



Google Search Queries, Foreclosures, and House Prices

Damian S. Damianov¹  · Xiangdong Wang¹ · Cheng Yan²

Published online: 29 August 2020
© The Author(s) 2020

Abstract

We study whether Google search behavior for “mortgage assistance” and “foreclosure help” aggregated in the mortgage default risk indicator (MDRI) of Chauvet et al. (2016) helps predict future house prices and foreclosures in local residential markets. Using a long-run equilibrium model, we disaggregate house prices into their fundamental and bubble components, and we find that MDRI dampens both components of house prices. This negative relationship is robust to various model specifications and time horizons. A higher intensity of search online, however, is associated with lower future foreclosure rates. We also find that foreclosure rates increase after a decline in the fundamental component of home values, but are not sensitive to their transitory (bubble) component. Foreclosure rates are higher in metropolitan areas located in non-recourse states. We interpret these findings as evidence for strategic household behavior. Our paper sheds new light on the predictive power of household sentiment derived from Google searches on prices and foreclosure rates in local housing markets.

JEL Classification D12 · D14 · E51 · G21 · G33 · L85 · R31

Keywords Mortgage default risk · Foreclosures · House prices

✉ Xiangdong Wang
xiangdong.wang@dur.ac.uk

Damian S. Damianov
damian.damianov@durham.ac.uk

Cheng Yan
Cheng.Yan.1@cass.city.ac.uk

¹ Durham University Business School, Mill Hill Lane, Durham DH1 3LB, UK

² Essex Business School, University of Essex, Colchester CO4 3SQ, UK

Introduction

The subprime mortgage crisis serves as a powerful reminder of the seismic impact that the financial behavior of homeowners can exert on the U.S. financial system and economy. In the aftermath of the financial crisis, a voluminous literature has developed that aims to shed light on a key relationship in the run-up to the crisis: the interdependence between downward spiraling house prices and rising mortgage default rates. A better grasp of this issue was a matter of urgency during the housing market downturn as policymakers evaluated initiatives to curb the wave of foreclosures and help ‘underwater’ homeowners to stay in their homes (Calomiris et al. 2013; Foote et al. 2008). Yet, the topic remains high on the public policy agenda as it lays bare the tension between housing affordability and financial stability, and carries implications for mortgage market design and macro-prudential regulation.

In the post-crisis period, there has also been substantial interest in the development of mortgage default risk indicators which can serve as a “warning signal” for ensuing future turmoil in housing and mortgage markets. The construction of such forward-looking sentiment indices from household survey data (such as the consumer sentiment survey of the University of Michigan) however has proven elusive. Household surveys are constrained with respect to geographical coverage and number of participants. Their reliability is further complicated by the reluctance of respondents to truthfully answer sensitive questions particularly related to their financial affairs (Singer and Ye 2013), and hence they are of limited use as a predictive tool particularly in the context of housing and mortgage markets.

A viable alternative that has increasingly been pursued in recent research is the creation of sentiment indices from internet search queries. Da et al. (2011, 2015) develop an investor sentiment indicator for the stock market while Beracha and Wintoki (2013) and Van Dijk and Francke (2018) create a proxy for housing demand and show that online behavior has predictive power for house prices and liquidity in local residential markets. More recently, Chauvet et al. (2016) construct a mortgage default risk index (MDRI) based on the intensity of online searches for keywords such as “mortgage help” and “foreclosure assistance” captured by Google Trends. They show that this broad-based index predicts house price returns, returns on subprime credit default swaps and other relevant mortgage indicators, and conclude that MDRI “acts as a leading indicator of the most up-to-date, real-time measures of housing market performance.”

Despite the advantages of MDRI as a predictive tool relative to survey-based alternatives, little is known about the identity, reasons, or intentions of the households whose online searches are aggregated in the index. As Chauvet et al. (2016) point out, “searches are derived from all households, a universe that includes both owners and renters,” yet, it may be assumed that “the bulk of such searches likely emanate from property owners as they are most likely to be concerned with mortgage default.”

While this assumption seems plausible, it is unknown how households process the information they gather in their online searches. One possibility, suggested by Chauvet et al. (2016) is that MDRI captures “household concerns about mortgage failure or foreclosure.” Another plausible alternative is that households learn by searching for relevant terms online and condition their behavior on the information they gathered. That is, as a result of the information they collect online, households may adapt their

behavior when dealing with financial distress, learning how to take advantage of government programs, or interacting with their mortgage lenders. Tetlock (2007) for example, hypothesizes a similar bi-directional relationship when studying the effect of negative media coverage on investor sentiment: While news printed in the Wall Street Journal might convey investor attitudes toward stocks not yet impounded in asset prices, they might also directly shape investors' perception of stocks.¹ Similarly, online searches might divulge information and at the same time convey information to economic agents who then act on this information. Indeed, top results from online searches include information on government programs to avert foreclosure as well as legal information. The mechanism of information acquisition by online searchers, however, is likely different from the one discussed by Tetlock (2007). While investor reaction to media content described by Tetlock is consistent with noise trader theories implying irrational behavior, the information gathering by households via online searches could be rational. Online searches can help households chart an optimal plan of action given the legal and institutional framework available in the state in which they reside as well as provide guidance on how to take advantage of government assistance programs.

A third possible scenario is that some searches are originating from prospective home buyers or home sellers who are trying to time their transactions or from investors trying to form expectations about the future performance of mortgage-related assets. Online searches thus might reflect the expectations about future market trends of this group of agents.

From a theoretical perspective, these three hypotheses are consistent with different causal relationships. The first hypothesis would predict *an increase in foreclosures* while the second hypothesis would predict *a decrease in foreclosures* as a result of a surge in the MDRI. The third hypothesis would imply *no relationship* between MDRI and *foreclosures* but a *negative relationship* between MDRI and future house prices as agents reveal their negative expectations about future market trends when searching online. Currently little is known about which of these hypotheses applies to local housing markets as most of the analysis by Chauvet et al. (2016) is conducted at the national level (local level analysis is restricted in terms of geographical coverage and does not account for metropolitan-area specific demographic and economic conditions).

The main objective of this paper is to explore the relationship between MDRI and future house prices and foreclosure rates in local housing markets. We advance previous research by expanding the set of metropolitan areas and accounting for the differences in appreciation rates between house price segments within the same geographical area. Furthermore, we take into account local economic conditions as well as relevant aspects related to mortgage lending at the regional level. Using a large set of metropolitan-area specific fundamental factors, we estimate a long-run equilibrium model and disaggregate house prices into their fundamental (equilibrium) component and bubble (deviation from equilibrium) component. We then study the relationship between MDRI and future house prices as well as their fundamental and bubble components. Further, we use this house price disaggregation to provide a more detailed analysis of the impact of the fundamental and bubble components on foreclosures.

¹ Tetlock (2007) finds evidence for the latter causal direction but not for the former one.

Related Literature

This paper contributes to two distinct strands of literature. The first strand examines the predictive power of online search intensity on real economic activity. Almost a decade ago, Hal Varian (Google's Chief Economist) suggested that Google Trends data on the search volume for specific keywords helps predict information contained in future government data releases.² Since then academics have explored the predictive power of Google's Search Volume Index (SVI) in other domains such as business activity and financial markets. Da et al. (2011, 2015) show that SVI captures investor attention and predicts stock prices at 2-week horizons. Beracha and Wintoki (2013) show that search intensity for terms such as "real estate" and "rent" help predict home prices. Chauvet et al. (2016) construct a mortgage default risk index from the search intensity of SVI for terms such as "mortgage assistance" and "foreclosure help," and show that this index helps predict housing return, mortgage delinquency indicators, and subprime credit default swaps. In this paper, we examine the predictive power of this index for city-level housing appreciation rates in different market segments while taking into account local fundamental factors, mortgage market conditions, and mortgage market legislation in the states in which metropolitan areas are located (i.e. whether mortgage contracts are recourse or non-recourse). In recourse states, lenders can pursue borrowers for the mortgage balance remaining after foreclosed properties are sold while in non-recourse states they cannot.

The second strand of literature, which developed rapidly in the aftermath of the subprime mortgage crisis, explores the impact of house prices on foreclosure rates. Studies on the contributing role of price declines to mortgage defaults examine the extent to which household behavior conforms to the "option-theoretical" model of mortgage default. A key prediction of this theory is that households find it optimal to walk away from their investment as soon as their equity falls below a certain (negative) threshold (Foster and Van Order 1984; Kau et al. 1994). Closely related research on the 'double trigger' hypothesis has developed which aims to disentangle the contributing role of the strategic motive from that of affordability issues and cash flow problems of households (e.g. income shock related to job loss, divorce, or unforeseen healthcare expenses). Empirical studies conducted before the financial crisis find that negative equity is indeed a significant determinant of default (see e.g., Deng et al. 2000; Bajari et al. 2008; and Foote et al. 2008). Using the data from the financial crisis, Elul et al. (2010) present the estimates for the contributions of negative equity, illiquidity (measured by credit card utilization), unemployment shocks and the existence of a second mortgage to the probability of default. More recently, Kelly and McCann (2016) find that short-term arrears are primarily driven by unemployment, negative income shocks or divorce, while long-term arrears are much more likely to be due to negative equity. Using post-crisis data, Mocetti and Viviano (2017) identify job losses as a primary reason for mortgage delinquencies. Ghent and Kudlyak (2011) find that borrowers are 30% more likely to default in non-recourse states, whereby this effect is much stronger for homeowners of high-value homes. Moreover, Guiso et al. (2013) use survey data to demonstrate that the willingness to default increases in the home-equity shortfall. Further, the exposure to

² Choi and Varian (2012) provide evidence that Google Trends data help predict "turning points" in home sales, automotive sales, and international travel.

people who recently defaulted for strategic reasons increases default probabilities because it shows that lenders are unlikely to pursue a deficiency judgment against borrowers.

The interdependence between foreclosures and house prices has also received considerable attention in the literature. Lin et al. (2009) explore how the distance to a foreclosed property in space and time impacts home values by focusing on the role of comparables in the price formation process. LaCour-Little et al. (2020) argue that the value of comparables need to be adjusted for loan assumability, particularly in areas with a higher concentration of Federal Housing Administration insured loans. Calomiris et al. (2013) provide evidence that foreclosures dampen house prices, yet the negative impact of prices on foreclosures is much greater, in line with the theory of strategic borrower behavior. In contrast, Bhutta et al. (2017) find that emotional and behavioral factors are more important in the decision-making process of households than option-theoretic considerations. Gerardi et al. (2018) use data from the Panel Study of Income Dynamics (PSID) to assess the relative importance of negative equity versus ability to pay. While they find that strategic effects are important, changes in the ability to pay (e.g., job losses) have large estimated effects.

In this paper, we add to these studies by disaggregating house prices into their fundamental and bubble components and differentiating between recourse and non-recourse states. Consistent with strategic motives for default, we find that homes are foreclosed at higher rates in Metropolitan Statistical Areas (MSAs) located in non-recourse states. Furthermore, foreclosures increase when fundamental home values decline but are not sensitive to transitory deviations from equilibrium (bubble component of house prices).

The remainder of this paper is organized as follows. In “[Methodology](#)” section, we present the methodology and in “[Data, Variable Construction, and Summary Statistics](#)” section we describe the data. The empirical results are presented in “[Results](#)” section, and the concluding remarks in “[Conclusion](#)” section.

Methodology

We begin our analysis by estimating a fundamental house price model. We assume that house prices converge toward their equilibrium values in the long run, yet may exhibit deviations from equilibrium in the short run. Furthermore, as different segments of the housing market (i.e. starter homes and trade-up homes) might react differently to changes in fundamentals, we allow for different functional relationships between fundamentals and Top tier and Bottom tier house prices. That is, the relationships between Top and Bottom house price tiers and fundamentals are given by the functions

$$P_{i,t}^{j*} = f_j(X_{i,t}) \quad (1)$$

where $P_{i,t}^{j*}$ is the logarithm of the fundamental value of the house in tier $j \in \{T, B\}$ (Top and Bottom) in MSA i , in month t and $X_{i,t}$ is the vector of fundamental variables. Following Abraham and Hendershott (1996) and Capozza et al. (2004), we consider population, income, employment rate, user cost, and construction cost of housing in the MSA as fundamental factors. Further, because house prices are also affected by regional geographical and regulatory constraints, we add the land supply elasticity

estimates derived by Saiz (2010) as a fundamental factor.³ These supply elasticity indices vary across MSAs but not across time.

The objective of the fundamental model is to estimate the relationships $f_j(\cdot)$ yet a key concern with the estimation is that the levels of the house price indices and (some of) the fundamental factors might be non-stationary. A standard approach to address this issue is the estimation of an error correction framework, and the literature has proposed various specifications for the long-run relationship between house prices and fundamentals as well as the short-run dynamics of house prices (see, e.g. Drake 1993, Ashworth and Parker 1997, Kasparova and White 2001, and Stevenson 2008). In this paper, we estimate versions of the error correction mechanism proposed by Abraham and Hendershott (1996). This estimation method accounts for the serial correlation and the mean reversion in the time series of US housing returns that are widely documented in the literature (see, e.g. Case and Shiller 1989, 1990).

We denote the actual appreciation rates of house prices (i.e. continuously compounded returns) of the two house tiers by $\Delta P_{i,t}^j = P_{i,t}^j - P_{i,t-1}^j$, and the appreciation rates of the fundamental house prices to be estimated by $\Delta P_{i,t}^{j*}$. Further, we assume that the way prices respond to fundamental factors is given by a linear relationship

$$\Delta P_{i,t}^j = \alpha_0^j + \alpha_1^j \Delta X_{i,t} + \theta_{i,t}^j \quad (2)$$

Hereby $\alpha_0^j + \alpha_1^j \Delta X_{i,t}$ is the change in the fundamental value, which we denote by $P_{i,t}^{j*}$, and $\theta_{i,t}^j$ denotes the “error term” which accounts both for momentum and mean reversion effects and is given by the equation

$$\theta_{i,t}^j = \lambda_0^j + \lambda_1^j \Delta P_{i,t-1}^j + \lambda_2^j (P_{i,t-1}^{j*} - P_{i,t-1}^j) + \varepsilon_{i,t}^j \quad (3)$$

In this equation, the coefficient λ_1^j measures the momentum (serial correlation) while the coefficient λ_2^j measures the speed of adjustment to the long-run equilibrium. Combining eqs. (2) and (3) we obtain:

$$\Delta P_{i,t}^j = \gamma_0^j + \alpha_1^j \Delta X_{i,t} + \lambda_1^j \Delta P_{i,t-1}^j + \lambda_2^j (P_{i,t-1}^{j*} - P_{i,t-1}^j) + \varepsilon_{i,t}^j \quad (4.OLS)$$

where $\gamma_0^j = \alpha_0^j + \lambda_0^j$. In addition to an OLS specification, we estimate fixed-effects models that allow for heterogeneity among MSAs and/or time⁴

$$\Delta P_{i,t}^j = \gamma_0^j + \alpha_1^j \Delta X_{i,t} + \lambda_1^j \Delta P_{i,t-1}^j + \lambda_2^j (P_{i,t-1}^{j*} - P_{i,t-1}^j) + \vartheta_i + \varepsilon_{i,t}^j \quad (4.MSA - FE)$$

³ Previous literature has considered related measures such as the percentage of land available for development (see, e.g. Rose 1989 or Capozza and Seguin 1996).

⁴ As the housing supply elasticity is time-invariant for each metropolitan area, in the MSA-level fixed effect regression specifications (4.MSA-FE) and (4.MSA&Time-FE) we exclude this variable from the vector of fundatament factors. For notational convenience, we still use $\Delta X_{i,t}$ to denote the relevant fundamental factors in these two model specifications.

$$\Delta P_{i,t}^j = \gamma_0^j + \alpha_1^j \Delta X_{i,t} + \lambda_1^j \Delta P_{i,t-1}^j + \lambda_2^j (P_{i,t-1}^{j*} - P_{i,t-1}^j) + \mu_t + \varepsilon_{i,t}^j \quad (4. \text{Time} - \text{FE})$$

$$\Delta P_{i,t}^j = \gamma_0^j + \alpha_1^j \Delta X_{i,t} + \lambda_1^j \Delta P_{i,t-1}^j + \lambda_2^j (P_{i,t-1}^{j*} - P_{i,t-1}^j) + \vartheta_i + \mu_t + \varepsilon_{i,t}^j \quad (4. \text{MSA} \& \text{Time-FE})$$

One difficulty with this estimation is that the fundamental values $P_{i,t-1}^{j*}$ depend on the estimates of the different versions of eqs. (4) while at the same time they are part of the error correction term which is used as an explanatory variable in these equations. We resolve this issue using the iterative procedure proposed by Abraham and Hendershott (1996). We assume that the observed house price in December 1999 corresponds to its fundamental value (i.e. $P_{i,t}^{j*} = P_{i,t}^j$ for $t = \text{December 1999}$) and recover the fundamental value time series from the relationship $P_{i,t}^{j*} = P_{i,t-1}^{j*} + \Delta P_{i,t}^{j*}$. We then re-estimate eqs. (4) and recalculate fundamental prices repeatedly until the estimates stabilize (typically we need to perform up to five iterations).⁵

We then analyze how the current (and past) values of the mortgage default risk index, $MDRI_{i,t}$ impacts the future values of the fundamental component $P_{i,t}^{j*}$ and the bubble component $B_{i,t}^j = P_{i,t}^j - P_{i,t}^{j*}$ of local house prices as well as the foreclosure rates $HF_{i,t}$. Furthermore, we use the house price decomposition to explore how changes in the fundamental $P_{i,t}^{j*}$ and the bubble $B_{i,t}^j$ components of home values affect foreclosure rates $HF_{i,t}$.

Data, Variable Construction, and Summary Statistics

The estimation of the fundamental house price model is based on a panel of 107 MSAs located in 29 U.S. states. A map with the location of these MSAs is presented in Fig. 1. For each MSA we observe the monthly growth rate of house prices and local fundamental factors. Further, in our analysis of the effect of the mortgage default risk index on house prices and foreclosures, we include additional variables that account for the mortgage market conditions in each MSA.

All MSAs in the dataset are listed in Table 7 in the Appendix along with the state in which they are located. The table also contains the classification of states into the recourse and non-recourse categories depending on whether states allow lenders to pursue a deficiency judgment against foreclosed borrowers (we use the classification of Ghent and Kudlyak 2011). In this figure, the recourse states are depicted in dark blue and the non-recourse states are represented in light blue color.

⁵ The initial fundamental value time series $P_{i,t}^{j*}$ are obtained by estimating equation (4) without the error correction term.

Local House Prices and Fundamental Factors

Although there is a variation across regional housing market segments, on average the indices peak in late 2006, and then decline and reach their lowest values between 2009 and 2012. They recover thereafter by almost reaching

⁸ For more information on the Zillow methodology see <https://www.zillow.com/research/zhvi-methodology/>

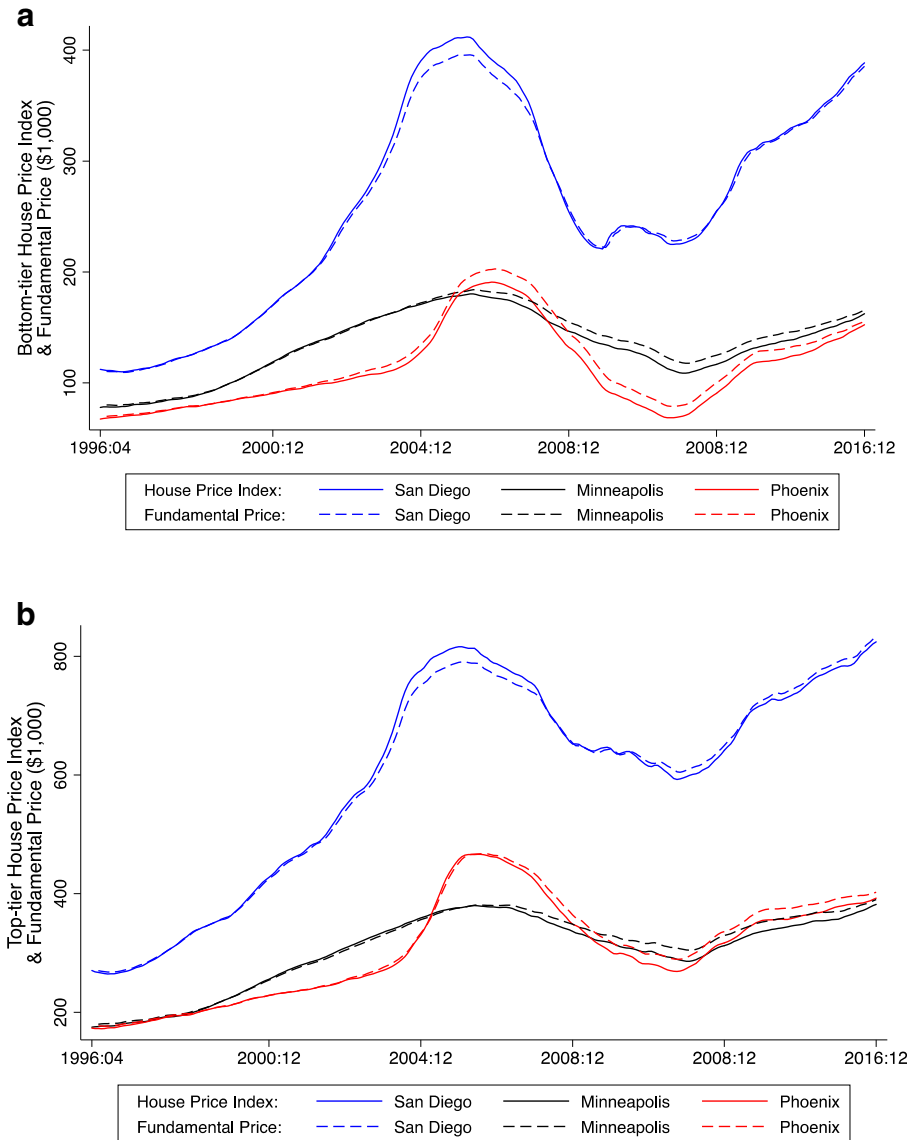


Fig. 2 Dynamics of house price indices for selected MSAs. **a** Bottom-tier house price index (solid lines) and fundamental price (dash lines). **b** Top-tier house price index (solid lines) and fundamental price (dash lines). **c** Bubble component of the bottom-tier house price index. **d** Bubble component of the top-tier house price index. *Notes:* San Diego, Minneapolis, and Phoenix represent examples of a cyclical, a steady, and a bubble market, respectively, according to the classification of Mayer (2011). The bubble component is calculated as the deviation from the fundamental house price, i.e. the difference between the logs of the house price index and its fundamental component: $B_{i,t}^j = P_{i,t}^j - P_{i,t}^{j*}$

their pre-crisis period values around 2016. In our analysis, we use the log differences of the price indices (i.e., the continuously compounded returns) for the two market segments.

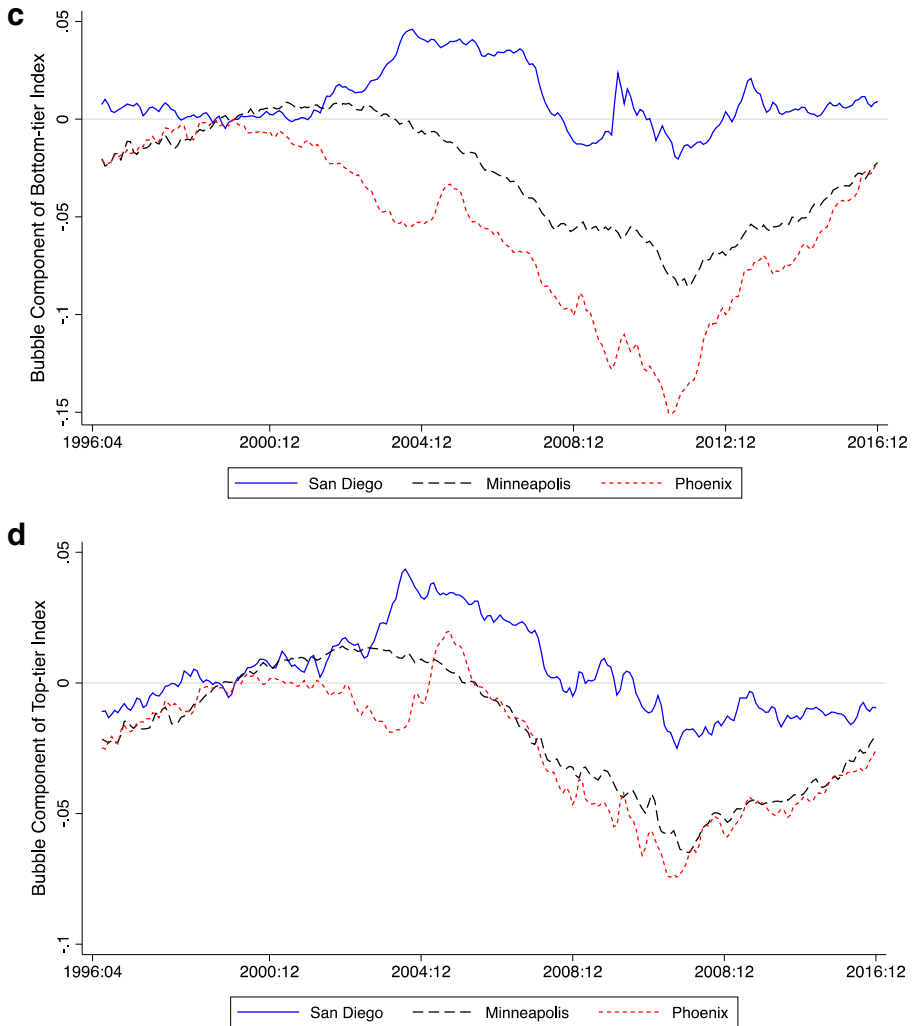


Fig. 2 (continued)

The fundamental variables used include the population, personal income per capita, total non-farm employment, construction cost, a derived user cost of homeownership, and the land supply elasticity index in the MSA. Descriptive statistics of these variables, except for land supply elasticity which is time-invariant, along with unit root tests are presented in Table 1.

The population and personal income data are collected from the Bureau of Economic Analysis. We use cubic spline interpolation (De Boor 1978) to derive monthly values from the original annual observations. The total non-farm employment, available at the state level, is collected from Datastream and used for all metropolitan areas located in the same state. The construction cost is measured by the price index of new single-family

Table 1 Descriptive statistics

Variables	Sample period	Data frequency	Mean	SD	Min	Max	ADF test	
							Levels	First difference
Top house price tier	Apr 1996 - Dec 2016	Monthly	320,162	199,068	70,400	1,785,000	0.8811	-48.3741***
Bottom house price tier	Apr 1996 - Dec 2016	Monthly	127,811	83,657	28,200	636,600	-0.8656	-45.0178***
MDRI	Jan 2004 - Dec 2016	Monthly	122.13	62.98	13.74	554.60	6.9389	-83.7958***
Homes foreclosed	Jan 1998 - Dec 2016	Monthly	5.46	7.35	0.01	106.20	-4.3333***	-
Population	Apr 1996 - Dec 2016	Monthly (from Annual)	1,050,383	1,607,180	77,482	13,300,000	6.3028	-4.6023***
Employment	Apr 1996 - Dec 2016	Monthly	6096.61	5138.90	437.00	16,637.90	6.4994	-51.7601***
Personal income	Apr 1996 - Dec 2016	Monthly (from Annual)	35,931.53	10,535.51	16,424.82	107,936.40	22.221	-46.1546***
Construction cost	Apr 1996 - Dec 2016	Monthly	93.43	15.03	67.70	124.00	19.476	-75.7401***
User cost	Apr 1996 - Dec 2016	Monthly	0.14	0.02	0.10	0.22	-35.9084***	-
Loan supply	Jan 2007 - Dec 2016	Annual	7,156,725	12,100,000	154,869	119,000,000	-1.355*	-6.7123***
Subprime mortgages (%)	Jan 2007 - Dec 2016	Annual	0.0259	0.0458	0.0003	0.2356	-16.7981***	-

The table reports descriptive statistics and stationarity tests. The Top and Bottom house price tiers are measured in US Dollars. The MDRI is constructed from 11 Google Trends search items such as “foreclosure,” “mortgage help,” or “government assistance,” (see Chauvet et al. 2016, Table 1, for the full list). The Homes foreclosed are the number of foreclosures per 10,000 homes. The monthly observations of Population and Personal income are derived from annual data via cubic spline interpolation (De Boor 1978). The User cost is constructed from equation (5). The tests for stationarity are Fisher-type Augmented Dickey-Fuller (ADF) tests for unbalanced panels that have as the null hypothesis that all panels contain a unit root (Choi 2001). The ADF tests are performed on the logarithms of the variables. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

houses under construction, which is available from the U.S. Census Bureau. As only the national index is available in monthly frequency, the change in construction costs varies over time but not across MSAs. As a measure of land supply elasticity of MSAs, we use the land supply estimates derived by Saiz (2010).⁹

To facilitate comparison to previous research, we construct the user cost by the method of Capozza et al. (2004) which accounts for mortgage rates, taxes, expected appreciation as well as annual maintenance and depreciation of properties. That is, the user cost is constructed by the formula

$$\begin{aligned} \text{User cost} = & (\text{Mortgage Rate} + \text{Property Tax Rate}) \\ & \times (1 - \text{Income Tax Rate}) - \text{Inflation} + 0.03 \end{aligned} \quad (5)$$

Here the “Mortgage Rate” is the 30-Year fixed-rate mortgage average in the United States, collected from the Federal Reserve Bank of St. Louis. The “Property Tax Rate,” collected from Wallethub,¹⁰ is the effective real-estate state tax rate. The “Income tax rate” is the sum of the average federal income tax rate and average state income tax rate for the middle quintile of households. The federal income tax rate is collected from the Urban-Brookings Tax Policy Center,¹¹ while the state income tax rate is collected from the National Bureau of Economic Research.¹² For inflation, we use the CPI provided by the Federal Reserve Bank of St. Louis. The annual maintenance and obsolescence of properties are set at 3% as indicated in eq. (5).

Mortgage Lending

To account for local mortgage market conditions we construct two variables from the Home Mortgage Disclosure Act (HMDA) data¹³: the total amount of mortgage loans in a given year in each MSA (Loan supply) and the percentage of loans that are subprime, or higher-priced mortgage loans in each MSA (Subprime). Loans are categorized as subprime following the classification of Mayer and Pence (2008) according to which a mortgage is a subprime mortgage if its rate spread exceeds 3% for first-lien mortgages and 5% for junior lien mortgages.¹⁴

⁹ As this elasticity measure has only limited coverage, we are left with only 93 MSAs in our sample. Another way to account for differences across MSAs is to estimate a model with MSA-level fixed effects while leaving out the supply elasticity as a regressor. In Table 2 we report results for both the OLS and the fixed effect model, but use the estimates of the fixed effect model in the subsequent analysis because this model allows us to use all 107 MSAs in our sample.

¹⁰ Property tax rates are collected from: <https://wallethub.com/edu/states-with-the-highest-and-lowest-property-taxes/11585/>

¹¹ The average federal income tax rate is downloaded from:

<https://www.taxpolicycenter.org/statistics/historical-average-federal-tax-rates-all-households>

¹² The state income tax rate is downloaded from: <http://users.nber.org/~taxsim/state-tax-rates/>. We apply the rates for a family income of \$50,000.

¹³ The HMDA data contains over 80% of home loans and is the most comprehensive source of data on mortgage loans (Avery et al. 2007).

¹⁴ The rate spread is the difference between the Annual Percentage Rate (APR) and a survey-based estimate of APRs offered on prime mortgage loans of a comparable type utilizing the “Average Prime Offer Rates” fixed table or adjustable table, action taken, amortization type, lock-in date, APR, fixed term (loan maturity) or variable term (initial fixed-rate period), and reverse mortgage.

Mortgage Default Risk

The Mortgage Default Risk Index (MDRI hereafter) of Chauvet et al. (2016) is constructed from the Search Volume Index (SVI) data for terms such as “foreclosure help” and “government mortgage help” in US states published by Google Trends.¹⁵ The MDRI is obtained from the UCLA ZIMAN Center for Real Estate.¹⁶ Zillow also publishes a Homes Foreclosed index (HF hereafter) which gives the number of homes foreclosed per 10,000 homes in metropolitan areas each month. As an illustration, in Fig. 3 (see Panels A and B) we present the dynamics of the MDRI and HF in three of the MSAs in our sample – San Diego, Minneapolis, and Phoenix. Both indicators start to increase in early 2007 and reach their peak around 2008 in San Diego, and around 2009 in Minneapolis and Phoenix. The descriptive statistics of these variables are presented in Table 1.

Results

We first explore how real house prices respond to changes in local fundamental factors by estimating the models given in eq. (4). In particular, we consider population, personal income, employment, as well as the variable we created for the user cost, construction cost, and the land supply elasticity of the MSA (cf. Capozza et al. 2004; Stevenson 2008). As a preliminary step, we perform unit root tests on the tiered house price indices as well as the fundamental variables (see the last two columns in Table 1). These variables are non-stationary in levels and stationary in first differences. This points to the inherent difficulties that would be present if we tried to directly use the levels of these variables in our statistical analysis. Furthermore, it justifies our focus on growth rates and the use of an error correction modeling approach.

Long-Run Equilibrium Relationship

The regression results of the error correction models specified in the four versions of eq. (4) are presented in Table 2. They include OLS estimates as well as estimates of fixed-effect models in which we control for MSA and time fixed effects.

The coefficient estimates for all fundamental variables have the anticipated sign and are statistically significant at the 1% or 5% level. As expected, growth in population, personal income, and employment have a positive impact on house prices. An increase in user cost, a significant component of which constitutes the mortgage interest rate, is associated with lower house price growth. Similarly, an increase in construction cost

¹⁵ For the construction of their monthly MDRI, Chauvet et al. (2016) used “foreclosure assistance+foreclosure help+government assistance mortgage+home mortgage assistance+home mortgage help+housing assistance+mortgage assistance program+mortgage assistance+mortgage foreclosure help+mortgage foreclosure+mortgage help” to obtain the joint SVI.

¹⁶ As the city-level MDRI data is only available for 20 cities, we use the state-level MDRI data for all the MSAs in our sample. The data on the MDRI indices are available at: https://github.com/ChandlerLutz/MDRI_Data

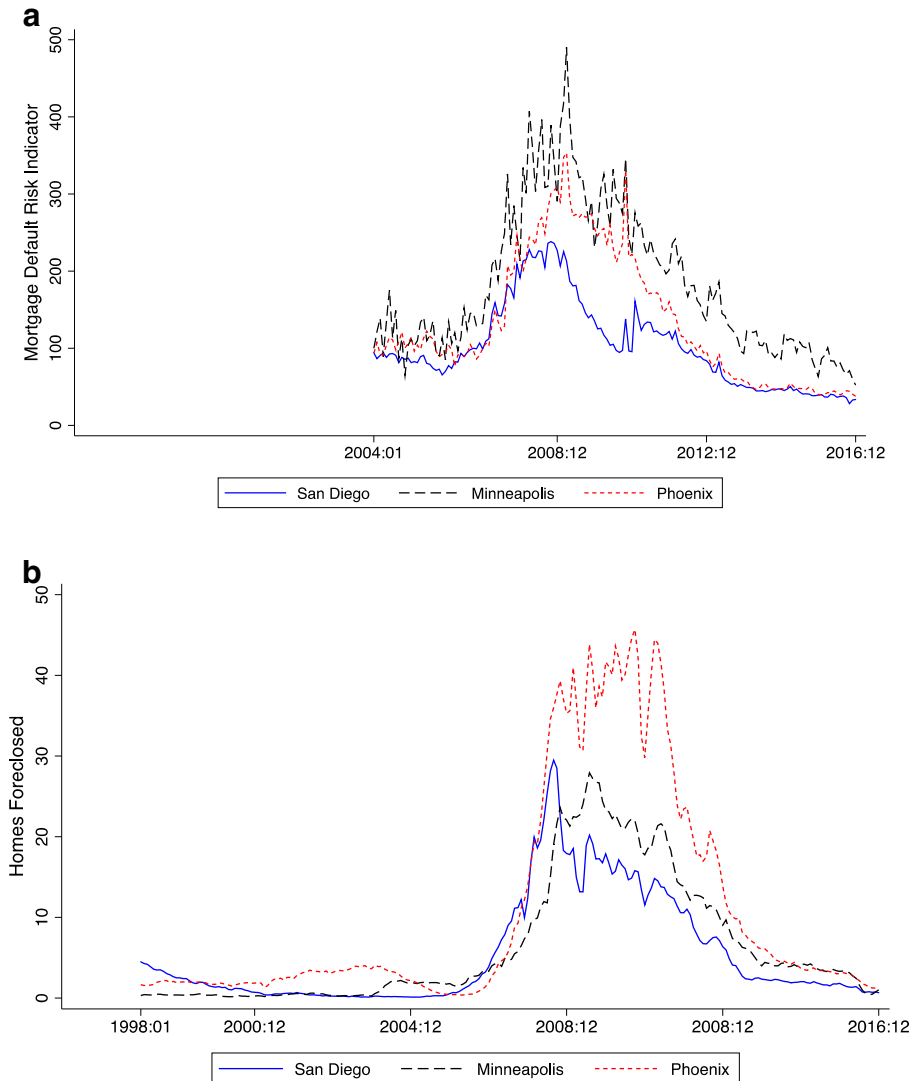


Fig. 3 Default risk indices.

leads to an increase in home values. Further, the relationship between the land supply index and house prices is negative, as had been found in previous literature. The error correction term is significant indicating that both the top and the bottom house price tiers adjust to their long-run equilibrium values. Similarly, the autoregressive coefficient is positive and statistically significant, indicating the presence of momentum in housing returns for both house price tiers.

The magnitude of the coefficients suggests that bottom tier homes are more sensitive to changes in population, employment, user cost, and construction cost as well as exhibit a stronger momentum. We formally test whether the coefficients for the top tier and the bottom tier are significantly different from each other using

Table 2 Estimates of the fundamental house price model

Dependent Variable	Model 1: Pooled OLS		Model 2: MSA Fixed effects		Model 3: Time fixed effects		Model 4: MSA and Time Fixed effects	
	Δ Bottom	Δ Top	Δ Bottom	Δ Top	Δ Bottom	Δ Top	Δ Bottom	Δ Top
Δ Population	0.0080***	0.0072***	0.0324***	0.0347***	0.0121***	0.0115***	0.0452***	0.0446***
Δ Employment	0.0419***	0.0365***	0.0821***	0.0838***	0.0527***	0.0327***	0.0544***	0.0399***
Δ Personal income	0.0409***	0.0520***	0.0260**	0.0352	0.1096***	0.0906***	0.0715***	0.0604***
Change of user cost	-0.0299***	-0.0228***	-0.0239***	-0.0193***	-0.0130**	-0.0122**	-0.0131**	-0.0138***
Δ House Price _{<i>t-1</i>}	0.9032***	0.8909***	0.8944***	0.8779***	0.8664***	0.8431***	0.8506***	0.8343***
Error correction _{<i>t-1</i>}	0.0073***	0.0070***	0.0319***	0.0369***	0.0056***	0.0050***	0.0369***	0.0277***
Δ Construction cost	0.0581***	0.0336***	0.0526***	0.0312***				
Supply elasticity	-0.0001**	-0.0001***			-0.0001***	-0.0001***		
Fixed effect (Time)	No	No	No	No	Yes	Yes	Yes	Yes
Fixed effect (Metro)	No	No	Yes	Yes	No	No	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs	22,815	22,804	26,259	26,248	22,815	22,804	26,259	26,248
Adjusted R-squared	0.8425	0.8242	0.8399	0.8205	0.8462	0.8292	0.8453	0.824

The table presents regression results of the fundamental house price model defined by the four versions of equation (4). The continuously compounded returns of the Bottom and Top tier house price indices are denoted by Δ Bottom and Δ Top, respectively. For the population, employment, personal income, and construction cost variables, the continuously compounded growth rates are used as regressors. Following Abraham and Hendershot (1996), the change of the user cost is used as a regressor. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

the OLS model specification (Model 1 in Table 2). In particular, we construct a dummy variable “Toptier,” which takes on the value of one for the top tier and zero for the bottom tier index. We include it as a regressor along with the interactions of this variable with the fundamental variables. We estimate this regression using Abraham and Hendershott’s (1996) iterative method described in the “Methodology” section by pooling the top tier and bottom tier observations together. We find that only the coefficients for the interaction variables ($\text{Toptier} * \Delta \text{House Price}_{t-1}$) and ($\text{Toptier} * \Delta \text{Construction cost}$) are significant. They have a negative sign indicating starter homes exhibit a stronger momentum effect and their response to construction cost is greater compared to trade-up homes.

The fundamental house price model allows us to disaggregate house price indices into their fundamental and bubble components. Using the estimates of Model 2 (Panel Fixed Effects) in Table 2, we calculate these two components of house price and represent their dynamics for three of the MSAs, (San Diego, Minneapolis, and Phoenix) in Fig. 2. In the following subsections, we analyze whether the MDRI helps predict these components of house prices and whether these components affect future foreclosures.

Effect of MDRI on House Prices

As a next step, we explore how household sentiment revealed by the mortgage default risk index (MDRI) impacts house prices. To account for mortgage market conditions, we add as regressors two variables that we constructed from HMDA data: the total amount of mortgage lending in the previous year (Total Loans), and the percentage of mortgage loans that are classified as subprime (Subprime). The results are reported in Table 3.

We find that an increase in the MDRI index lowers house price growth in the following three to 6 months. In the regression in which all lags are included (see model 8), the coefficients for the lags between 3 and 6 months are statistically significant and range between -0.00017 and -0.0012 . Further, as anticipated we find that the amount of mortgage credit that flows into the area serves to increase home values, while subprime lending in the previous year dampens home values in the current year.

In the Appendix, we present regression results for the 2007–2012 and the 2013–2016 subsamples (see Table 8, Panel A and Panel B, respectively), and we find that the predictive power of MDRI applies mostly for the first subsample that includes the subprime mortgage crisis. In addition to considering actual growth rates of the house price tiers, we also analyze the decomposition of house prices into their fundamental and bubble components. We find that the MDRI dampens both the fundamental component (see Table 9) and the bubble component (see Table 10) of house prices.

The regression results also indicate that foreclosure legislation has a significant effect on home value appreciation rates. In particular, house price growth is on average lower in the metropolitan areas located in recourse states where lenders can pursue a deficiency judgment against borrowers. The coefficient for the recourse dummy variable in Table 3 is statistically significant at conventional levels and equals -0.0002 across all specifications. One possible explanation is that buying a home with a mortgage is less attractive to borrowers in

Table 3 Predictive power of MDRI for the house price appreciation rates (Δ HP)

Dependent Variable	House Price appreciation rate (Δ HP)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ MDRI _{<i>t-1</i>}	-0.0003						-0.0003	-0.0003
Δ MDRI _{<i>t-2</i>}		0.0005					0.0002	-0.0001
Δ MDRI _{<i>t-3</i>}			-0.0008***				-0.0008**	-0.0013***
Δ MDRI _{<i>t-4</i>}				-0.0009***				-0.0017***
Δ MDRI _{<i>t-5</i>}					-0.0003			-0.0013***
Δ MDRI _{<i>t-6</i>}						-0.0008**		-0.0012***
Δ HP _{<i>t-1</i>}	0.8355***	0.8355***	0.8353***	0.8352***	0.8354***	0.8352***	0.8353***	0.8338***
HF _{<i>t-1</i>}	-0.0002***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
Δ Loan supply _{<i>t-12</i>}	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***
Subprime _{<i>t-12</i>}	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
Recourse	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs	19,254	19,254	19,254	19,254	19,254	19,254	19,254	19,254
Adjusted R-squared	0.7696	0.7697	0.7697	0.7697	0.7696	0.7697	0.7697	0.7701

The table presents regression estimates of the effect of the MDRI index on house price appreciation rates. Loan supply_{*t-12*} is the total supply of mortgage loans in the previous year in the MSA (derived from HMDA data). Subprime_{*t-12*} is the percentage of the total amount of mortgage loans in the MSA that are subprime. The Recourse variable takes on the value of one, if the MSA is in a recourse state, and zero otherwise. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

a recourse state where contracts are lacking the put option associated with mortgage default.

Effect of MDRI on Foreclosure Rates

Ghent and Kudlyak (2011), Table 1 provide an overview of foreclosure legislation across US states and present statistics of the timeline of different stages in the foreclosure process in each state. If there are no delays, a non-contested non-judicial foreclosure can take as little as 60 days, yet often the process takes longer. Furthermore, foreclosures are followed by a redemption period with a duration of another 6 months. It could only be speculated when delinquent borrowers start searching online for help and how the intensity of their searches varies over time. To allow for different timing of online search we explore alternative specifications. In Table 4 we present results for search behavior with lags between 1 and 6 months.

We find that an increase in the MDRI lowers foreclosures for horizons between 2 and 6 months.¹⁷ These coefficients are statistically significant and range between -0.0425 and -0.1645 (see model 8). In Table 5 we aggregate the Google searches for periods of 3 months and considers regressions with lags of up to a year.

We find also for this setting that Google searches reduce foreclosures (the coefficients for the lags between 1 and 3 months and lags between 9 and 12 months are statistically significant). As a robustness check, in Table 11 presented in the Appendix we consider the 2007–2012 and the 2013–2016 subsamples and largely find an inverse relationship between MDRI index and future foreclosures.

These findings are consistent with the hypothesis that the MDRI index captures learning effects. That is, by searching online some households may access information that helps them avert foreclosure. Further, consistent with the theory of strategic default, we find that foreclosure rates are lower in recourse states. Similar findings are reported in the recent empirical literature on the effect of recourse on default. For example, Ghent and Kudlyak (2011), Table 3 report that the probability of default of loans made in recourse states is on average 6.2% smaller (although their coefficient estimate is not significant).

Effect of House Prices on Foreclosure Rates

In Table 6 we report results for several alternative specifications. We disaggregate house prices to their fundamental and bubble components and study the contributing effect of these two components on foreclosures.

¹⁷ While in the regression specification including only one lag the MDRI coefficient is positive, it is only marginally significant. Notably, the coefficient for one lag is insignificant when more lags of the MDRI variable are included in the regression. Furthermore, for the specifications including more than one lag, the MDRI coefficient is negative and highly statistically significant.

Table 4 Predictive power of MDRI for the Homes Foreclosed (monthly lags)

Dependent Variable	Homes Foreclosed (HF)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	0.0369**						-0.02	-0.0185
ΔMDRI_{t-2}		-0.1272***					-0.1611***	-0.1645***
ΔMDRI_{t-3}			-0.0587***				-0.1046***	-0.1167***
ΔMDRI_{t-4}				0.0181				-0.0549***
ΔMDRI_{t-5}					-0.0006			-0.0425**
ΔMDRI_{t-6}						-0.0931***		-0.0969***
ΔHP_{t-1}	-1.1015***	-1.1235***	-1.1190***	-1.1026***	-1.1082***	-1.1407***	-1.1508***	-1.2188***
HF_{t-1}	0.9729***	0.9730***	0.9729***	0.9730***	0.9729***	0.9728***	0.9730***	0.9727***
$\Delta \text{Loan supply}_{t-12}$	-0.0508***	-0.0492***	-0.0509***	-0.0506***	-0.0508***	-0.0515***	-0.0489***	-0.0505***
Subprime_{t-12}	0.0093***	0.0097***	0.0093***	0.0093***	0.0093***	0.0093***	0.0097***	0.0097***
Recourse	-0.0102***	-0.0093**	-0.0097***	-0.0100***	-0.0100***	-0.0095**	-0.0085**	-0.0076**
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs	19,148	19,148	19,148	19,148	19,148	19,148	19,148	19,148
Adjusted R-squared	0.9425	0.9426	0.9425	0.9425	0.9425	0.9425	0.9427	0.9428

The table presents regression estimates of the effect of the MDRI index on the log of homes foreclosed. The MDRI is included with monthly lags. Loan supply_{t-12} is the total supply of mortgage loans in the previous year in the MSA (derived from HMDA data). Subprime_{t-12} is the percentage of the total amount of mortgage loans in the MSA that are subprime. The Recourse variable takes on the value of one, if the MSA is in a recourse state, and zero otherwise. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

Table 5 Predictive power of MDRI for the Homes Foreclosed (three-month lags)

Dependent Variable	Homes Foreclosed (HF)					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔMDRI_{1-3}	-0.0616***				-0.0624***	-0.0621***
ΔMDRI_{4-6}		0.0119			0.015	0.0089
ΔMDRI_{7-9}			0.0167			0.0109
$\Delta \text{MDRI}_{10-12}$				-0.0434***		-0.0451***
ΔHP_{t-1}	-1.1262***	-1.1002***	-1.0922***	-1.1308***	-1.1166***	-1.1339***
HF_{t-1}	0.9730***	0.9730***	0.9730***	0.9727***	0.9730***	0.9728***
$\Delta \text{Loan supply}_{t-12}$	0.0094***	0.0093***	0.0093***	0.0095***	0.0094***	0.0096***
Subprime_{t-12}	-0.0500***	-0.0506***	-0.0506***	-0.0527***	-0.0497***	-0.0517***
Recourse	-0.0093**	-0.0101***	-0.0101***	-0.0097***	-0.0094**	-0.0092**
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs	19,148	19,148	19,148	19,148	19,148	19,148
Adjusted R-squared	0.9425	0.9425	0.9425	0.9425	0.9425	0.9425

The table presents regression estimates of the effect of the MDRI index on the log of Homes Foreclosed. The MDRI is included with three-month lags (e.g. ΔMDRI_{1-3} is the change in the log the MDRI index for the previous three months, etc.). $\text{Loan supply}_{t-12}$ is the total supply of mortgage loans in the previous year in the MSA (derived from HMDA data). Subprime_{t-12} is the percentage of the total amount of mortgage loans in the MSA that are subprime. The Recourse variable takes on the value of one, if the MSA is in a recourse state, and zero otherwise. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

The finding that is robust across all specifications is that the foreclosures respond to changes in fundamental values. A 1 % drop in fundamental home values increases the log of the homes foreclosed by 1.2, i.e. leads to about 3.3 extra foreclosures in the following month for every 10,000 homes. The bubble component, on the other hand, does not appear to have an impact on the proportion of defaulting homeowners. We interpret this as further evidence for strategic sophistication by homeowners. A shock to the fundamental component of home values would have a long-term effect on future house prices while a shock to the bubble component would disappear over time as home values revert to their long-run equilibrium. Indeed, note that the speed of adjustment coefficient in the fundamental equation reported in Table 2 is significant and has the anticipated sign. We further examine strategic default behavior by constructing a dummy variable for house price declines of more than 5% in the past twelve months ($\text{PriceDecline} > 5\%$). We do not find evidence that foreclosures are driven primarily by option-theoretic defaults: the coefficients for both the Recourse dummy and the interaction term of the Recourse dummy with the ($\text{PriceDecline} > 5\%$) dummy are not significant. These findings correspond to results reported in the recent empirical literature documenting the relative rarity of defaults due solely to strategic motives, and the relative importance of affordability constraints. Bhutta et al. (2017) find that homeowners do not walk away from their investments unless they are

Table 6 Predictive power of house price declines for the Homes Foreclosed

Dependent Variable	Homes Foreclosed (HF)			
	(1)	(2)	(3)	(4)
$\Delta \text{Fundamental}_{t-1}$	-1.1965***	-1.1971***	-1.2142***	-1.2071***
Bubble_{t-1}	-0.0895**	-0.0862*	-0.0898**	-0.0863*
HF_{t-1}	0.9704***	0.9695***	0.9703***	0.9695***
$\Delta \text{Loan supply}_{t-12}$	-0.0480***	-0.0480***	-0.0479***	-0.0479***
Subprime_{t-12}	0.0090***	0.0092***	0.0089***	0.0092***
$\text{Price Decline} > 5\%$	0.0146***	0.0203***	0.0128*	0.0192**
ΔMDRI_{t-1}	-0.017	-0.0159	-0.0171	-0.016
ΔMDRI_{t-2}	-0.1634***	-0.1622***	-0.1635***	-0.1622***
ΔMDRI_{t-3}	-0.1169***	-0.1157***	-0.1171***	-0.1158***
ΔMDRI_{t-4}	-0.0559***	-0.0549***	-0.0561***	-0.0550***
ΔMDRI_{t-5}	-0.0434**	-0.0426**	-0.0436**	-0.0427**
ΔMDRI_{t-6}	-0.0987***	-0.0980***	-0.0989***	-0.0981***
Recourse		-0.0024		-0.0024
Recourse * Price Decline > 5%		-0.0146		-0.0143
Top tier			0.0003	0.0004
Top tier * Price Decline > 5%			0.0045	0.0026
Constant	Yes	Yes	Yes	Yes
Number of Obs	19,148	19,148	19,148	19,148
Adjusted R-squared	0.9428	0.9428	0.9428	0.9428

The table presents regression estimates of the effect of house prices on foreclosures. Price Decline > 5% takes on the value of one if house prices in the MSA declined more than 5% in the last 12 months and zero otherwise. The dummy variable Top tier equals one for the top-tier house prices, and zero for the bottom-tier house prices. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

substantially more ‘underwater’ than option-theoretical models would predict, while Gerardi et al. (2018) find that default is primarily driven by income shocks rather than strategic motives. We further explore how foreclosures respond to house price declines of more than 5% in the previous year and find that the foreclosures increase more in the MSAs sustaining such declines. There is no evidence of a differential impact of such declines in recourse and non-recourse states. Furthermore, the effect on foreclosures is about the same regardless of whether the decline has originated in the top tier or the bottom tier of the local housing market.

Conclusion

As of 2020, Google commands more than 92% of the search engine market share worldwide with an estimated number of approximately 2 trillion global

searches per year (www.hubspot.com). Extant research has established that Google searches provide timely indicators for social and economic activity in a variety of domains ranging from automotive sales to the spread of infections, to asset returns in financial and housing markets. Da et al. (2011) find that search volume data predict stock returns and conclude that “search data has the potential to objectively and directly reveal to empiricists the underlying belief of an entire population of households.” Chauvet et al. (2016) constructed a mortgage default risk index from data on Google search volumes for keywords such as “mortgage help” and “foreclosures assistance” and demonstrated that this index has predictive power for the returns on housing and mortgage-related assets.

In this paper, we analyze how this mortgage default risk index is related to house prices and foreclosure rates in local housing markets. Using a long-run equilibrium model, we disaggregate local house prices into their fundamental and bubble components. We then explore how the mortgage default risk index relates to future housing market outcomes such as house prices and foreclosures. In line with previous literature, we find that an increase in the mortgage default risk index leads to lower house price appreciation rates. Perhaps somewhat surprisingly, we also find that an increase in the mortgage default risk index reduces the percentage of foreclosures for various time horizons. One interpretation of these findings is that economic agents not only reveal their sentiments through their search behavior but also collect and process the information they access and as a consequence adapt their behavior. That is, through online searches for “foreclosure help” and “mortgage assistance” households can access relevant information that helps them avert foreclosures.

We also report new results on the interaction between housing and mortgage markets which suggest some degree of household strategic behavior. In particular, we find that declines in the fundamental component of house prices lead to an increase in foreclosure rates while declines in the transitory component of house prices have no statistically significant effect. Furthermore, foreclosures tend to be higher in metropolitan areas situated in non-recourse states where lenders cannot pursue a deficiency judgment against borrowers.

In addition to exploring its predictive properties, one can also use online search data to empirically test economic models that incorporate the learning of economic agents and view equilibria as the outcome of adaptive behavior. The empirical assessment of such models is left for future research.

Acknowledgements We thank Nigar Hashimzade, Alla Koblyakova, Michael LaCour-Little, Lok Man Michelle Tong, Gianluca Marcato, and participants at the 2018 AREUEA International Conference, the 25th Annual Conference of European Real Estate Society, the 5th Money, Macroeconomics and Finance (MMF) Ph.D. conference, the MMF and Cass Business School conference on Investment, Funding and Industrial Policy, and the 2018 FMA Annual meeting for their comments. Cheng Yan would like to gratefully acknowledge the financial support received by the Zhejiang Provincial Natural Science Foundation of China [Grant No: LZ20G010002].

Appendix

Table 7 MSAs and classification of states according to state foreclosure law

Metropolitan area	State ¹	Abbreviation	Recourse / Non-recourse
Akron	Ohio	OH	Recourse
Albany	New York	NY	Recourse
Allentown	Pennsylvania	PA	Recourse
Atlantic City	New Jersey	NJ	Recourse
Bakersfield	California	CA	Non-Recourse
Baltimore	Maryland	MD	Recourse
Bellingham	Washington	WA	Non-Recourse
Bend	Oregon	OR	Non-Recourse
Binghamton	New York	NY	Recourse
Bloomington	Illinois	IL	Recourse
Boulder	Colorado	CO	Recourse
California-Lexington Park	Maryland	MD	Recourse
Canton	Ohio	OH	Recourse
Charlotte	North Carolina	NC	Non-Recourse
Chico	California	CA	Non-Recourse
Cincinnati	Ohio	OH	Recourse
Cleveland	Ohio	OH	Recourse
Colorado Springs	Colorado	CO	Recourse
Columbia	South Carolina	SC	Recourse
Columbus	Ohio	OH	Recourse
Corvallis	Oregon	OR	Non-Recourse
Crestview-Fort Walton Beach-Destin	Florida	FL	Recourse
Cumberland	Maryland	MD	Recourse
Dallas-Fort Worth	Texas	TX	Recourse
Dayton	Ohio	OH	Recourse
Denver	Colorado	CO	Recourse
Eugene	Oregon	OR	Non-Recourse
Fayetteville	North Carolina	NC	Non-Recourse
Flagstaff	Arizona	AZ	Non-Recourse
Fort Collins	Colorado	CO	Recourse
Fresno	California	CA	Non-Recourse
Glens Falls	New York	NY	Recourse
Grand Junction	Colorado	CO	Recourse
Green Bay	Wisconsin	WI	Non-Recourse
Greenville	South Carolina	SC	Recourse
Hanford	California	CA	Non-Recourse
Hartford	Connecticut	CT	Recourse
Johnson City	Tennessee	TN	Recourse
Knoxville	Tennessee	TN	Recourse
Lafayette-West Lafayette	Indiana	IN	Recourse

Table 7 (continued)

Metropolitan area	State ¹	Abbreviation	Recourse / Non-recourse
Lancaster	Pennsylvania	PA	Recourse
Las Vegas	Nevada	NV	Recourse
Lincoln	Nebraska	NE	Recourse
Little Rock	Arkansas	AR	Recourse
Los Angeles-Long Beach-Anaheim	California	CA	Non-Recourse
Madera	California	CA	Non-Recourse
Madison	Wisconsin	WI	Non-Recourse
Medford	Oregon	OR	Non-Recourse
Memphis	Tennessee	TN	Recourse
Merced	California	CA	Non-Recourse
Merced	California	CA	Non-Recourse
Milwaukee	Wisconsin	WI	Non-Recourse
Minneapolis-St Paul	Minnesota	MN	Non-Recourse
Mobile	Alabama	AL	Recourse
Modesto	California	CA	Non-Recourse
Morristown	Tennessee	TN	Recourse
Napa	California	CA	Non-Recourse
Nashville	Tennessee	TN	Recourse
New Haven	Connecticut	CT	Recourse
New London	Connecticut	CT	Recourse
North Port-Sarasota-Bradenton	Florida	FL	Recourse
Oklahoma City	Oklahoma	OK	Recourse
Olympia	Washington	WA	Non-Recourse
Philadelphia	Pennsylvania	PA	Recourse
Phoenix	Arizona	AZ	Non-Recourse
Pittsburgh	Pennsylvania	PA	Recourse
Pittsfield	Massachusetts	MA	Recourse
Portland	Oregon	OR	Non-Recourse
Prescott	Arizona	AZ	Non-Recourse
Providence	Rhode Island	RI	Recourse
Pueblo	Colorado	CO	Recourse
Raleigh	North Carolina	NC	Non-Recourse
Redding	California	CA	Non-Recourse
Reno	Nevada	NV	Recourse
Richmond	Virginia	VA	Recourse
Riverside	California	CA	Non-Recourse
Sacramento	California	CA	Non-Recourse
Salem	Oregon	OR	Non-Recourse
Salinas	California	CA	Non-Recourse
Salisbury	Maryland	MD	Recourse
San Diego	California	CA	Non-Recourse
San Francisco	California	CA	Non-Recourse

Table 7 (continued)

Metropolitan area	State ¹	Abbreviation	Recourse / Non-recourse
San Jose	California	CA	Non-Recourse
San Luis Obispo	California	CA	Non-Recourse
Santa Cruz	California	CA	Non-Recourse
Santa Maria-Santa Barbara	California	CA	Non-Recourse
Santa Rosa	California	CA	Non-Recourse
Seattle	Washington	WA	Non-Recourse
Spartanburg	South Carolina	SC	Recourse
Spokane	Washington	WA	Non-Recourse
Springfield	Massachusetts	MA	Recourse
Springfield	Ohio	OH	Recourse
Stamford	Connecticut	CT	Recourse
State College	Pennsylvania	PA	Recourse
Stockton	California	CA	Non-Recourse
Toledo	Ohio	OH	Recourse
Tucson	Arizona	AZ	Non-Recourse
Urban Honolulu	Hawaii	HI	Recourse
Utica	New York	NY	Recourse
Vallejo	California	CA	Non-Recourse
Ventura	California	CA	Non-Recourse
Virginia Beach	Virginia	VA	Recourse
Visalia	California	CA	Non-Recourse
Worcester	Massachusetts	MA	Recourse
Yakima	Washington	WA	Non-Recourse
York	Pennsylvania	PA	Recourse
Yuba City	California	CA	Non-Recourse
Yuma	Arizona	AZ	Non-Recourse

Table 8 Predictive power of MDRI for the house price appreciation rates (Δ HP)

Panel A. Subsample 2007–2012					
Dependent Variable	Δ HP	Dependent Variable	Δ HP	Dependent Variable	Δ HP
	(1)		(1)		(1)
Δ MDRI _{<i>t</i>-1}	-0.0003				
Δ MDRI _{<i>t</i>-2}		0.0011**			
Δ MDRI _{<i>t</i>-3}					
Δ MDRI _{<i>t</i>-4}					
Δ MDRI _{<i>t</i>-5}					
Δ MDRI _{<i>t</i>-6}					
Δ HP _{<i>t</i>-1}					
HF _{<i>t</i>-1}					
Subprime _{<i>t</i>-12}					
Δ Loan supply _{<i>t</i>-12}					
Recourse					
Constant					
Number of Obs					
Adjusted R-squared					
Panel B. Subsample 2013–2016					
Dependent Variable	Δ HP	Dependent Variable	Δ HP	Dependent Variable	Δ HP
	(1)		(1)		(1)
Δ MDRI _{<i>t</i>-1}	-0.0002				
Δ MDRI _{<i>t</i>-2}		-0.0002			
Δ MDRI _{<i>t</i>-3}					
Δ MDRI _{<i>t</i>-4}					
Δ MDRI _{<i>t</i>-5}					

Table 9 Predictive power of MDRI for fundamental house price appreciation rates

Dependent Variable	Fundamental house price appreciation rate							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	-0.0007**						-0.0007**	-0.0007**
ΔMDRI_{t-2}		0					-0.0002	-0.0003
ΔMDRI_{t-3}			0.0001				0	-0.0006*
ΔMDRI_{t-4}				-0.0007**				-0.0013***
ΔMDRI_{t-5}					-0.0012***			-0.0016***
ΔMDRI_{t-6}						0.0003		-0.0003
$\Delta \text{Fundamental}_{t-1}$	0.8393***	0.8394***	0.8394***	0.8391***	0.8390***	0.8395***	0.8393***	0.8380***
HF_{t-1}	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
Subprime_{t-12}	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
$\Delta \text{Loan supply}_{t-12}$	0.0008***	0.0008***	0.0008***	0.0007***	0.0007***	0.0008***	0.0008***	0.0007***
Recourse	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0002***
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs	19,254	19,254	19,254	19,254	19,254	19,254	19,254	19,254
Adjusted R-squared	0.784	0.784	0.784	0.784	0.7842	0.784	0.784	0.7844

Loan supply_{t-12} is the total supply of mortgage loans in the previous year in the MSA (derived from HMDA data). Subprime_{t-12} is the percentage of the total amount of mortgage loans in the MSA that are subprime. The Recourse variable takes on the value of one, if the MSA is in a recourse state, and zero otherwise. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

Table 10 Predictive power of MDRI for the bubble component of house prices

Dependent Variable	Change of bubble component							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔMDRI_{t-1}	0.0001						0.0001	0.0001
ΔMDRI_{t-2}		0.0004					0.0002	0.0002
ΔMDRI_{t-3}			-0.0007**				-0.0006**	-0.0007**
ΔMDRI_{t-4}				-0.0003				-0.0005
ΔMDRI_{t-5}					0.0005*			0.0001
ΔMDRI_{t-6}						-0.0008***		-0.0008**
$\Delta \text{Bubble}_{t-1}$	0.3056***	0.3056***	0.3057***	0.3055***	0.3057***	0.3058***	0.3058***	0.3059***
HF_{t-1}	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
Subprime_{t-12}	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**
$\Delta \text{Loan supply}_{t-12}$	0	0	0	0	0	0	0	0
Recourse	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**	-0.0001**
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,254	19,254	19,254	19,254	19,254	19,254	19,254	19,254
Adjusted R-squared	0.0966	0.0967	0.0969	0.0967	0.0968	0.097	0.0968	0.0971

The table presents regression estimates of the effect of the MDRI index on the difference in the bubble component of the house price index. $\text{Loan supply}_{t-12}$ is the total supply of mortgage loans in the previous year in the MSA (derived from HMDA data). Subprime_{t-12} is the percentage of the total amount of mortgage loans in the MSA that are subprime. The Recourse variable takes on the value of one, if the MSA is in a recourse state, and zero otherwise. One, two, and three asterisks represent significance at the 10, 5, and 1%, respectively

Table 11 Predictive power of MDRI for the homes foreclosed

Panel A. Subsample 2007–2012							
Dependent Variable	Homes Foreclosed (HF)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
							(8)
ΔMDRI_{t-1}	0.0071						-0.0607***
ΔMDRI_{t-2}		-0.1048***					-0.1633***
ΔMDRI_{t-3}			-0.0453***				-0.1080***
ΔMDRI_{t-4}				0.0356**			-0.0963***
ΔMDRI_{t-5}					0.0508***		0.0055
ΔMDRI_{t-6}						-0.0607***	0.0516***
ΔHP_{t-1}	-0.7085***	-0.7419***	-0.7191***	-0.6954***	-0.6907***	-0.7315***	-0.0258
HF_{t-1}	0.9729***	0.9731***	0.9729***	0.9729***	0.9730***	0.9728***	-0.7794***
Subprime_{t-12}	0.0055***	0.0060***	0.0055***	0.0054***	0.0055***	0.0055***	0.9733***
$\Delta \text{Loan supply}_{t-12}$	-0.0136*	-0.0068	-0.0123*	-0.0139*	-0.0140**	-0.0134*	0.0062***
Recourse	-0.0193***	-0.0194***	-0.0191***	-0.0193***	-0.0194***	-0.0190***	0.0011
Constant	Yes	Yes	Yes	Yes	Yes	Yes	-0.0190***
Number of Obs	9316	9316	9316	9316	9316	9316	Yes
Adjusted R-squared	0.9644	0.9646	0.9645	0.9645	0.9645	0.9645	9316
							0.9647
							0.9648
Panel B. Subsample 2013–2016							
Dependent Variable	Homes Foreclosed (HF)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
							(8)
ΔMDRI_{t-1}	0.0750**						0.0015
ΔMDRI_{t-2}		-0.1787***					-0.2257***
ΔMDRI_{t-3}			-0.0914***				-0.1780***
ΔMDRI_{t-4}				-0.018	-0.0963***		-0.1458***
ΔMDRI_{t-5}							-0.1791***

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abraham, J. M., & Hendershott, P. H. (1996). Bubbles in metropolitan housing markets. *Journal of Housing Research*, 191–207.
- Ashworth, J., & Parker, S. C. (1997). Modelling regional house prices in the UK. *Scottish Journal of Political Economy*, 44(3), 225–246.
- Avery, R. B., Brevoort, K. P., & Canner, G. B. (2007). The 2006 HMDA data, *Federal Reserve Bulletin*, December, A73-A109.
- Bajari, P., Chu, C. S., & Park, M. (2008). An empirical model of subprime mortgage default from 2000 to 2007. *NBER working paper* (no. w14625).
- Beracha, E., & Wintoki, M. B. (2013). Forecasting residential real estate price changes from search activity. *Journal of Real Estate Research*, 35(3), 283–312.
- Bhutta, N., Dokko, J., & Shan, H. (2017). Consumer ruthlessness and mortgage default during the 2007 to 2009 housing bust. *Journal of Finance*, 72(6), 2433–2466.
- Calomiris, C. W., Longhofer, S. D., & Miles, W. R. (2013). The foreclosure–house price nexus: A PVAR model for U.S. states, 1981–2009. *Real Estate Economics*, 41(4), 709–746.
- Capozza, D. R., & Seguin, P. J. (1996). Expectations, efficiency, and euphoria in the housing market. *Regional Science and Urban Economics*, 26, 369–386.
- Capozza, D. R., Hendershott, P. H., & Mack, C. (2004). An anatomy of price dynamics in illiquid markets: Analysis and evidence from local housing markets. *Real Estate Economics*, 32(1), 1–32.
- Case, K. E., & Shiller, R. (1989). The efficiency of the market for single-family homes. *American Economic Review*, 79, 125–137.
- Case, K. E., & Shiller, R. (1990). Forecasting prices and excess returns in the housing market. *AREUEA Journal*, 18, 253–273.
- Chauvet, M., Gabriel, S., & Lutz, C. (2016). Mortgage default risk: New evidence from internet search queries. *Journal of Urban Economics*, 96, 91–111.
- Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, 20(2), 249–272.
- Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *The Economic Record*, 88, 2–9.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, 66(5), 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), 1–32.
- De Boor, C. (1978). A practical guide to splines. Series: Applied Mathematical Sciences, 27, Springer.
- Deng, Y., Quigley, J. M., & Van Order, R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, 68(2), 275–307.
- Dorsey, R. E., Hu, H., Mayer, W. J., & Wang, H. (2010). Hedonic versus repeat-sales house price indexes for measuring the recent boom-bust cycle. *Journal of Housing Economics*, 19(2), 75–93.
- Drake, L. (1993). Modeling UK house prices using cointegration: An application of the Johansen technique. *Applied Economics*, 25, 1225–1228.
- Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., & Hunt, R. M. (2010). What ‘triggers’ mortgage default? *American Economic Review*, 100(2), 490–494.
- Foote, C. L., Gerardi, K., & Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics*, 64(2), 234–245.
- Foster, C., & Van Order, R. (1984). An option-based model of mortgage default, *Housing Finance Review*, 3(4), 351–372.
- Gerardi, K., Herkenhoff, K. F., Ohanian, L. E., & Willen, P. S. (2018). Can’t pay or won’t pay? Unemployment, negative equity, and strategic default. *Review of Financial Studies*, 31(3), 1098–1131.

- Ghent, A. C., & Kudlyak, M. (2011). Recourse and residential mortgage default: Evidence from U.S. states. *Review of Financial Studies*, 24(9), 3139–3186.
- Guiso, L., Sapienza, P., & Zingales, L. (2013). The determinants of attitudes toward strategic default on mortgages. *Journal of Finance*, 68(4), 1473–1515.
- Kasparova, D., & White, M. (2001). The responsiveness of house prices to macroeconomic forces: A cross country comparison. *European Journal of Housing Policy*, 1(3), 385–416.
- Kau, J. B., Keenan, D. C., & Kim, T. (1994). Default probabilities for mortgages. *Journal of Urban Economics*, 35(3), 278–296.
- Kelly, R., & McCann, F. (2016). Some defaults are deeper than others: Understanding long-term mortgage arrears. *Journal of Banking and Finance*, 72, 15–27.
- LaCour-Little, M., Lin, Z., & Yu, W. (2020). Assumable financing redux: A new challenge for appraisal? *Journal of Real Estate Finance and Economics*, 60, 3–39.
- Lin, Z., Rosenblatt, E., & Yao, V. (2009). Spillover effects of foreclosures on neighborhood property values. *Journal of Real Estate Finance and Economics*, 38, 387–407.
- Mayer, C. (2011). Housing bubbles: A survey. *Annual Review of Economics*, 3(1), 559–577.
- Mayer, C., & Pence, K. (2008). Subprime mortgages: What, where, and to whom?, NBER working Paper No. 14083.
- Mocetti, S., & Viviano, E. (2017). Looking behind mortgage delinquencies. *Journal of Banking and Finance*, 75, 53–63.
- Rose, L. A. (1989). Urban land supply: Natural and contrived restrictions. *Journal of Urban Economics*, 25, 325–345.
- Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics*, 125(3), 1253–1296.
- Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The Annals of the American Academy of Political and Social Science*, 645, 112–141.
- Stevenson, S. (2008). Modeling housing market fundamentals: Empirical evidence of extreme conditions. *Real Estate Economics*, 36(1), 1–29.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3), 1139–1168.
- Van Dijk, D. W., & Francke, M. K. (2018). Internet search behavior, liquidity and prices in the housing market. *Real Estate Economics*, 46(2), 368–403.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.