Online Appendix for

"Search and Predictability of Prices in the Housing Market"

May 2022

This appendix provides additional results and robustness checks of the analysis reported in the main paper. Below we provide a brief description of the robustness checks:

- Section A.1 illustrates robustness to the method of inference.
- Section A.2 reports in-sample results for conventional determinants of outcomes in housing markets.
- Section A.3 reports results from a placebo test.
- Section A.4 reports results from the selling side of market.
- Section A.5 shows out-of-sample results using alternative search indices.
- Section A.6 reports bootstrap results from a "useless" factor test.
- Section A.7 reports prediction results from more advanced machine learning procedures.
- Section A.8 analyzes the relation between HSI and speculative behavior proxies.
- Section A.9 reports results for one-way and two-way clustered standard errors in panel regressions.
- Section A.10 provides details of additional results for the Covid-19 period.
- Section A.11 reports results on the relation between search and price expectations.
- Section A.12 analyzes the REITs market.
- Section A.13 analyzes non-seasonally adjusted data.

- Section A.14 provides additional results on the predictive ability of local search.
- Section A.15 provides a description of MSAs sorted according to the degree of predictability by HSI.

A.1 Inference

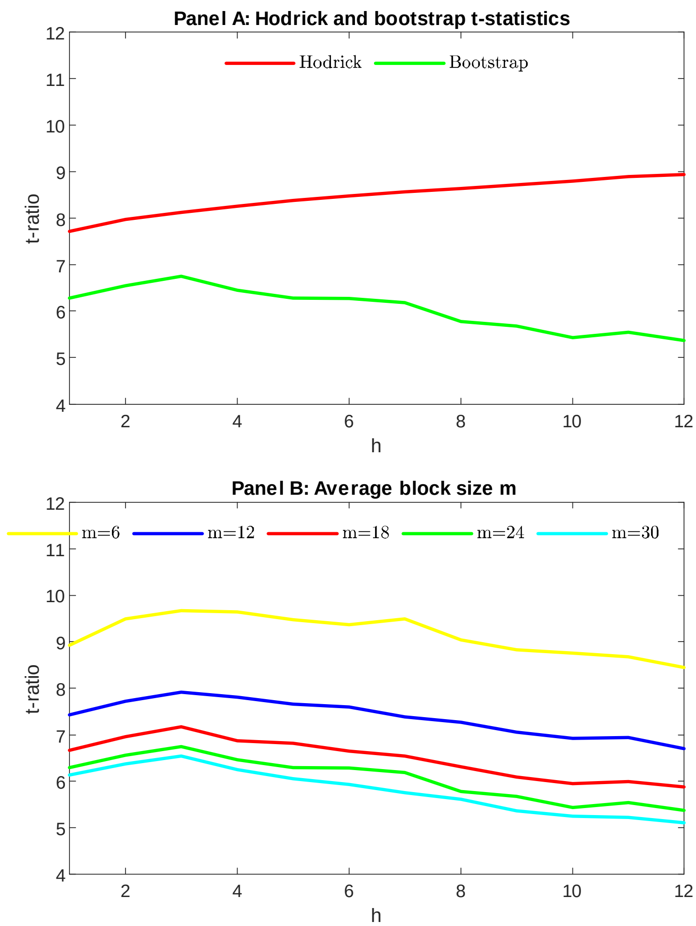
In the main paper, we use a circular block bootstrap to conduct inference when estimating overlapping multi-period ahead forecast regressions of the form, $p_{t+h} - p_t = \alpha + \beta' x_t + \varepsilon_{t+h}$, where p_t is the log of the FHFA house price index, x_t is a vector of predictors, and h is the forecast horizon in months. We resample the regressand and the regressor jointly in blocks with an average block size of $m = 24$, which is motivated by the optimal block length procedure of Politis and White (2001). We use 20,000 replications.

In the following, we analyze the robustness of the method used for inference. First, in Panel A of Figure A.1, we compare block bootstrap t-statistics with those of the Hodrick (1992) procedure, which is aimed at circumventing issues with overlapping data.[1] We use HSI as the predictor and consider horizons from one through 12 months. We find that the Hodrick t-statistics confirm the statistical significance of HSI across horizons, but that these t-statistics are less conservative compared to the block bootstrap t-statistics.

Second, in Panel B of Figure A.1, we consider various choices of the average block length: $m = 6, 12, 18, 24, 30$. Following Politis and White's (2001) automatic selection procedure, we set the average block size to $m = 24$ when generating the results reported in the main paper. The statistical significance of HSI is robust towards other reasonable choices of m as shown in Panel B of Figure A.1. From the figure, we see that using $m = 18$ or $m = 30$ gives bootstrap t-statistics close to those obtained with $m = 24$, whereas $m = 12$ and especially $m = 6$ generates somewhat higher bootstrap t-statistics. In any case, HSI remains statistically significant across the various specifications.

[1] Hodrick's (1992) procedure is based on reverse regressions. Wei and Wright (2013) show that the reverse regression methodology can help in circumventing size distortions in long-horizon forecasting regressions with near-unit roots.

Figure A.1. Robustness to the method of inference.



A.2 Commonly Used Housing Market Determinants

In the main paper, Table 3 provides in-sample evidence that HSI retains its predictive ability when controlling for a long list of predictive variables. Table A.1 shows predictive results from each of the 14 control variables. Among all variables, HSI generates by far the strongest predictive results in terms of statistical significance as well as R^2 values.

Table A.1. Predicting House Prices With 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 forecast horizon in months. For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the R^2 in square brackets. We compute standard errors using a circular block bootstrap. All predictive variables are standardized and slope coefficients are multiplied by 100 to facilitate comparison across variables. The sample period is 2004:1 to 2021:1.

	$h=1$	$h=3$	$h=6$	$h=12$
HSI	0.43	1.26	2.44	4.37
	(6.39)	(6.74)	(6.19)	(5.35)
	[56.91]	[67.37]	[70.41]	[64.42]
payrolls	0.03	0.06	0.32	2.72
	(0.12)	(0.08)	(0.23)	(1.59)
	[0.25]	[0.15]	[1.22]	[24.98]
infl	0.00	−0.12	−0.25	−0.34
	(0.03)	(−0.82)	(−0.97)	(−0.69)
	[0.00]	[0.58]	[0.77]	[0.40]
permits	0.17	0.38	0.62	1.34
	(4.13)	(3.01)	(2.73)	(2.75)
	[9.27]	[6.05]	[4.61]	[6.07]
starts	0.08	0.23	0.38	0.85
	(2.19)	(2.49)	(2.22)	(2.41)
	[2.06]	[2.16]	[1.74]	[2.42]
term	−0.06	−0.10	−0.04	0.59
	(−0.51)	(−0.27)	(−0.05)	(0.35)
	[1.17]	[0.42]	[0.02]	[1.17]
mort	−0.24	−0.72	−1.44	−2.95
	(−1.89)	(−1.95)	(−2.03)	(−2.16)
	[18.00]	[21.82]	[24.55]	[29.37]
pr	0.04	0.03	−0.16	−1.09
	(0.28)	(0.08)	(−0.19)	(−0.67)
	[0.45]	[0.05]	[0.30]	[3.99]
loans	−0.07	−0.00	0.36	0.34
	(−0.62)	(−0.02)	(0.62)	(0.22)
	[1.36]	[0.00]	[1.56]	[0.40]
sent	0.12	0.35	0.82	2.36
	(1.28)	(1.22)	(1.36)	(2.28)
	[4.65]	[5.36]	[7.97]	[18.77]
cfnai	0.15	0.31	0.31	1.78
	(0.77)	(0.49)	(0.26)	(2.53)
	[6.70]	[4.15]	[1.13]	[10.69]
ads	0.13	0.41	0.16	1.64
	(0.53)	(1.50)	(0.50)	(1.75)
	[4.90]	[3.96]	[0.59]	[9.10]
pd	0.20	0.47	0.76	1.27
	(2.16)	(1.78)	(1.61)	(1.34)
	[11.98]	[9.43]	[6.78]	[5.43]
ra	−0.22	−0.48	−0.85	−1.75
	(−2.68)	(−2.46)	(−1.97)	(−2.73)
	[15.36]	[9.75]	[8.60]	[10.37]
unc	−0.19	−0.46	−0.75	−1.64
	(−2.11)	(−1.90)	(−1.58)	(−1.95)
	[11.18]	[8.88]	[6.67]	[9.11]

A.3 Placebo Tests

In the main paper, we analyze the relation between HSI and a large set of commonly used housing market determinants by estimating the contemporaneous regression, $HSI_t = \alpha + x_t'\beta + \varepsilon_t$, where x_t contains a list of 14 variables. Using the full list of housing market determinants leads to an R^2 of about 70%, meaning that about 30% of the variation in HSI is left unexplained. Some of the time-series variation in these 14 variables might be correlated with HSI by chance. We analyze this possibility by running a placebo test that generates artificial times series by resampling from the panel of regressors. We use a circular block bootstrap to resample the panel of housing market determinants and use 20,000 replications. For each replication, we regress the non-resampled HSI on the randomly resampled panel of regressors and save the R^2. We set the average block size to $m = 6, 12, 18, 24$, and 30. Table A.2 reports the median R^2 across the different block sizes. We see that the placebo regressors generate a median R^2 in the range from 19.4% to 30.5% across block sizes, which suggests that the estimated $R^2 = 0.70$ obtained from regressing HSI on all 14 variables is inflated. There is therefore a risk of excluding valuable information when orthogonalizing HSI against all 14 housing market determinants.

Table A.2. Results of placebo test. We use a circular block bootstrap to resample the panel of control variables. For each replication, we regress the non-resampled HSI on the resampled panel of variables and save the R^2 from this regression. The table reports the median R^2 value across different block sizes.

	$m = 6$	$m = 12$	$m = 18$	$m = 24$	$m = 30$
R^2	19.4%	25.0%	27.8%	29.2%	30.5%

A.4 Buying versus Selling Side of the Housing Market

Wu and Brynjolfsson (2015) conjecture that house prices are difficult to predict using two predefined real estate categories since these categories can 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 index based on these search terms by HSI^{sell}.

Panel A in Table A.3 shows that this search index based on the selling side of the housing market holds limited predictive power over future house prices. The slope coefficients are only significantly different from zero for $h = 1$ and the R^2-values never exceed 8%. 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. However, when we control for HSI, the slope coefficients on HSI^{sell} turn negative except for $h = 12$ as shown in Panel B of Table A.3.

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.

Table A.3. 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 predictors, and h is the forecast horizon in months. For each regression, the table reports slope estimates, the corresponding t-statistics in parenthesis, and the R^2 in square brackets. We compute standard errors using a circular block bootstrap. All predictors are standardized to facilitate comparison of the β estimates and the log price change is multiplied by 100. Panel A shows the results for HSI^{sell}, which is an alternative search index based on the selling side of the housing market, and Panel B includes HSI and HSI^{sell} in a joint regression. The sample period is 2004:1 to 2021:1.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: Search index based on the selling side of the housing market				
HSI^{sell}	0.11	0.34	0.68	1.51
	(0.79)	(0.82)	(0.81)	(0.93)
	[3.68]	[4.97]	[5.43]	[7.72]
Panel B: HSI joint with the selling side of the housing market				
HSI	0.46	1.31	2.48	4.27
	(6.35)	(6.64)	(6.09)	(5.02)
HSI^{sell}	−0.07	−0.15	−0.15	0.33
	(−1.04)	(−0.75)	(−0.38)	(0.42)
	[58.05]	[68.22]	[70.64]	[64.76]

A.5 Out-of-Sample Results With Alternative Search Indices

Rather than using specific keywords, Wu and Brynjolfsson (2015) employ two predefined search categories supplied by Google Trends, namely "Real estate agencies" and "Real estate listings". In the main paper, we show that HSI contains stronger in-sample predictive power than these two alternative search indices. As a further robustness check, we analyze the out-of-sample evidence. In Table A.4, we compare the out-of-sample predictive power of HSI against the two predefined search categories. Panel A reports results for HSI, while Panel B reports results from the predefined search categories. Both "Real estate agencies" and "Real estate listings" generate negative out-of-sample R^2s when used on their own as well as in a joint specification. In contrast, HSI is able to outperform the historical mean benchmark. Thus, the out-of-sample evidence confirms the results obtained from the in-sample analysis.

Table A.4. Out-of-Sample Tests With Alternative Search Indices. The table reports the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OoS}) and in parenthesis the p-value from the Diebold and Mariano (1995) t-statistic, computed using the Newey and West (1987) estimator with h lags, where h is the forecast horizon in months. The null hypothesis is that the R^2_{OoS} is equal to zero or negative and the alternative hypothesis is that it is positive. Panel A reports results for HSI. Panel B reports results from the predefined search categories used by Wu and Brynjolfsson (2015).

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A				
HSI	51.37	64.54	63.91	54.25
	(0.00)	(0.00)	(0.00)	(0.02)
Panel B				
Real estate agencies	−1.08	−4.45	−6.68	−17.85
	(0.57)	(0.68)	(0.70)	(0.82)
Real estate listings	−4.36	−7.63	−12.21	−20.89
	(0.90)	(0.90)	(0.89)	(0.83)
Joint	−0.79	−4.04	−5.15	−12.97
	(0.56)	(0.68)	(0.67)	(0.71)

A.6 Bootstrap Analysis

To further validate the statistical significance of 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 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 HSI and generate out-of-sample forecasts, then save the R^2_{OoS} statistic. Because the resampled placebo Google search data should bear no relation to the realized house price growth rates, HSI should not be useful in forecasting growth in house prices. Basically, the resampled search indices represent random noise and so are "useless".

Table A.5. Bootstrapped p-values. Panel A reports the Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OoS}). In Panel B, we generate 10,000 bootstrap samples of the search indices used to compute HSI. For each bootstrap sample, we recursively estimate HSI based on 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 fraction 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 R^2				
HSI	51.37	64.54	63.91	54.25
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

We analyze the empirical distribution of the R^2_{OoS} statistic by computing empirical p-values. The simulations in Panel B in Table A.5 show that the fraction 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 procedure 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 additional tests corroborate the robustness of our findings on the highly significant out-of-sample predictive power of HSI over future movements in house prices.

A.7 Machine Learning

We construct HSI using a simple targeted PCA approach. However, it may be possible to achieve more accurate forecasts using more advanced machine learning techniques. To test this possibility, we perform an out-of-sample forecasting exercise using three popular machine learning techniques that have been shown to achieve strong predictive performance even on difficult forecasting problems such as asset risk premiums (Gu et al., 2020). The three methods we consider are Random Forest (Breiman, 2001), Gradient Boosted Trees (GBT)[2], and an artificial neural network (ANN) with a single hidden layer.[3] All of these methods incorporate the possibility of complex non-parametric nonlinear relations between the predictors and house prices but also run the risk of overfitting the data. To avoid this problem, we follow the most common approach in the literature and select *tuning* parameters adaptively from the data using a recursive cross-validation scheme. Subsection A.7.1 provides further details on our cross validation sample splitting scheme, choice of hyper-parameters and choice of neural network architectures. As in the main paper, we use the first three years as our initial estimation period and reserve the remaining sample for out-of-sample evaluation.

Machine learning methods are inherently designed to take advantage of high-dimensional data so it is possible that they require a larger set of predictors to achieve their forecasting potential. For this reason, when utilizing these methods we consider an extended set of predictors where we add the top 25 related search terms for each of our original set of 23 predictors used to form HSI. After removing duplicates, this expanded set contains a total of 292 Google Trends (GT) terms. We note that, although it would be possible to construct a factor like HSI using this expanded set of predictors, this factor would not have the simple demand search interpretation as HSI since many of the additional terms are not necessarily reflecting housing demand. Thus, our high-dimensional panel does not offer the same interpretability as the smaller data used to extract HSI from.[4]

Table A.6 reports the Campbell and Thompson (2008) out-of-sample R^2 values (R^2_{OoS}) and Diebold and Mariano (1995) t-statistics (t_{DM}) against the historical average. Overall, the forecasting per-

[2] In particular, we consider the powerful XGboost algorithm of Chen and Guestrin (2016), which builds on the non-parametric procedures developed by Breiman (1997) and Friedman (2001).

[3] In unreported results, we also considered a a deep artificial neural network with three hidden layers, but this ANN underperformed the network with a single layer.

[4] For example, the expanded set of predictor terms include terms like *"average cost of building a house"* and *"real estate news"*, which can contain predictive information for housing returns but are not necessarily reflecting demand.

formance of these methods using a high-dimensional panel is similar to what we obtained for HSI. For example, the GBT model which is the best performing machine learning approach, obtains an average R^2_{OoS} across horizons of 58.6%, which is only marginally above that obtained using HSI (58.5%).

The fact that an expanded set of GT predictors in combination with models that allow for non-linearities does not result in significantly improved predictive performance implies that most of the information embedded in Google search data is likely to be already captured by HSI and that the search terms are probably characterized by a strong factor structure. The exception to this rule appears to be forecasts at the one-month horizon, where the ML methods outperform HSI.

The panel we use for this forecasting experiment is based on terms that are somehow related to the original set of 23 GT terms used to form HSI. If the purpose is to obtain the best possible forecast, it is possible that a larger set of terms with broader coverage, in combination with ML methods, will result in better forecasting performance. Such an exercise would, however, result in even more complexity when trying to interpret the information captured by the forecasting model.

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

	$h=1$	$h=3$	$h=6$	$h=12$
RF	60.80	60.47	52.93	35.55
	(0.00)	(0.00)	(0.00)	(0.04)
GBT	58.85	64.83	60.55	50.29
	(0.00)	(0.00)	(0.00)	(0.00)
ANN	53.18	56.26	61.80	49.15
	(0.00)	(0.00)	(0.00)	(0.00)

A.7.1 Machine Learning Model Specifications

All models are implemented in `Python` with the choice of hyperparameters shown below. We use the recursive *walk forward* cross validation scheme implemented in the `TimeSeriesSplit` function of the `sklearn` package. We use 12 validation periods (splits) and set the gap equal to the forecast horizon. Using K-fold cross validation with $K = 5$ folds leads to similar results.

Gradient Boosted Trees The number of trees considered are $B = 500$ with a shrinkage parameter (or learning rate) set to $\tau = 0.1$. We validate the depth of the trees using the following grid of values: maximum tree depth $(L) \in [1, 2, 3, 5]$, the percentage of observations used in each tree in: $[60\%, 80\%, 100\%]$, the percentage of predictors used in each tree in: $[60\%, 80\%, 100\%]$, and the across-trees regularization parameter $\gamma \in [0, 0.1]$. The implementation is conducted with the `XGBRegressor` from the `XGboost` package. All other settings are set to the default of the package.

Random Forest The number of trees considered is $B = 500$, and the trees are allowed to grow fully deep (L can be large) consistent with Breiman (2001). With the number of predictors given by $p = 292$, we validate the number of predictors to be considered at each split using the following grid: $[p, 2/3p, 1/3p, \sqrt{p}, \log_2 p]$. The implementation is conducted with the `RandomForestRegressor` from the `sklearn` package. All other settings are set to the default of the package.

Artificial Neural Network We use a batch size of 12 in the stochastic gradient descent optimizer, set the learning rate to a (by default) constant value of 0.01, and use 1,000 epochs. The activation function is the rectified linear activation function (ReLU). We use the popular square root rule to select the number of neurons, hence with 292 predictors, the hidden layer has $\lfloor\sqrt{292}\rfloor = 17$ neurons. We cross-validate a single parameter which is the penalty parameter for the ℓ_2 regularization using the following grid of values: $[0.01, 0.001, 0.0001]$. The implementation is conducted using `MLPRegressor` from the `sklearn` package, using default settings for the ADAM stochastic gradient descent optimizer. The results are robust to changes in the number of epochs and learning rate.

A.8 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, 2018).[5] Given that HSI is a direct measure of peoples' intention to buy a

[5] Other factors contributing 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).

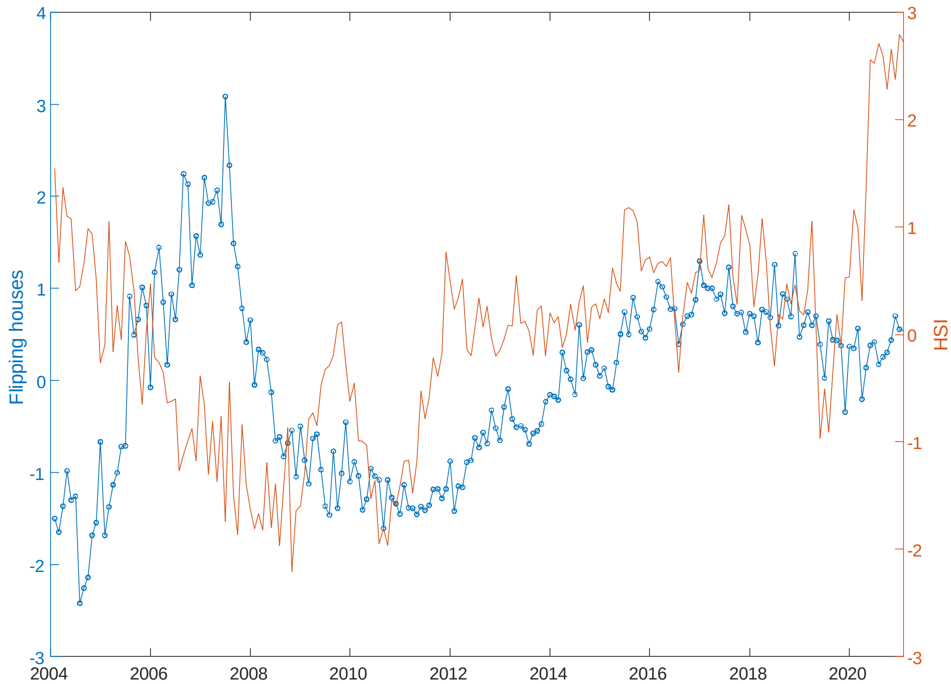
house and hence captures the demand side of the market, we may expect that the predictive power of HSI reflects both fundamental and non-fundamental sources of demand for housing.

As an attempt to analyze to what extent HSI captures speculation-driven demand, we construct a search-based measure of the interest in "flipping houses", which is likely to reflect speculative motives, and examine its relation to HSI.[6] We construct a "flipping houses" search index by using “flipping houses” as our main search term and then obtain the following list of related terms: "flipping homes", "flip houses", "house flipping", "house flip", "flip the house", "how to flip houses", and "real estate flipping". Next, we construct a house flipping search index by extracting the first principal component of these search indices.

We start by plotting HSI together with the house flipping index in Figure A.2. From the figure, we see that the house flipping index increases substantially in the mid 2000s and reaches its highest value at the beginning of 2007, in line with the literature that has shown that speculation in the housing market was an important driver of the boom in the early 2000s. HSI also takes on high values in the 2004-2005 period, but does not show strong comovement with the flipping houses index. Furthermore, while HSI increases substantially during the Covid-19 pandemic, the house flipping index shows a more modest increase. Over the full sample, the correlation between the two series is only 0.10. Thus, when orthogonalizing HSI with respect to the flipping houses index, we find that it retains basically all of its predictive ability. These results indicate that the predictive ability of HSI is not driven by speculative activity.

[6] Recent papers on house flipping include Goldstein (2018) and Bayer et al. (2020).

Figure A.2. Flipping Houses. The figure plots the flipping houses index together with the housing search index (HSI). The sample period is 2004:1-2021:1.



As another measure of speculation, we follow Gao et al. (2020) and compute the fraction of non-owner-occupied home purchases using the Home Mortgage Disclosure Act (HMDA) data from which it is possible 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 by decisions to buy a primary home. The HDMA data is only available annually, but we can still compare general movements in the housing speculation measure of Gao et al. (2020) with those of HSI. For the aggregate U.S. housing market, Gao et al. (2020) find that the degree of speculation peaked in 2005 with a share of non-owner occupied home purchases of about 15%. This is also the period where the house pricing boom was close to reaching its peak. Interestingly, the fraction of investment properties purchased is now substantially lower than it was during the first part of the 2000s.

Across the 77 MSAs in our sample, the median value of the share of investment properties purchased was 6.8% in 2020, a drop from 7.9% in 2019 and 8.8% in 2018. This relatively low level of housing speculation during the pandemic holds across the 77 MSAs in our sample. For example, Gao et al.

(2020) show that the share of non-owner occupied home purchases was almost 30% in Las Vegas in 2005. We find that this number has dropped to about 9% in 2020. These strong movements in speculation stand in stark contrast to the time-series movements in HSI, which reached an all-time high during the pandemic.

In conclusion, these results suggest that HSI captures general movements in demand and is not dominated by speculative activity.

A.9 Clustered Standard Errors

We use two-way clustered t-statistics when estimating fixed-effects panel regressions in the main paper. In Table A.7, we show results from both one-way and two-way clustered t-statistics (clustered by time and MSA).

Table A.7. Predicting Local House Prices With Local Housing Search: Evidence From Panel Regressions. The table reports results from fixed effects panel regressions of the form, $p_{it+h} - p_{it} = \alpha_i + \beta HSI_{it} + \varepsilon_{it+h}$, 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, and h is the forecast horizon in months. For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the within R^2 in square brackets. We compute standard errors using Thompson (2011) clustered robust-statistics with h lags. The table shows results using both single-clustered and double-clustered t-statistics. HSI is standardized to facilitate interpretation of the β estimates. The sample period is 2004:1 to 2021:1.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
β	0.42	1.22	2.32	4.19
t (cluster by time)	(13.26)	(9.12)	(6.98)	(5.98)
t (cluster by MSA)	(12.88)	(12.56)	(12.16)	(11.93)
t (cluster by both time and MSA)	(9.43)	(7.59)	(6.26)	(5.58)
R^2	[37.20]	[37.89]	[36.89]	[33.81]

A.10 Additional Results on Covid-19 Period

In the main paper, we reports results from the following panel regression

$$p_{it+1} - p_{it} = \alpha_i + \beta_D HSI_{it} + \beta_S S_{it} + \gamma' Z_{it} + \varepsilon_{it+1}, \tag{A1}$$

where $p_{it+1} - p_{it}$ is the one-month change in the log house price index for MSA i in month $t + 1$, HSI_{it} is our housing search index, S_{it} is a proxy for the housing supply given by the Zillow for-sale-inventory of houses, and Z_{it} is a vector of controls including the Covid-19 stringency index of Hale et al. (2021) and the number of Covid-19 cases, all measured for MSA i in month t. To visualize the model fit, Figure A.3 plots realized month-on-month house price growth rates during the Covid-19 pandemic together with predicted values from the full specification in (A1). We focus on the four largest MSAs as measured by population. Across MSAs, growth in house prices were in general low (even negative) in March and April of 2020 but bounced back sharply between May and August and remained high subsequently. The figure illustrates that demand and supply effects along with Covid-19 restrictions combine to capture a substantial part of the variation in house prices across MSAs during the pandemic. If we only use supply as regressor, the model fit deteriorates, but once we account for HSI the predicted values are much closer to the actual values.

To give a sense of how search activity changed during the Covid-19 period in the various MSAs, Figure A.4 shows scatter plots of one-month-ahead price changes ($h = 1$) against search, marking Covid-19 points in red and pre-Covid-19 points in blue. The strong positive relation between one-month lagged search and house price changes is present both in the pre-Covid-19 period and during the pandemic. The relation remains remarkably stable both for MSAs in which the pandemic triggered the highest search activity and price growth observed in the entire sample (New York and Dallas) as well as for MSAs with a more modest pandemic impact (Los Angeles and Chicago).

While our data on housing supply from Zillow does not allow us to go as far back in time as HSI data, it allows us to compare the period from November 2017 to January 2021. Figure A.5 shows that the pandemic was associated with a notably lower housing supply across the top-four MSAs with particularly strong reductions for New York and Chicago. Notice also how the housing supply constraint gradually tightens from month to month during 2020 – a pattern quite different from that seen in HSI during the pandemic.

To further analyze the effect of HSI during the pandemic period, we estimate the panel regression model, $p_{it+1} - p_{it} = \alpha_i + \beta HSI_{it} + \varepsilon_{it+1}$, using both the pre-pandemic period and the pandemic period. Table A.8 shows the results. We find that the effect of HSI is relatively stable across sample periods.[7]

[7] We also analyzed the relation between search activity and time-on-market across pre- and post-pandemic periods and find that the effect of HSI is magnified for the covid-19 period. However, our data on time-on-market cover a relatively short time period.

Figure A.3. House Price Movements During the Covid-19 Pandemic. The figure plots realized month-on-month house price growth rates during the Covid-19 pandemic together with predicted values from equation (A1). As regressors, we use either supply on its own, demand (HSI) on its own, or the full specification equation (A1). The figure shows results for the four largest MSAs as measured by population.

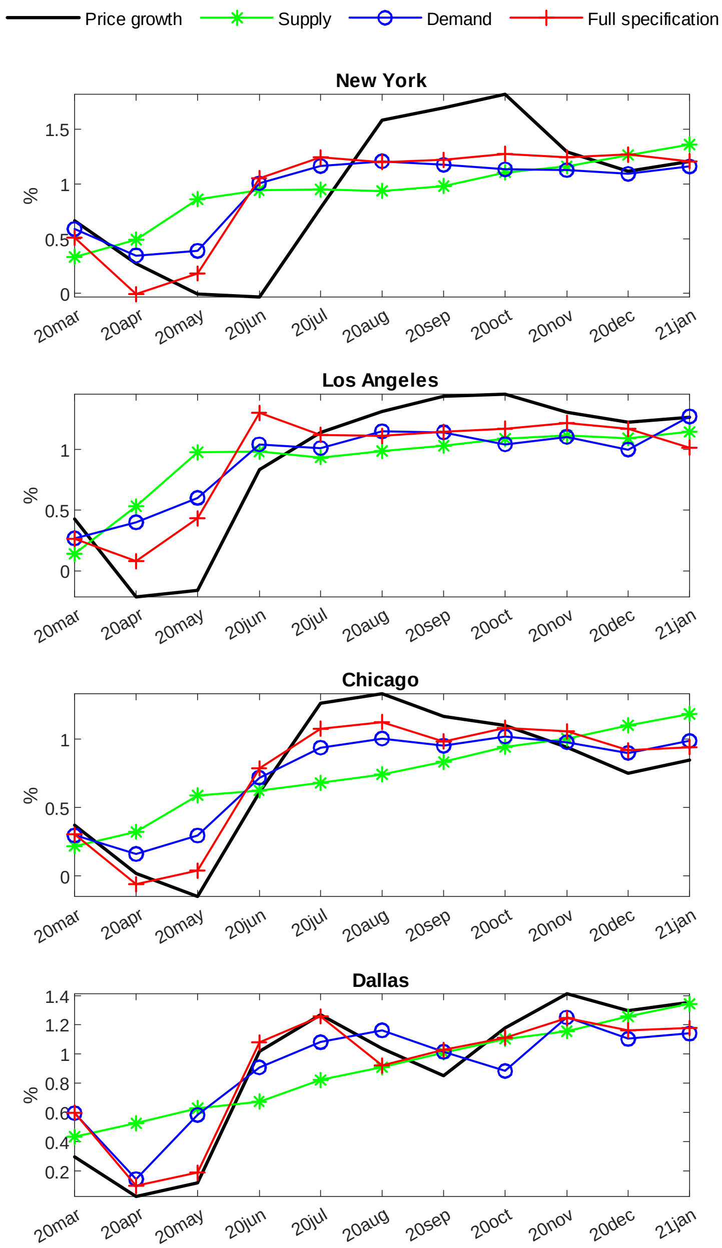


Figure A.4. Price Changes vs. Search. The figure plots realized house price growth rates (in percent) against one-month lagged values of HSI (standardized). Covid-19 points are marked in red with pre-Covid-19 points marked in blue. The figure shows results for the four largest MSAs as measured by population.

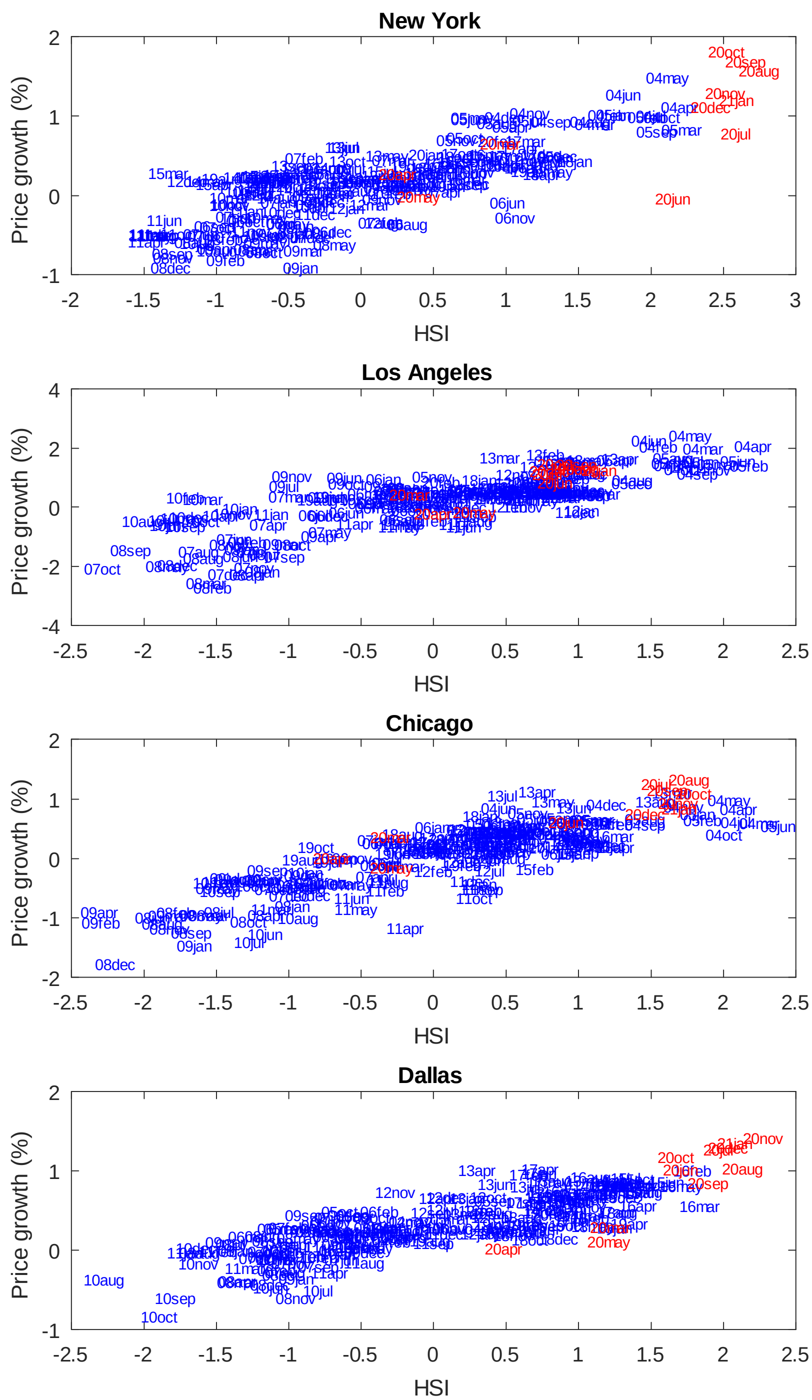


Figure A.5. Price Changes vs. Supply. The figure plots realized house price growth rates (in percent) against one-month lagged values of supply as measured by the inventory of houses for sale (standardized). Covid-19 points are marked in red with pre-Covid-19 points marked in blue. The figure shows results for the four largest MSAs as measured by population.

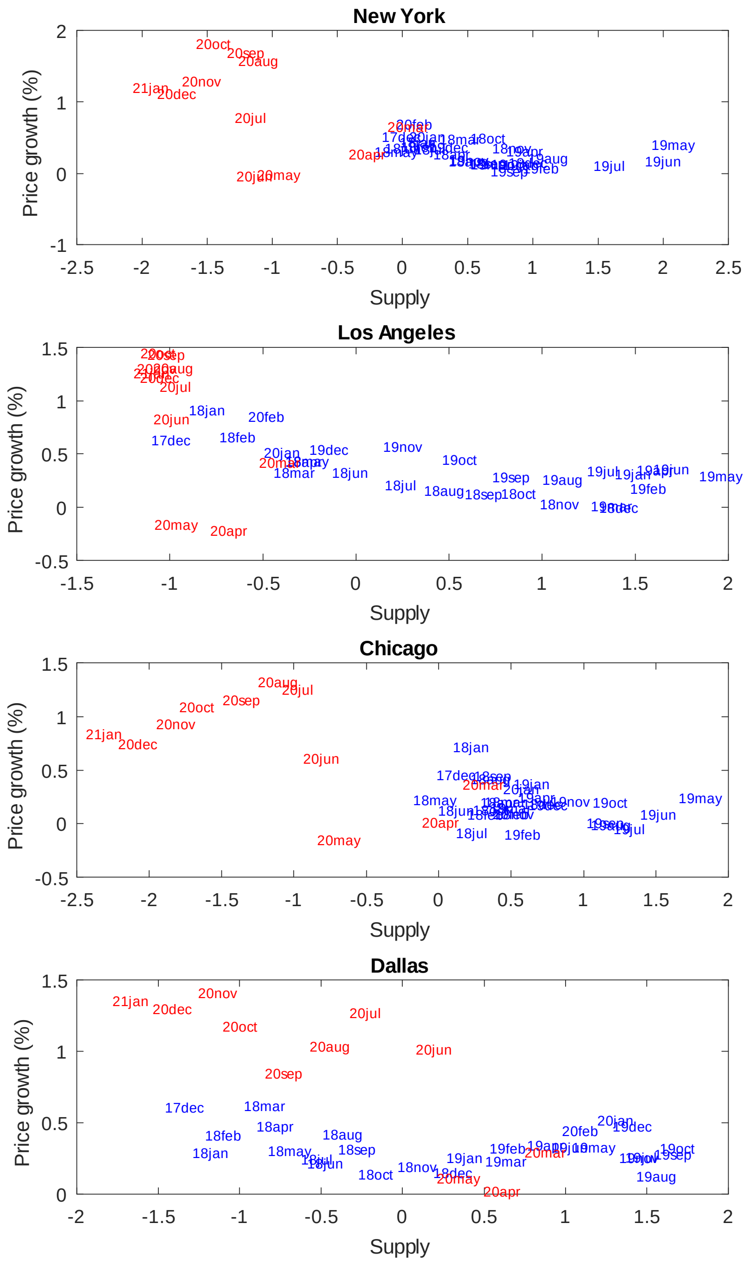


Table A.8. Predicting Local House Prices With Local Housing Search: Subsample Analysis. The table reports results from fixed effects panel regressions of the form, $p_{it+1} - p_{it} = \alpha_i + \beta HSI_{it} + \varepsilon_{it+1}$, where p_{it} is the log of the Freddie-Mac house price index in MSA i, and HSI_{it} is the housing search index in MSA i. 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. HSI is standardized to facilitate interpretation of the β estimates.

	Pre-covid-19 period	Covid-19 period
HSI	0.38	0.32
	(8.32)	(9.65)
	[33.34]	[37.02]

A.11 Search and Price Expectations

Our estimates of changes in house prices based on search activity are model-based and so may be subject to model misspecification biases. 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 to 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 0.56 (Northeast) to 0.78 (North Central). 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.

A.12 REITs

Prices of residential real estate investment trusts (REITs) provide a direct market-based view of real estate conditions. Unlike residential house price indices such as the purchase-only FHFA index, it is possible for individuals to invest and trade in REITs. Because REIT prices should reflect investor expectations about future fundamentals, they may potentially also contain relevant forward looking information about future FHFA house price growth rates.

To analyze this possibility, we obtain information about publicly traded residential REITs from the S&P Global SNL Real Estate Database and merge it with stock return data obtained from CRSP. The residential REITs are those that have primary property type specified as "Multifamily Apartments." We use data on 21 residential REITs that have been traded during our sample period from 2004:1 to 2020:12. Seven out of these are traded during the whole period, while the rest have missing observations in certain parts of the sample. To get an aggregate measure of the residential REIT market, we compute an equal-weighted average of the firm-specific REIT returns. Next, we estimate predictive regressions of the form, $p_{t+h} - p_t = \alpha + \beta R_t^{REIT} + \varepsilon_{t+h}$, where p_t is the log of the FHFA house price index, R_t^{REIT} is the average return on the residential REIT market, and h is the forecast horizon in months.

In Panel A of Table A.9, we see that there is a positive relation between R_t^{REIT} and future growth rates on the FHFA index, but it is not statistically significant according to the bootstrap-based t-statistics. These results imply that although REIT returns represent a marked-based view of real estate conditions, they are not useful in forecasting residential house prices.

Table A.9. REITs. Panel A reports results from predictive regressions of h-step-ahead log growth rates on the FHFA house price index using residential REIT returns as predictors. Panel B reports results from predictive regressions of h-step-ahead log excess returns on residential REITs using HSI as the predictor. For each regression, the table reports the estimate of β, the corresponding t-statistic in parenthesis, and the R^2 in square brackets. We compute standard errors using a circular block bootstrap. All predictive variables are standardized and slope coefficients are multiplied by 100. The sample period is 2004:1-2020:12.

$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A			
0.11	0.17	0.28	0.52
(1.88)	(1.18)	(0.88)	(0.78)
[3.67]	[1.17]	[0.96]	[0.92]
Panel B			
0.77	2.19	4.53	9.42
(0.98)	(0.95)	(1.18)	(1.37)
[1.23]	[3.18]	[5.51]	[11.23]

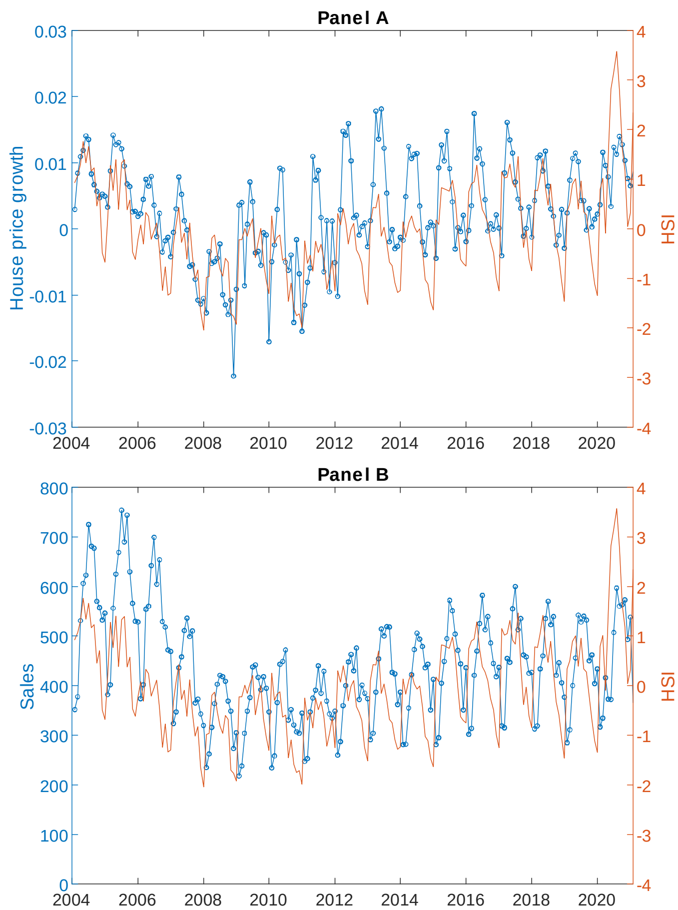
The REIT market also provides an opportunity to analyze the source of predictability emanating from HSI. If HSI captures a time-varying risk premium component due to time-varying risk or risk aversion, we would expect that its predictive power for residential house price indices carries over to the REIT market. In contrast, if the predictive power from HSI stems from search frictions, it should hold no predictive power for REITs since returns on REITs are largely unaffected by such frictions. In Panel B, we show results from forecasting h-step-ahead log excess returns on the residential REIT market using HSI as our predictive variable. We see that HSI has very limited predictive power over REIT returns. The sign on HSI is positive in accordance with the predictive results obtained when forecasting house prices on the residential real estate market. However, HSI is insignificant across all horizons. We view these results as indicative evidence that the predictive power of HSI over future house prices does not arise from a risk compensation channel, but is more likely to reflect sluggish price adjustments in the residential real estate market due to frictions.

A.13 Non-Seasonally Adjusted Data

In the main paper, we analyze seasonally-adjusted houses prices, but it is well-known that house prices contain a strong seasonal component with high prices during spring and summer and low prices during fall and winter (e.g. Ngai and Tenreyro, 2014). Fundamentals such as rental flows, the level of the mortgage rate, or credit variables do not follow such a seasonal cycle. Growth in output and related economic variables typically have a strong seasonal cycle, but with a boom in the fourth quarter (e.g. Barsky and Miron, 1989), which does not match the seasonal pattern on the housing market. In addition, it seems unlikely that sentiment/beliefs about the outlook for the housing market should move up and down each quarter according to the season. In the paper, we measure sentiment as the fraction of respondents who answer that now is a "good time" to buy a house from the University of Michigan's Survey of Consumers (Cox and Ludvigson, 2019). This series does not show any pronounced seasonality. If HSI captures seasonal variation in house prices, it can therefore be taken as evidence that it contains information beyond what is contained in typically used housing market determinants. To check this, we therefore reconstructed HSI based on seasonally unadjusted data using the approach outlined in Section 2.2 of the paper.

In panel A of Figure A.6, we plot the non-seasonally-adjusted search index together with non-seasonally adjusted house price growth. As is evident from the figure, the two series tend to move closely together with a strong seasonal cycle in both search and house prices. Search activity tends to reach its lowest points in November and December, while the peak points vary more from year to year but always occur in the spring or early summer months. In comparison, house prices tend to be lowest in December and often peak in May or June. Panel B illustrates the seasonal variation in sales, which typically peaks in June and reaches its lowest point in January, which also implies that search tends to move before sales.

Figure A.6. Non-Seasonally Adjusted Housing Search Index. Panel A shows the non-seasonally adjusted housing search index (HSI) along with the log growth rate in the non-seasonally adjusted purchase-only FHFA house price index. Panel B shows the non-seasonally adjusted HSI along with the non-seasonally adjusted sales of existing single-family housing units from the National Association of Realtors. The sample period is 2004:1-2021:1.



In Panel A of Table A.10, we use the seasonally unadjusted HSI to predict seasonally unadjusted house price growth rates. We see that the predictability results are strong for $h = 1$ but weaken for $h > 1$. By increasing the forecast horizon, we often predict from one season into another and, in addition, the use of overlapping growth rates removes the extent of seasonality. In Panel B, we therefore use both the non-seasonally and seasonally adjusted HSI as predictive variables. We find that the non-seasonally and seasonally adjusted HSI jointly have strong predictive power for non-seasonally adjusted house prices across all forecast horizons.

Table A.10. Predicting Seasonally Unadjusted House Prices. The table reports results from predictive regressions of the h-step-ahead growth rate in house prices, $p^{NSA}_{t+h} - p^{NSA}_t$, where p^{NSA} is the log of the non-seasonally adjusted FHFA house price index. We use both the seasonally adjusted and unadjusted housing search index as predictors. For each regression, the table reports the slope estimates, the corresponding t-statistics in parenthesis, and the R^2 in square brackets. We compute standard errors using a circular block bootstrap. All predictors are standardized and slope coefficients are multiplied by 100 to facilitate comparison across variables. The sample period is 2004:1-2021:1.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Panel A: Seasonally unadjusted HSI				
NSA HSI	0.52	1.18	1.29	2.48
	(6.12)	(5.47)	(2.81)	(3.31)
	[47.23]	[34.93]	[14.50]	[20.73]
Panel B: Seasonally unadjusted and adjusted HSI				
NSA HSI	0.42	0.61	−0.50	−0.00
	(5.48)	(3.27)	(−3.07)	(−0.00)
SA HSI	0.15	0.85	2.78	4.35
	(1.76)	(3.55)	(6.88)	(5.06)
	[49.51]	[44.97]	[53.88]	[64.23]

A.14 Local Search

There is widespread evidence that housing markets are local in nature and segmented (see, e.g., Del Negro and Otrok, 2007, Gyourko et al., 2013, Glaeser et al., 2014, and Hernández-Murillo et al., 2017). Consistent with this evidence, in Table 8 of the main paper, we show that local housing search stays statistically significant across all forecast horizons when controlling for national-level housing search. To further verify that local search, unrelated to the aggregate search patterns of the economy, plays a significant role in predicting local house price changes, we orthogonalize local search with respect to US-level aggregate search:

$$HSI_{it} = b_i HSI_{USt} + HSI^{Local}_{it} \tag{A2}$$

where HSI^{Local}_{it} is the part of MSA-level search, HSI_{it}, that is unrelated to the aggregate US-level search behavior, HSI_{USt}. Next, we examine whether expected local house price changes have both a local and national component by estimating

$$p_{it+h} - p_{it} = \alpha_i + \beta_{Local} HSI^{Local}_{it} + \beta_{US} HSI_{USt} + \varepsilon_{it+h}, \tag{A3}$$

where β_{Local} measures the impact of the part of local search that is orthogonal to aggregate national-level search behavior. Table A.11 shows the results of estimating ($A3$) on our sample of MSAs. Across horizons, we see that both national-level search and the part of local search that is unrelated to aggregate search patterns are strongly statistically significant. These results confirm that local-specific search, orthogonal to aggregate search patterns of the economy, plays a significant role in the prediction of local house price changes.

Table A.11. Predicting Local House Prices With Local Search Orthogonal to US search: Evidence From Panel Regressions. The table reports results from fixed effects panel regressions of the form, $p_{it+h} - p_{it} = \alpha_i + \beta_{Local} HSI_{it}^{Local} + \beta_{US} HSI_{USt} + \varepsilon_{it+h}$, 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}^{Local} is the part of MSA-level search that is orthogonal to HSI_{USt}, and h is the forecast horizon in months. We compute standard errors using Thompson (2011) two-way clustered robust-statistics with h lags. HSI is standardized to facilitate interpretation of the β estimates. The sample period is 2004:1 to 2021:1.

	$h = 1$	$h = 3$	$h = 6$	$h = 12$
U.S. HSI	0.43	1.28	2.44	4.38
	(11.80)	(10.36)	(8.51)	(6.37)
Local $HSI^{\perp}$	0.20	0.59	1.13	2.08
	(8.29)	(7.07)	(6.45)	(7.28)
	[48.11]	[49.86]	[49.26]	[45.43]

A.15 Difference in Predictive Power Across MSAs

In the main paper, we document that HSI is a strong predictor across most MSAs but also that the degree of predictive power varies somewhat as measured both by the slope coefficient and the R^2 value. In analyzing the differences in predictive power, we pay special attention to supply elasticity and show that the relation between variation in online search activity and house prices is significantly stronger in MSAs with low elasticity compared to those with high elasticity.

Table A.12 provides further insights with respect to the differences in predictive power across MSAs. In Panel A we sort the MSAs into two groups based on the size of the R^2 values in the MSA-level regressions. For each of the two groups, we report the average R^2 value, supply elasticity, population, income per capita and population density.[8] In addition to low supply elasticity, the MSAs that experience relatively high predictive power of HSI are also characterized by high

[8]We obtain population density for each MSA based on the 2010 Census from the U.S. Census Bureau. Population density is the 'Population-Weighted Density by Distance from City Hall'. For population and income per capita, we use the 2012 estimate from the U.S. Census Bureau.

population and high population density, whereas income per capita is similar across MSAs with high and low predictability. Panel B shows that we obtain the same result if we instead sort based on the size of the slope coefficients in MSA-level regressions.

Table A.12. Differences in Predictive Power Across MSAs. In Panel A, we sort the MSAs into two groups based on the R^2 values from the MSA-level regressions, $p_{it+1} - p_{it} = \alpha_i + \beta HSI_{it} + \varepsilon_{it+1}$, where p_{it} is the log of the Freddie-Mac house price index in MSA i and HSI_{it} is the housing search index in MSA i. We use the median to distinguish the two groups. For each group we report the average R^2 value, supply elasticity, population, income per capita and population density measured both 0 and 10 miles from the city hall. In Panel B, we sort based on the slope coefficient in the MSA-level regression. The sample period is 2004:1 to 2021:1.

	Low	High
Panel A: Sorting based on R^2		
R^2	32.87%	61.16%
Supply elasticity	2.19	1.81
Population	1,519,948	3,138,546
Income per capita	42084	41250
Population density 0 miles	9212	12310
Population density 10 miles	2762	4953
Panel B: Sorting based on β		
β	2.68	7.43
Supply elasticity	2.36	1.64
Population	1,491,210	3,168,039
Income per capita	42153	41179
Population density 0 miles	7772	13828
Population density 10 miles	2404	5312

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.

Barsky, R.B., Miron, J.A. (1989). The seasonal cycle and the business cycle. *Journal of Political Economy* 97, 503-534.

Bayer, P., Geissler, C., Mangum, K., Roberts, J. (2020). Speculators and middlemen: The strategy and performance of investors in the housing market. *Review of Financial Studies* 33, 5212–5247.

Breiman L. (1997). *Arcing the edge.* Technical report 486, Department of Statistics, University of California Berkeley.

Breiman, L. (2001). Random forests. *Machine learning* 45, 5-32.

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.

Chen, T., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Working paper.

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

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

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.

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

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.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29, 1189-1232.

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

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

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

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

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

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

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

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

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

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

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 Financ*e 54, 203-235.

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.

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

Ngai, L. R., Tenreyno, S. (2014). Hot and cold seasons in the housing market. *American Economic Review* 104, 3991-4026.

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

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

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

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 firm and time. *Journal of Financial Economics* 99, 1-10.

Wei, M., Wright, J. (2013). Reverse regressions and long-horizon forecasting. *Journal of Applied Econometrics* 28, 353-371.

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.