) for out-of-sample testing.[20]

Panel A in Table 5 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

[19] For the AR(1) model, we estimate, $p_t - p_{t-1} = \alpha + \beta(p_{t-1} - p_{t-2}) + \varepsilon_t$, and then use the residuals $\hat{\varepsilon}_t$ as the dependent variable.

[20] We use an expanding estimation window but obtain similar results with rolling windows.

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 the HSI is able to explain 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 $h = 6$ with $R^2_{OoS} = 66\%$, declining to $R^2_{OoS} = 57\%$ for $h = 12$. The Diebold-Mariano tests strongly reject the null hypothesis that $R^2_{OoS} \leq 0$ at all forecast horizons.

Further, Panel A shows that forecasts from the HSI strongly outperform forecasts based on popular determinants of house prices across all horizons. In most cases, these variables generate R^2_{OoS} statistics that are close to zero or negative. The main exception is the sentiment variable which generates a positive R^2_{OoS} across all horizons (see also Cox and Ludvigson, 2019, and Bork et al., 2020). However, the R^2_{OoS} generated by sentiment is notably lower than that of the HSI and this variable fails to generate statistically significant Diebold-Mariano test statistics.

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 (CSSFE) for $h = 1$ in Figure 5. 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. Positive values at the end of the sample show that a given forecasting model produced more accurate out-of-sample forecasts, on average, than the benchmark with negative values suggesting the reverse. For comparison we also show the results for the employment and sentiment variables. HSI (top panel) performs especially well during the housing bust from 2007 to 2009 but online search activity also holds important information about future house prices in less turbulent times and there are no periods with notable underperformance against the benchmark. The employment and sentiment variables (middle and bottom panels) also performed well during the housing bust period. However, in contrast to HSI these variables fail to predict time-variation in house prices in the subsequent sample, including the 2009-2012 recovery in house prices.

In conclusion, our out-of-sample analysis confirms the strong in-sample predictive ability of the 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 to the standard house price determinants that generally have difficulties predicting future house prices out-of-sample.

3.3.1 Bootstrap Analysis

To further validate the statistical significance of the HSI in forecasting house prices, we consider a simulation experiment that is comparable to the "useless" factor tests of Kan and Zhang (1999a,b). In particular, we generate 10,000 bootstrap samples by row-wise resampling from the panel of 23 search indices (with replacement). The resampled panels have the same length as the original panel of search indices. For each bootstrap sample, we recursively estimate the HSI and generate out-of-sample forecasts, then save the R^2_{OoS} statistic. As the resampled placebo Google search data should bear no relation to the realized house price growth rates, the HSI should not be useful in forecasting growth in house prices. Basically, the resampled search indices represent random noise and so are "useless".

We analyze the empirical distribution of the R^2_{OoS} statistic by computing empirical p-values. The simulations in Panel B in Table 5 show that the share of bootstrapped R^2_{OoS} statistics that exceed their empirical counterparts from Panel A equals zero across all horizons. Hence, the chance of obtaining the same goodness-of-fit with random Google data as we find with the actual data is virtually zero.

As further robustness checks, we consider two alternative bootstrap procedures that take into account the persistence in the data. The first one uses a parametric bootstrap in which we estimate an AR(1) model for each Google series and retain the estimated coefficients along with the residuals from each regression to construct a panel of placebo series that have the same autoregressive coefficient and variance as the series in the Google Trends panel. In addition, we consider a non-parametric circular block bootstrap procedure similar to the row-wise resampling above. However, instead of drawing one row at a time, for each series we build the placebo series from blocks of size m. For each series, we select the optimal value of m, using the automatic selection procedure developed by Politis and White (2004). Results from these two alternative bootstrap procedures are also shown in Panel B. The findings are identical to those obtained from the row-resampling bootstrap, implying that it is extremely unlikely that the observed R^2_{OoS} were due to chance.

These robustness tests corroborate the robustness of our findings on the highly significant out-of-sample predictive power of the HSI over future movements in house prices.

3.4 Comparison with Wu and Brynjolfsson (2015)

Our construction of the 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. In a related paper, 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 not completely clear how we should interpret these categories other than they relate to the topic given by the name of the category. Wu and Brynjolfsson (2015) find that these two search categories hold limited predictive power for future house prices compared to future house sales.[21]

In their empirical analysis, Wu and Brynjolfsson (2015) forecast house prices in levels, which complicates a direct comparison with our results which use log-changes in house prices. Furthermore, their sample period ends in 2011.

To facilitate a direct comparison with Wu and Brynjolfsson (2015), Table 6 explores the predictive power of the two search categories “Real estate agencies” and “Real estate listings” and compares these to HSI.[22] Panel A shows that the two predefined categories hold no predictive power for growth in house prices. The slope coefficients are not significantly different from zero and the R^2 values range between 0.6% and 2%. In contrast, Panel B shows that HSI retains its strong predictive power after we control for the two predefined search categories. 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 it is clear that "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 search 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.

In summary, we confirm Wu and Brynjolfssons (2015) finding that house prices are difficult to predict using the predefined categories “Real estate agencies” and “Real estate listings”. 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 the HSI suggest an additional explanation, namely that the predefined categories contain too much irrelevant information which distorts the predictive power

[21] Dietzel (2016) takes a similar approach and uses subcategories related to real estate to analyze turning points in housing markets.

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

of search activity.[23]

3.5 Buying versus Selling Side of the Housing Market

Wu and Brynjolfsson (2015) conjecture that house prices are difficult to predict using the two predefined categories since they potentially reflect both the buying and selling sides of the housing market. This motivates us to explore the predictive power of a search index based on the main search term "selling a house" instead of "buying a house". We follow the approach used in constructing HSI described in Section 2.2, but now use a keyword intended to capture the selling side of the housing market. The related search terms are: "when selling a house", "selling a home", "selling your house", "selling my house", "selling a house taxes", "how to sell a house", "selling your home", "tax on selling a house", "selling house by owner", "cost of selling a house", "capital gains", "taxes on selling a house", "closing costs", "capital gains tax", and "selling a house tips". These search terms are all directly related to the home selling process. We denote the first principal component of these search terms HSI^{sell}.

Panel C in Table 6 shows that this search index based on the selling side of the housing market also contains important information about future house prices. All slope coefficients are significantly positive with R^2-values ranging between 22% for $h = 12$ and 37% for $h = 3$. However, once we control for our original HSI measure (based on housing demand), the predictive power of HSI^{sell} declines and at longer horizons its slope coefficient is no longer significantly different from zero (Panel D). Comparing the R^2-values from the model that contains both the HSI and HSI^{sell} measures to those based only on the HSI in Table 3, we observe only a very small increase from adding HSI^{sell} to a model that already contains HSI.

[23] We have 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 suggested procedure has considerably more predictive content for future house prices than that of Beracha and Wintoki (2013). For example, in Miami, Toledo, and Houston, our local HSI's generate R^2s of 69%, 64%, and 48% at the one-month horizon compared to 10%, 7%, and 0% when using "real estate Miami", "real estate Toledo", and "real estate Houston", respectively.

The slope coefficients for HSI^{sell} have the same positive sign as for HSI, i.e. an increase in search activity on the selling side of the housing market is associated with an increase in future house prices. Accordingly, we are careful not to interpret HSI^{sell} as a measure of housing supply. To better understand the nature of HSI^{sell}, we plot the search index together with HSI and the log growth rate in the FHFA House Price Index in Panel D of Figure 1. HSI generally tends to lead HSI^{sell}, which suggests that households tend to search for a new house before selling their existing home.

In conclusion, search activity on the buying side of the housing market appears to dominate search activity on the selling side in terms of predictive power over movements in future house prices.

4 Variation in Search across Local Housing Markets

While national accounts data are often limited in geographic scope, 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 the HSI, we estimate MSA-level regressions,

$$p_{it+h} - p_{it} = \alpha_i + \beta_i HSI_{it} + \varepsilon_{it+h}, \tag{5}$$

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 6 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 2.6. To ease comparisons across MSAs all search indices are standardized and the slope coefficients are multiplied with 1,200, such that they 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 the HSI at the national level reappears in individual local housing markets with slope coefficients that are significantly positive for all except one MSA and with 54 MSAs generating R^2 values exceeding 40%.

Across the 77 MSAs, the estimated slope coefficients range from 0.45 (Rochester) 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 12.4% increase in expected house price growth in Miami the following month, while the corresponding response is only 1.4% in Wichita. Figure 6 illustrates these differences across local housing markets.

To further corroborate our interpretation of the HSI as capturing latent demand for housing, we consider this cross-sectional dispersion in local housing markets along two dimensions. First, we would expect the effect on house prices of variation in housing demand measured through HSI to be stronger in housing markets where the supply is more constrained. Following Saiz (2010), we therefore analyze whether the HSI is associated with stronger positive effects on house prices in MSAs with a more inelastic housing supply. Second, since speculation represents a source of housing demand (Gao et al., 2020), we should see larger effects of HSI in markets where there are more housing speculation.[24] We next explore these two dimensions empirically. After that we analyze the effect of variation in national-level versus MSA-level search and, finally, we analyze potential economic gains from exploiting the predictability in house prices and the link between search activity and house price expectations.

[24]Piazzesi and Schneider (2009) use a search model to show that a small number of optimistic investors can have a large price impact.

4.1 Local Variation in Supply Elasticities

To address the hypothesis that variation in housing demand affects house prices more strongly in local markets with lower supply elasticity, 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 (5). 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.

The more difficult it is to expand the housing supply, the greater the effect of variation in housing demand on house prices. Accordingly, we split the MSAs in two groups based on their 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

$$p_{it+h} - p_{it} = \alpha_i + \left(\beta + \beta_E \times I_i^E\right) HSI_{it} + \varepsilon_{it+h}, \tag{6}$$

where I_i^E is a dummy variable that is equal to 1 if the supply elasticity in MSA i is below median. Thus, β_E measures the additional effect of HSI on house prices in low supply-elasticity MSAs. Panel B in Table 7 shows the results. For both high and low supply-elasticity MSAs, we find a significant relation between housing search and future growth rates in house prices. However, the economic significance of the relation between variation in demand and house prices is notably stronger in MSAs with low supply elasticity compared to those with high supply elasticity. For example, at the annual horizon, the impact on house price growth from a one standard deviation change in the HSI is on average $3.32\%+2.11\% = 5.43\%$ in low supply-elasticity MSAs compared to an average of 3.32% in high supply-elasticity MSAs. To visualize these results, Figure 6 shows the ten most supply-constrained MSAs in red and the ten least supply constrained MSAs in green. We see a clear clustering of the MSAs in accordance with the panel results in Table 7.

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 notably larger economic impact on house prices in MSAs with a more constrained supply of housing.

4.2 Local Variation in Speculative Demand

During the early 2000s house prices in many MSAs increased dramatically and reached record high levels, which was followed by a collapse in house prices and a severe crisis in the U.S. economy. A growing literature suggests that speculation in the housing market was an important driver of the boom and argues that economic fundamentals accounted for just a small fraction of the changes in prices during the housing boom (e.g. Akerlof and Shiller, 2009, Chinco and Mayer, 2016, and Nathanson and Zwick, 2017).[25] Given that the HSI is a direct measure of peoples' intention to buy a house and hence captures the demand side of the market, we would expect that the predictive power of HSI is systematically linked to the degree of speculation across MSAs.[26]

[25] Other contributing factors to the boom and bust in house prices have been put forward in the literature, including credit conditions (Mian and Sufi, 2009, and Favilukis et al., 2017) and low interest rates resulting from excessively loose monetary policy (Taylor, 2014).

[26] An advantage of the HSI as a predictor is that it simply reflects peoples' interest in buying a house. As such it can be used to capture both fundamental and non-fundamental sources of demand for housing.

Following Gao et al. (2020), we measure speculation as the fraction of non-owner-occupied home purchases using the Home Mortgage Disclosure Act (HMDA) dataset which allows us to map individual mortgage level data to the 77 MSAs in our sample. As Gao et al. (2020) point out, decisions to buy a non-owner-occupied home are to a greater extent driven by speculative motives than decisions to buy a primary home.

Figure 7 plots the estimate of the predictive coefficient β_i from the MSA-level regressions in (5) against the degree of housing speculation as measured by the fraction of non-owner-occupied home purchases across MSAs. The degree of housing speculation is strongly positively correlated with the size of the estimated predictive coefficient of the HSI. For example, in Stockton and Las Vegas – both among the MSAs with the greatest magnitude of housing speculation – a one standard deviation increase in the HSI leads to an increase in the expected growth rate of local house prices of more than 14 percentage points per year. In contrast, in Hartford and Fort Wayne – both among the MSAs with the least housing speculation – a one standard deviation increase in the HSI leads to an increase in the expected house price growth rate of less than four percentage points per annum.

To more formally test whether there is an additional effect from the HSI in areas with more intense housing speculation, we estimate

$$p_{it+h} - p_{it} = \alpha_i + \left(\beta + \beta_S \times I_i^S\right) HSI_{it} + \varepsilon_{it+h}, \tag{7}$$

where I_i^S is a dummy variable that equals one if the fraction of non-owner-occupied home purchases in MSA i is above median. Thus, β_S measures the additional effect of HSI_{it} on house prices in MSAs with high degrees of housing speculation. Consistent with the visualization in Figure 7, the estimates of β_S in Panel C of Table 7 imply that changes in the HSI have a significantly larger impact on house prices in MSAs with stronger degrees of housing speculation. Moreover, both the magnitude of the estimated coefficients as well as their explanatory power is very similar to that found for the regression that accounts for supply elasticity (equation (6)): moving from MSAs with below-median levels of speculation to MSAs with above-median speculation, the impact on local growth in house prices of a one standard deviation increase in the HSI is 0.2% at the one-month horizon, 1.2% at the six-month horizon, and 2.1% at the 12-month horizon. These results relate to the work of Piazzesi and Schneider (2009), showing that a small number of optimistic investors can have a large price impact.

Taken together, our results demonstrate that the HSI is a better predictor of future movements in home prices in housing markets with (i) inelastic supply; and (ii) greater speculative activity. These channels may of course be related, so to see if they are individually important, we next consider a panel regression model that includes both I_i^E and I_i^S. From the results in Panel D of Table 7, we see that both effects are statistically significant across all horizons.[27]

We conclude from these results that local housing search generally has strong predictive ability for growth in local house prices but that the HSI is a particularly strong predictor of house prices in markets with inelastic housing supply and high speculation.

4.3 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 our estimates of the effect of local supply elasticity and local

[27] I_i^E and I_i^S have a correlation of 0.22, suggesting that they capture different aspects of the variation in local housing markets.

speculative investment activity are 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 as well as in MSAs more prone to speculation in the housing market.

4.4 Economic Significance

Our results show that there is strong variation in search activity over time and across MSAs and that search activity as a proxy for demand predicts future house prices – a finding that is in line with the theoretical search-and-matching modelling framework. An obvious question is then whether the time-variation in expected house prices provides economic opportunities for potential homebuyers. Suppose, for instance, that house prices are predicted to increase by, say, 4% in the next six months in a given MSA. Does this represent an economic opportunity? If the potential homebuyer buys now, it will lead to expected savings of 4%, but the potential homebuyer will not have the chance to engage in time-consuming search which increases the probability of buying a house that is not a good match. If the potential homebuyer waits, house prices will probably go up, but the potential homebuyer gets to inspect a larger set of houses which come on the market in the interim. These types of considerations are not a concern in markets with homogeneous goods, but of course housing is a heterogeneous good involving a complicated and time-consuming search process.

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 2,194 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 0.6% at the one-month horizon, 1.8% at the three-month horizon, 3.5% at the six-month horizon and 6.9% at the annual horizon. These realized house price changes suggest that significant economic gains can be achieved from being an early buyer in a market with increasing demand. Those gains come with a higher risk of buying the wrong house and so should be balanced against the opportunity of engaging in a more thorough (time-consuming) search and finding a better match.

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 mean, the 25th percentile growth rate is 0.4%, 1.3%, 2.6%, and 5.2% at the one-month, three-month, six-month, and one-year ahead horizons, while for the 75th percentile, the price changes are 0.9%, 2.7%, 5.2%, and 9.9%, respectively.[28]

Given that the HSI captures local changes in housing demand, we would also expect episodes with low search activity to coincide with subsequent downward pressures on house prices. We identify periods with low search activity as months in which the local housing search falls one standard deviation of its mean. Across MSAs, we find 2,606 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%, -1.0%, -1.9%, and -3.4% at the one-month, three-month, six-month, and one-year ahead horizons. This means that it is risky to buy early during times when the HSI is low. That is, there is a strong incentive to suspend or reduce search efforts because it is possible to buy a house at a lower price, the longer the potential homebuyer waits.

Overall, our results suggest that the potential economic gains from exploiting the predictability in house prices can be quite large. However, we of course cannot rule out that the predicted house price gains are just large at times when there is little transparency in the market so that the expected gains from buying early is just a risk premium/compensation for the elevated risk of buying a poor-quality home.

[28] The MSAs in the 1st quartile have an average supply elasticity of about 2.5, while those in the 4th quartile typically are more inelastic markets with an average supply elasticity of about 1.2.

4.5 Search and Price Expectations

The estimates of changes in house prices based on search activity are model-based and so there is no guarantee that they accurately represent home buyers' expectations. As we next discuss, however, we can obtain more direct measures of households' expectations about future house prices.

Following the analysis in Shimer (2004), analogous with search in the labor market, we can think of home buyers' optimal search intensity as depending on three factors, namely (i) the sensitivity of the likelihood of finding a desirable home with respect to variation in search intensity. If the chance of finding a suitable home is highly sensitive to the amount of search, home buyers should be more willing to vary their search efforts in response to shifts in the housing market. Conversely, if the probability of finding a home is either very low (due to a tight housing market) or very high (due to an excess of supply), home buyers are unlikely to vary their search by much due to such shifts; (ii) the expected present value of rents or user benefits from owning a home, including shifts in expectations of future house prices. If home buyers expect future prices to be much higher, they should increase their search intensity, expecting to benefit from such price increases; (iii) the marginal cost of searching. This may change, e.g., as a result of new online search tools being launched (decreasing search costs) but could also simply reflect variation in the marginal value of time across economic recessions and expansions.

The third factor is likely to vary less over time than the first two. Provided that the cost of search is relatively constant, variation over time in search intensity should predominantly be driven by movements in the returns to search, i.e., the first two factors.

While we do not directly observe the likelihood of finding a house, we can construct a measure of house price expectations. Since 2007 the University of Michigan Surveys of Consumers has asked homeowners: "what do you think will happen to the prices of homes like yours in your community over the next 12 months?". To get a measure of house price expectations, we use the monthly time series of the fraction of people saying that house prices will increase minus the fraction responding that they will decrease. The survey data are available at the regional level (West, North Central, Northeast, and South), which we match with the MSA-level house search indices by taking averages within each region. We then analyze the relation between search activity and house price expectations by computing correlation coefficients across regions. The results indicate that housing search intensity is strongly linked to house price expectations as the correlation coefficients range from 0.79 (Northeast) to 0.82 (West). These results are consistent with the notion that home buyers increase their search intensity and, thus, their demand, in part because of higher expected future house prices.

5 Concluding Remarks

Fluctuations in house prices have 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"). Producing reliable and accurate predictions of house prices is evidently of great importance and can, for example, be used for early warning of an incipient bubble in the housing market.

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 conventional measures of variation in housing demand such as employment, interest rates, or sentiment. 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.

References

Akerlof, G., Shiller, R. (2009). Animal spirits: How human psychology drives the economy and why it matters for global capitalism. Princeton, NJ: Princeton University Press.

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

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

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

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. (2020). In search of a job: Forecasting employment growth using google trends. *Journal of Business & Economic Statistics*, forthcoming.

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.

Chinco, A., Mayer, C. (2016). Misinformed speculators and mispricing in the housing market. *Review of Financial Studies* 29, 486-522.

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

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.

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, Francke, M. (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.

Favilukis, J., Ludvigson, S., Van Nieuwerburgh, S. (2017). The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium. *Journal of Political Economy* 125, 140-223.

Fischer, J.D., Geltner, D.M., Webb, R.B. (1994). Value indices of commercial real estate: A comparison of index construction methods. *Journal of Real Estate Finance and Economics* 9, 137-164.

Gao, Z., Ren, H., Zhang, B. (2020). Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis* 55, 549-580.

Gao, Z., Sockin, M., Xiong, W. (2020). Economic consequences of housing speculation. *Review of Financial Studies* 33, 5248-5287.

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

Getmansky M., Lo, A.W, Makarov, I. (2004). An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics* 74, 529-609.

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.

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.

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.

Kan, R., Zhang, C. (1999a). GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics* 54, 103-127.

Kan, R., Zhang, C. (1999b). Two-pass tests of asset pricing models with useless factors. *Journal of Finance* 54, 203-235.

Krainer, J. (2001). A theory of liquidity in residential real estate markets. *Journal of*

Urban Economics 49, 32-53.

Mian, A, Sufi, A. (2009). The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *Quarterly Journal of Economics* 124, 1449-1496.

Miller, M.H, Muthuswamy, J., Whaley, R.E. (1994). Mean reversion of Standard & Poor's 500 index basis changes: Arbitrage-induced or statistical illusion? *Journal of Finance* 49, 479-513.

Nathanson, C.G., Zwick, E. (2018). Arrested development: Theory and evidence of supply-side speculation in the housing market. *Journal of Finance* 73, 2587-2633.

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

Piazzesi, M., Schneider, M. (2009). Momentum traders in the housing market: Survey evidence and a search model. *American Economic Review, Papers and Proceedings* 99, 406-411.

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.

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

Shimer, R. (2004). Search intensity. Working Paper, University of Chicago.

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

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

Taylor, J.B. (2014). Causes of the financial crisis and the slow recovery: A 10-year perspective. Maily M, Taylor JB, eds. Across the Great Divide: New Perspectives on the Financial Crisis (Hoover Press, Stanford, CA), 51-65.

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.

Yu, L., Zhao, Y., Tang, L., Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google Trends. *International Journal of Forecasting* 35, 213-223.

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 or home sales. The table shows p-values from the joint test that $\gamma_1 = \gamma_2 = ... = \gamma_p = 0$. The sample period is 2004:1 to 2019:9.

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	0.021	0.107	0.358	0.165
HSI does not Granger cause home sales	0.000	0.002	0.004	0.026
Home sales do not Granger cause HSI	0.849	0.639	0.544	0.584

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 based on one lag. All variables are standardized to ease comparison of the β estimates. The sample period is 2004:1 to 2019:9.

	Univariate	Multivariate
employ	0.38	0.34
	(4.26)	(3.27)
	[14.57]	
infl	0.03	−0.07
	(0.24)	(−1.02)
	[0.10]	
permits	0.27	0.11
	(3.08)	(1.63)
	[7.03]	
starts	0.18	0.08
	(2.92)	(1.55)
	[3.22]	
term	0.17	−0.02
	(1.58)	(−0.17)
	[2.86]	
mort	0.09	0.02
	(0.96)	(0.29)
	[0.77]	
pr	−0.09	0.25
	(−0.92)	(2.50)
	[0.79]	
loans	−0.11	−0.18
	(−1.22)	(−2.40)
	[1.25]	
sent	0.44	0.46
	(5.46)	(4.38)
	[18.93]	[36.67]

Table 3. Predicting House Prices With Housing Search and Alternative Variables. 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 predictive variable, and h is the forecasting horizon in months. $HSI^{\perp}$ is the part of HSI that is orthogonal to the other predictive variables. 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 the Newey and West (1987) estimator based on h lags. All predictive variables are standardized and slope coefficients are multiplied with 100 to ease comparison across variables. The sample period is 2004:1 to 2019:9.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
HSI	0.40	1.20	2.42	4.55
	(12.12)	(12.41)	(11.85)	(9.24)
	[52.93]	[66.54]	[72.19]	[68.52]
employ	0.25	0.74	1.41	2.57
	(4.06)	(4.54)	(4.09)	(4.42)
	[20.47]	[25.37]	[24.47]	[21.80]
infl	−0.02	−0.15	−0.20	−0.36
	(−0.29)	(−0.99)	(−0.58)	(−0.60)
	[0.12]	[1.05]	[0.51]	[0.43]
permits	0.14	0.33	0.64	1.34
	(3.11)	(2.83)	(3.19)	(3.34)
	[6.87]	[5.05]	[5.11]	[5.91]
starts	0.04	0.16	0.37	0.78
	(1.08)	(2.08)	(2.45)	(2.37)
	[0.60]	[1.15]	[1.66]	[2.01]
term	−0.02	0.05	0.28	1.27
	(−0.32)	(0.22)	(0.49)	(0.84)
	[0.10]	[0.10]	[0.94]	[5.31]
mort	−0.00	0.03	0.10	0.28
	(−0.06)	(0.32)	(0.56)	(1.32)
	[0.00]	[0.05]	[0.13]	[0.27]
pr	0.01	−0.03	−0.25	−1.21
	(0.24)	(−0.15)	(−0.48)	(−0.93)
	[0.04]	[0.04]	[0.74]	[4.87]
loans	0.01	0.12	0.25	0.49
	(0.23)	(0.63)	(0.53)	(0.38)
	[0.07]	[0.63]	[0.78]	[0.79]
sent	0.19	0.58	1.30	2.98
	(3.92)	(3.40)	(2.92)	(2.92)
	[11.36]	[15.43]	[20.72]	[29.48]
$HSI^{\perp}$	0.26	0.82	1.70	3.25
	(6.73)	(6.38)	(5.97)	(4.81)
	[22.89]	[31.22]	[35.79]	[34.98]

Table 4. Predicting Unsmoothed House Prices With Housing Search. The table reports results from h-month ahead predictive regressions of unsmoothed house priced changes using the HSI as a predictive variable. 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 the Newey and West (1987) estimator based on h lags. We unsmooth the house price changes using either an MA(3) model (Panel A) or an AR(1) model (Panel B). The sample period is 2004:1 to 2019:9.

$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: MA(3) model			
0.18	0.55	1.10	2.04
(6.43)	(9.64)	(10.69)	(8.27)
[20.66]	[52.96]	[66.54]	[64.33]
Panel B: AR(1) model			
0.12	0.35	0.69	1.27
(4.17)	(7.20)	(8.69)	(6.58)
[9.11]	[38.84]	[57.88]	[57.23]

Table 5. 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 (t_{DM}), which is 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. In Panel B we generate 10,000 bootstrap samples of the 23 search indices used to compute the HSI. For each bootstrap sample, we recursively estimate the HSI as the first principal component of the resampled search indices, generate out-of-sample forecasts, and then compute the R^2_{OoS}. The table reports the empirial p-value, which is the share of artificial R^2_{OoS} statistics that exceed the actual R^2_{OoS} statistic. Results are shown for three different bootstrap techniques: row resampling, a parametric bootstrap, and a block bootstrap.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: Out-of-sample results				
HSI	50.06	64.85	66.03	57.15
	(0.00)	(0.00)	(0.00)	(0.01)
employ	−2.51	−18.60	−37.52	−22.34
	(0.53)	(0.70)	(0.77)	(0.77)
infl	0.74	−2.63	−1.40	−0.91
	(0.37)	(0.78)	(0.85)	(0.87)
permits	−6.38	0.80	−0.55	1.17
	(0.69)	(0.47)	(0.53)	(0.42)
starts	0.14	1.65	1.48	0.85
	(0.47)	(0.21)	(0.19)	(0.27)
term	−7.73	−11.81	−14.32	−16.03
	(0.83)	(0.79)	(0.74)	(0.66)
mort	−1.77	−1.84	0.13	−0.60
	(0.98)	(0.94)	(0.44)	(0.84)
pr	−16.03	−30.61	−44.76	−73.22
	(0.96)	(0.97)	(0.96)	(0.91)
loans	−6.98	−14.32	−27.87	−50.87
	(0.81)	(0.85)	(0.89)	(0.86)
sent	7.32	8.44	13.90	26.29
	(0.15)	(0.22)	(0.19)	(0.11)
Panel B: Bootstrapped p-values for HSI				
Row resampling	0.00	0.00	0.00	0.00
Parametric AR(1)	0.00	0.00	0.00	0.00
Block bootstrap	0.00	0.00	0.00	0.00

Table 6. 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 forecasting 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 the Newey and West (1987) estimator based on h lags. All predictive variables are standardized to ease comparison of the β estimates and the log price change is multiplied with 100. Panel A shows results for the predefined search categories used by Wu and Brynjolfsson (2015), Panel B includes the HSI jointly with the predefined search categories, Panel C shows the results for HSI^{sell}, which is an alternative search index based on the selling side of the housing market, and Panel D includes HSI and HSI^{sell} in a joint regression. The sample period is 2004:1 to 2019:9.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: Predefined search categories				
Real estate agencies	−0.08	−0.33	−0.77	−1.06
	(−0.96)	(−1.12)	(−1.11)	(−0.76)
Real estate listings	0.08	0.34	0.71	0.74
	(0.83)	(0.95)	(0.83)	(0.40)
	[0.60]	[1.44]	[1.96]	[1.07]
Panel B: HSI joint with predefined search categories				
HSI	0.41	1.22	2.43	4.59
	(11.87)	(12.39)	(11.88)	(9.51)
Real estate agencies	0.05	0.04	−0.03	0.35
	(0.86)	(0.33)	(−0.12)	(0.78)
Real estate listings	0.00	0.10	0.24	−0.19
	(0.05)	(0.74)	(1.00)	(−0.36)
	[53.67]	[67.42]	[72.79]	[68.67]
Panel C: Search index based on the selling side of the housing market				
HSI^{sell}	0.31	0.90	1.64	2.58
	(7.62)	(6.56)	(4.86)	(3.01)
	[32.08]	[37.39]	[33.13]	[22.06]
Panel D: HSI joint with the selling side of the housing market				
HSI	0.33	1.02	2.19	4.51
	(8.90)	(8.22)	(7.41)	(5.94)
HSI^{sell}	0.13	0.33	0.41	0.07
	(3.40)	(3.09)	(1.62)	(0.10)
	[56.66]	[70.06]	[73.64]	[68.53]

Table 7. Predicting Local House Prices With Local Housing Search: Evidence From Panel Regressions. The table reports results from cross-section fixed effects panel regressions of the form, $p_{it+h} - p_{it} = \alpha_i + \beta HSI_{it} + \beta_E HSI_{it} \times I_i^E + \beta_S HSI_{it} \times I_i^S + \varepsilon_{it+h}$, 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, I_i^E is a dummy variable that is equal to 1 if the supply elasticity in MSA i is below median, I_i^S is a dummy variable that is equal to 1 if the fraction of non-owner-occupied home purchases in MSA i is above median, and h is the forecasting 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 h lags. HSI is standardized to ease interpretation of the β estimates. The sample period is 2004:1 to 2019:9.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: HSI				
HSI	0.39	1.18	2.31	4.37
	(42.03)	(29.52)	(23.11)	(19.40)
	[33.73]	[34.98]	[35.41]	[34.68]
Panel B: Supply Elasticity				
HSI	0.29	0.87	1.72	3.32
	(31.66)	(22.57)	(17.66)	(14.76)
$HSI \times I^{Elasticity}$	0.21	0.62	1.19	2.11
	(11.62)	(8.02)	(6.18)	(4.87)
	[36.02]	[38.73]	[37.75]	[36.71]
Panel C: Speculation				
HSI	0.30	0.88	1.72	3.32
	(40.21)	(29.88)	(23.65)	(19.25)
$HSI \times I^{Speculation}$	0.20	0.60	1.18	2.12
	(10.93)	(7.83)	(6.15)	(4.90)
	[35.86]	[37.26]	[37.73]	[36.73]
Panel D: Supply Elasticity and Speculation				
HSI	0.23	0.69	1.35	2.66
	(25.60)	(18.96)	(15.25)	(13.06)
$HSI \times I^{Elasticity}$	0.17	0.51	0.97	1.73
	(9.68)	(6.91)	(5.26)	(4.09)
$HSI \times I^{Speculation}$	0.16	0.49	0.97	1.74
	(9.22)	(6.63)	(5.20)	(4.11)
	[37.36]	[38.81]	[39.23]	[38.02]

Table 8. Predicting Local House Prices With National and Local Housing Search: Evidence From Panel Regressions. The table reports results from cross-section fixed effects panel regressions of the form, $p_{it+h} - p_{it} = \alpha_i + \beta_{US} HSI_{USt} + (\beta + \beta_E \times I_i^E + \beta_S \times I_i^S) \times HSI_{it} + \varepsilon_{it+h}$, 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, I_i^E is a dummy variable that is equal to 1 if the supply elasticity in MSA i is below median, I_i^S is a dummy variable that is equal to 1 if the fraction of non-owner-occupied home purchases in MSA i is above median, and h is the forecasting 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 h lags. The local and U.S. HSI are standardized to ease comparison of the β estimates. The sample period is 2004:1 to 2019:9.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: U.S. vs. Local HSI				
U.S. HSI	0.27	0.83	1.67	3.15
	(28.94)	(22.10)	(17.92)	(14.53)
Local HSI	0.24	0.68	1.33	2.49
	(25.95)	(18.81)	(15.01)	(13.13)
	[43.87]	[46.33]	[47.44]	[46.36]
Panel B: Supply Elasticity and Speculation				
U.S. HSI	0.26	0.81	1.63	3.08
	(28.62)	(21.97)	(17.90)	(14.51)
Local HSI	0.09	0.24	0.46	0.97
	(9.20)	(6.15)	(4.62)	(4.24)
Local $HSI \times I^E$	0.16	0.48	0.92	1.62
	(9.91)	(7.07)	(5.38)	(4.21)
Local $HSI \times I^S$	0.14	0.44	0.86	1.55
	(8.90)	(6.46)	(5.08)	(4.05)
	[46.95]	[49.57]	[50.65]	[49.14]

Figure 1. Housing Search Index. Panel A shows the Housing Search Index (HSI) along with the log growth rate in the Federal Housing Finance Agency (FHFA) House Price Index. Panel B shows the 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. Panel D shows the HSI constructed based on search terms related to the buying side of the housing market along with HSI^{sell} constructed based on search terms related to the selling side of the housing market as well as the log growth rate in the FHFA House Price Index. The sample period is 2004:1 to 2019:9.

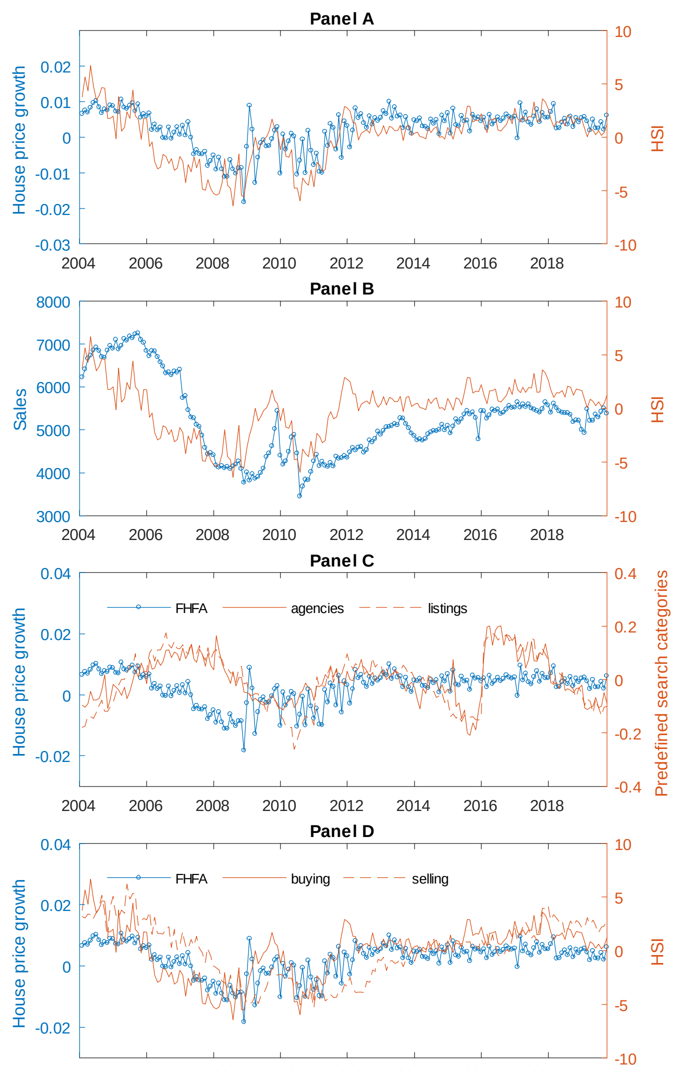

Figure 2. Lead-Lag Relations. Panel A shows regression slope coefficients, associated t-statistics and R^2 values of monthly price changes from $t-1$ to t on HSI_{t+j} for $j \in \{-12, 12\}$. Panel B shows the results from regressing monthly house sales at time t on HSI_{t+j} for $j \in \{-12, 12\}$.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

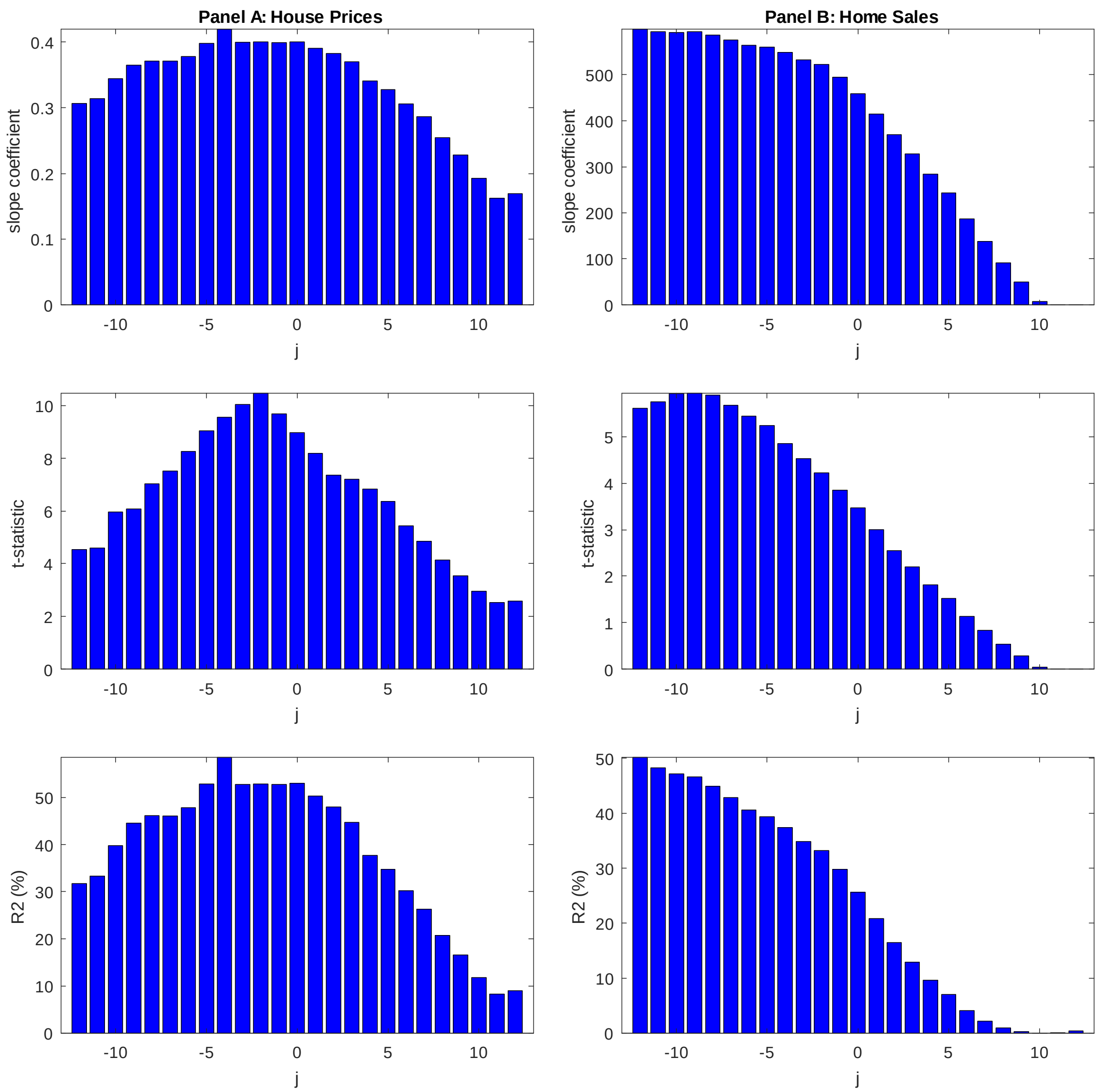

Figure 3. Local Housing Search. Panel A and B show the local HSI and log growth rate in house prices in Miami (FL) and Wichita (KS), respectively. The sample period is 2004:1 to 2019:9.

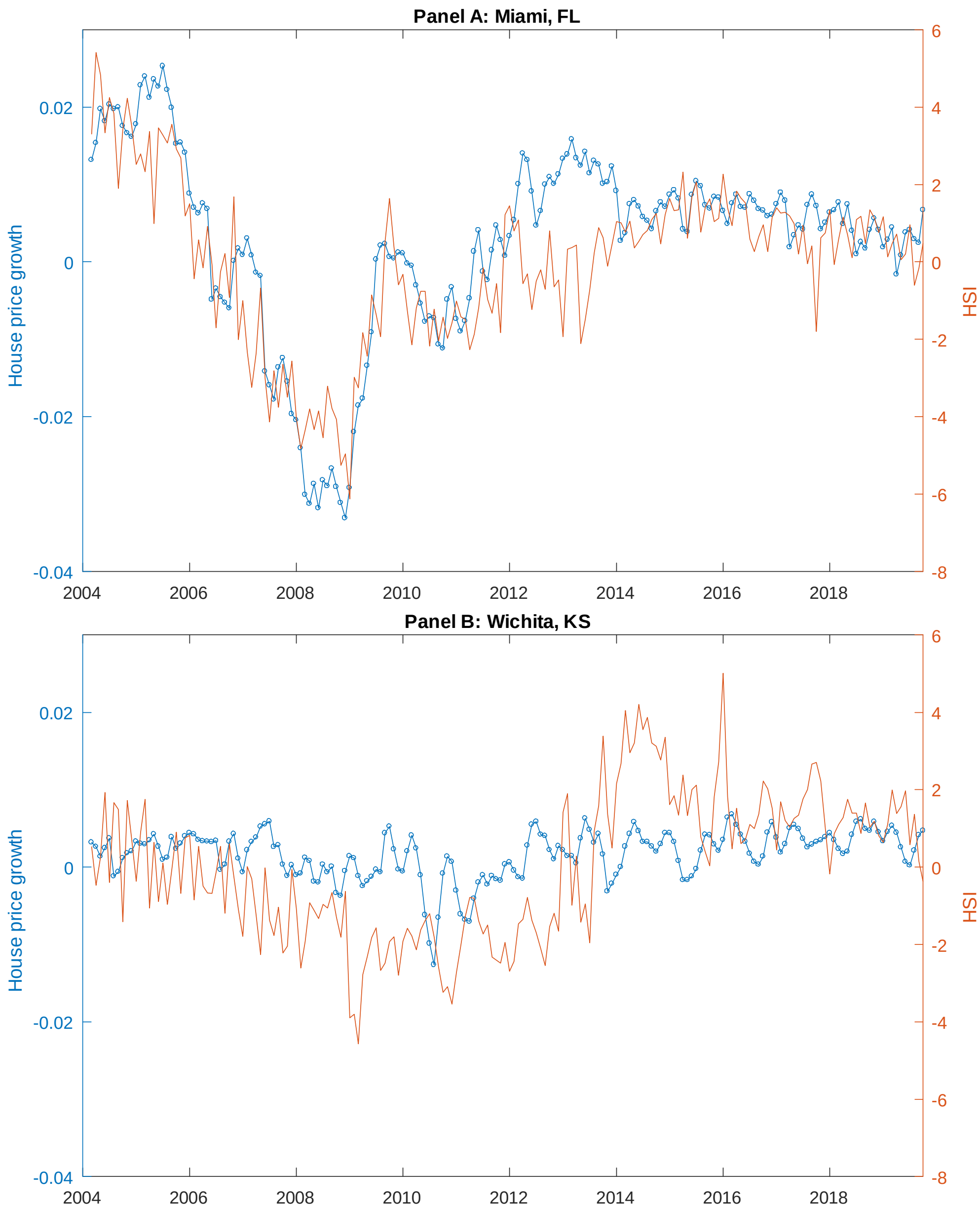

Figure 4. Long-Horizon Predictability. The figure shows regression slope coefficients, associated t-statistics and R^2 values from the regression, $p_{t+h} - p_t = \alpha + \beta HSI_t + \varepsilon_{t+h}$, as a function of h. We compute standard errors using the Newey and West (1987) estimator based on h lags. The sample period is 2004:1 to 2019:9.

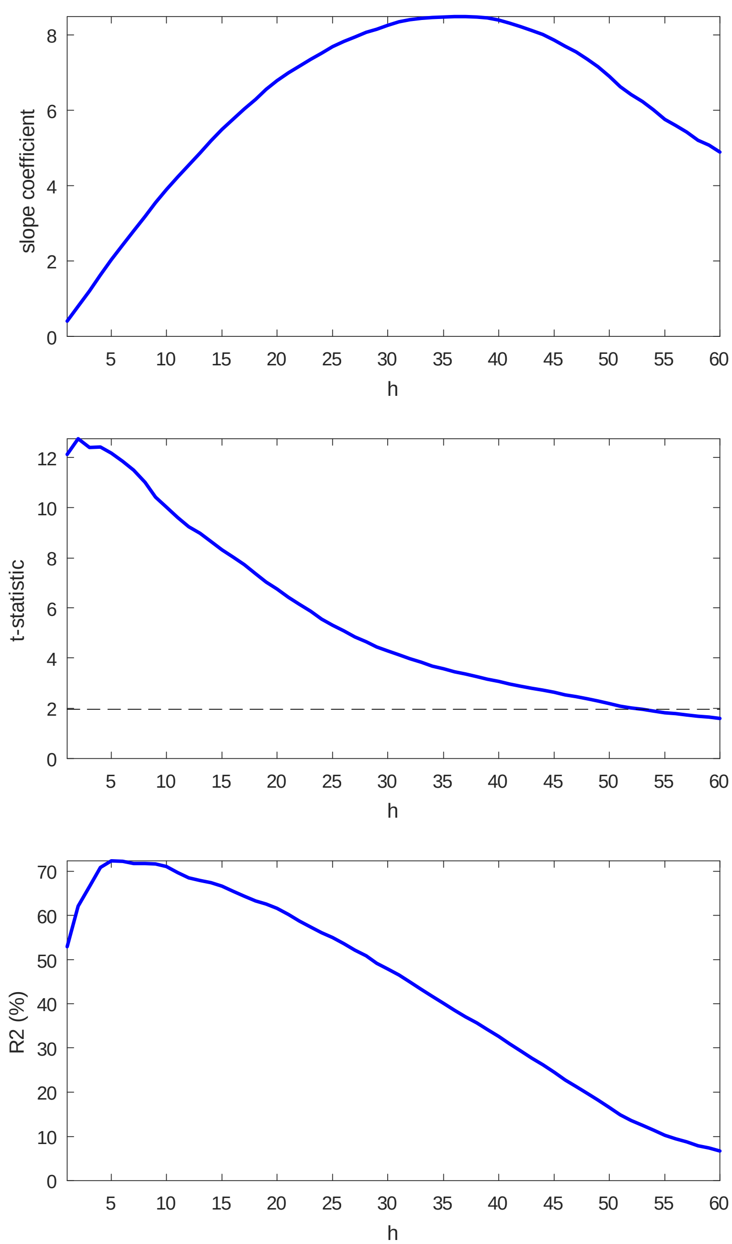

Figure 5. Out-of-Sample Forecast Errors. The figure shows the cumulative sum of squared forecast errors of the no-predictability benchmark minus the cumulative sum of squared forecast errors of a given predictor variable. The forecast horizon is $h = 1$ month and the out-of-sample period runs from 2007:1 to 2019:9.

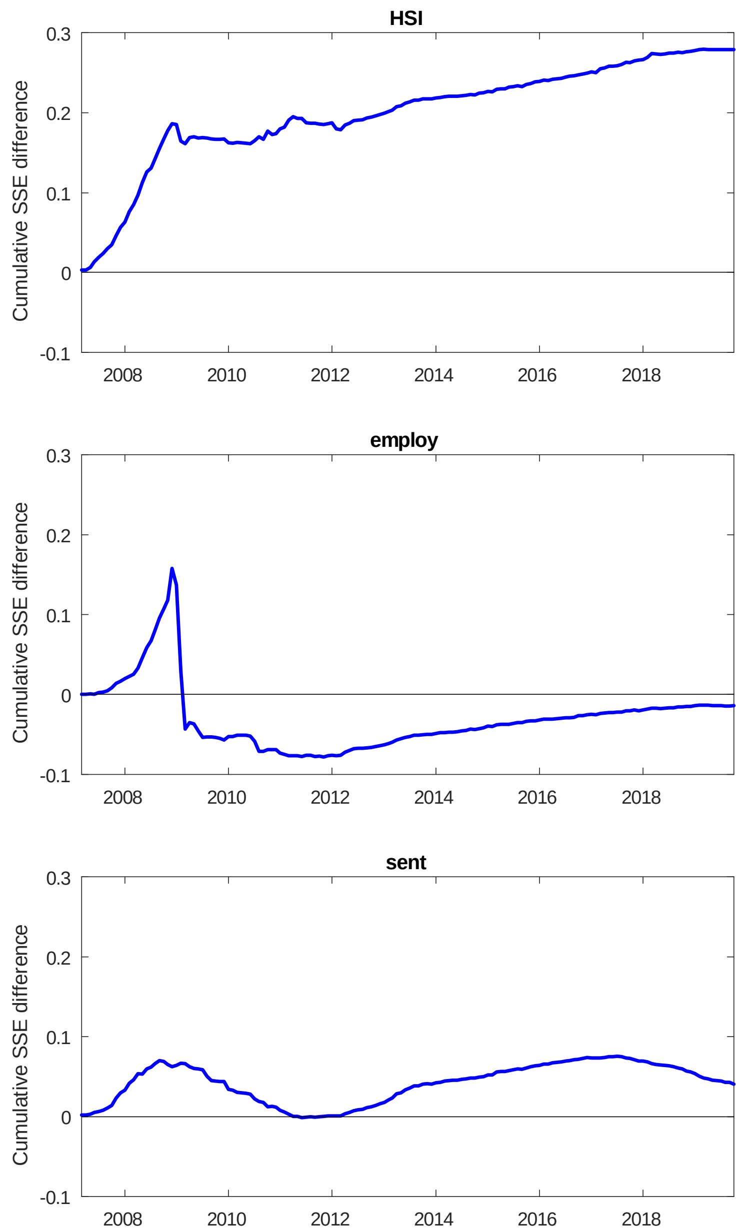

Figure 6. Local Predictive Power of *HSI*. We run MSA-by-MSA forecasting regressions, $p_{it+1} - p_{it} = \alpha_i + \beta_i HSI_{it} + \varepsilon_{it+1}$, and plot the estimate of the predictive coefficient β_i against the R_i^2 across MSAs. The HSI is standardized and all β_i estimates are annualized. The ten most supply-constrained MSAs, cf. Saiz (2010), are shown in red and the ten least supply-constrained MSAs are shown in green.

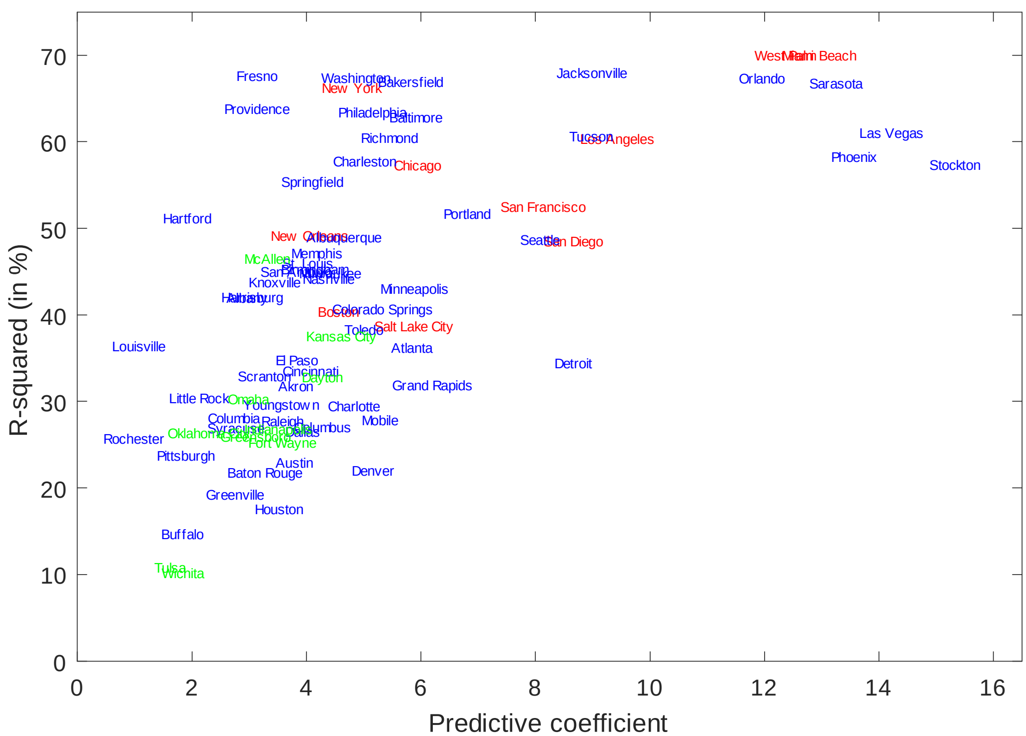

Figure 7. Housing Speculation and the Predictive Ability of $\boldsymbol{HSI}$. We run MSA-by-MSA forecasting regressions, $p_{it+1} - p_{it} = \alpha_i + \beta_i HSI_{it} + \varepsilon_{it+1}$, and plot the estimate of the predictive coefficient β_i against the fraction of non-owner-occupied home purchases across MSAs. The HSI is standardized and all β_i estimates are annualized.

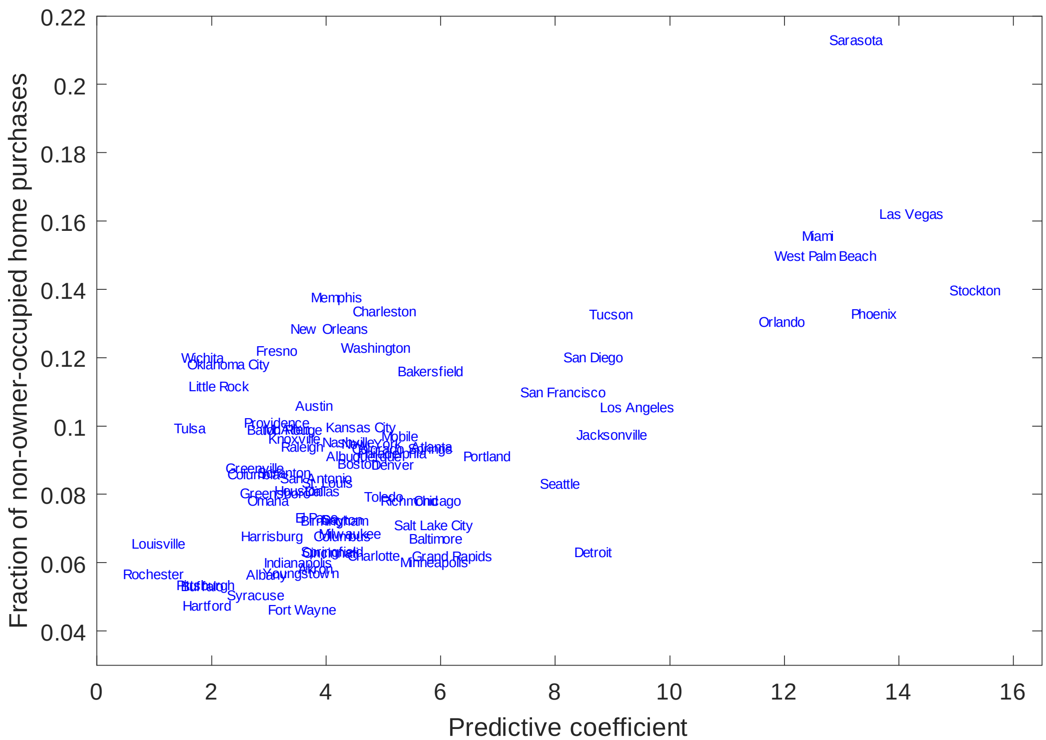

Figure 8. Economic Value. Panel A shows 1st quartile, median and 3rd quartile growth rates in house prices after episodes where HSI is one standard deviation above the local mean. Panel B shows the results following events when HSI is one standard deviation below the local mean. The horizons range from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